# Controlling drift when extending index series of consumer price change over time


Antonio G. Chessa

Department of Consumer Prices
Statistics Netherlands



Paper prepared for the 64th ISI World Statistics Congress
Ottawa, Canada

July 2023

# Abstract

The increased acquisition of transaction data sets from retail chains encourages national statistical institutes to study multilateral methods in order to better exploit the detailed sales data and product dynamics when compiling the Consumer Price Index (CPI). In addition, multilateral methods have the advantage of producing time series of aggregate price change that are transitive. Such time series are free of "chain drift", which means that direct and indirect price comparisons between two periods give the same result.

Unfortunately, transitivity no longer holds when an initial index series is extended over time after including data from future periods. Index series compiled on successive time windows can be linked to previously calculated index series in numerous ways. This paper characterises and formalises these splicing or extension methods, which are compared in a study that includes transaction data sets on a broad range of product types.

The results show that linking on the most recently updated or recalculated index of a half year ago produces drift-free indices and 12-month rates of change (inflation) at all-items level for 25-month windows, while also 13-month windows give very good results. These methods also produce accurate results at more detailed product aggregate levels, in particular for 25-month windows. The mean splice method, which averages over all linking periods, also gives good results but may suffer from window censoring effects. Linking on published indices of one year ago produces accurate 12-month rates of change, but monthly rates of change can be inaccurate. The choice of linking period can be optimised each month, but improvements over the aforementioned methods are rather small.



**Keywords**

Consumer price index, inflation, multilateral methods, index extension, transaction data.

# Contents

# 1 Introduction

National statistical institutes (NSIs) compile and publish figures on a wide range of economic and other indicators that are used as instruments for policy making and indexation purposes. Examples are Gross Domestic Product (GDP), which is used as a measure of economic growth, and the Consumer Price Index (CPI), which is considered as the most important indicator of inflation. Most NSIs compile and publish their CPI each month, so the process from data collection and processing to index compilation and publication follows a tight schedule.

Data acquisition primarily consists of collecting prices for physical goods and services ("products") offered to or actually purchased by consumers for sets of products that fit within the international COICOP classification system.[1] Most countries still collect prices in traditional ways, which consists of recording prices for relatively small sets of products in physical outlets by pencil and paper or with electronic devices.

Over the past 20 years, data collection has started to gradually shift from traditional price collection to automated collection of large amounts of electronic data. Web scraping of consumer prices emerged about 10 years ago, which became increasingly popular during the COVID-19 pandemic as NSIs were not allowed to visit physical outlets in periods of lockdown. The surge of web scraping was preceded by efforts of acquiring product sales data from retail chains, which were pioneered by the statistical offices of Norway and the Netherlands more than 20 years ago.

This paper only focuses on sales data, which are widely known as *transaction data* or by the more restrictive term *scanner data*, which specifically refers to goods and services that are uniquely identified by a bar code. A growing number of NSIs are trying to acquire transaction data, which are now used by more than 10 countries in their CPI compilation. The share of transaction data used in the CPI of Statistics Netherlands, measured in terms of COICOP weights, will surpass 40 per cent in the course of 2023.

Transaction data constitute a very attractive data source for replacing traditionally collected prices for different reasons. Sample sizes in traditional price collection are typically in the order of several tens or hundreds of products per retail chain, while transaction data sets may easily contain several tens or hundreds of thousands of sold products at the article code or bar code level, which is officially known as the Global Trade Item Number (GTIN). In this paper, products at this level will be referred to as *items*, while the term *product* is more generic and may be broader than the item level, for instance, in order to capture price increases associated with product relaunches that are assigned new bar codes [9].

A second important feature of transaction data is the inclusion of expenditures and numbers of products sold at the GTIN or item level, which are usually specified by week, aggregated over consumers. This allows the derivation of transaction prices, which are the prices paid by an average consumer for each sold item in different time periods. Traditionally collected prices and web scraped prices are offer or shelf prices. Expenditures

[1] See `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:COICOP_HICP`

and numbers of sold items are not collected in these forms of data collection.

A major benefit of transaction data is the possibility of defining expenditure based product weights. This is very important, since the price development of each product will be reflected in the price index of the corresponding product aggregate in accordance with the composition of consumers' baskets and changes in basket composition over time. Price index formulas that make use of both product prices and sold quantities are therefore obvious options to consider when processing transaction data compared to index formulas that only make use of prices.

The class of index formulas that include both prices and quantities is quite broad. An important distinctive feature is the number of periods used when computing aggregate price change between two periods. *Bilateral* formulas make use of data from two periods, while *multilateral* index formulas incorporate data from more than two periods ([3], [12], [16]).

Index formulas have to be applied in each successive period in order to compile index numbers with newly collected data. An important question is how to link the indices calculated in each period to index values in past periods. There are different ways of linking, which follow from the different choices that can be made for the reference period in an index formula. The term "reference period" denotes the period from which prices and quantities are compared with the values in the second period, which is typically the current period in practice. Index formulas that yield the same result for each linking choice are called *transitive* and are said to be "free of chain drift" [25]. Transitive index series are independent of the choice of reference period, which is a desirable property. The problem of linking subsequent indices to past index values becomes irrelevant for index formulas that are transitive.

Unfortunately, bilateral index formulas that make use of both product prices and quantities are sensitive to chain drift. These methods also include the Fisher index (see e.g. [26]), which is the only index formula that satisfies all 20 "axioms" or "tests" for bilateral indices listed in the CPI manual ([25], pp. 292-297). The problem of chain drift can be resolved to a certain extent with multilateral methods. However, although these methods produce transitive index series on fixed time intervals, transitivity will be lost once the initial transitive index series is extended when linking index series compiled with data from successive periods ([7], [8], [30]). The problem of chain drift is illustrated in Section 2.

Since multilateral methods make use of data from multiple periods and the number of periods (window length) can be varied, these methods offer a wider range of choices for linking successive index series than bilateral methods. In theory, this means that multilateral methods are better equipped for controlling chain drift. For this reason, the focus in the remainder of this paper will be on multilateral methods, with the objective to identify index extension methods that are capable of controlling chain drift.

Section 3 starts with a characterisation of the index extension problem. The identification of the choice variables serves as a basis for defining extension methods. Some well-known methods and more sophisticated optimisation methods are described afterwards.

Section 4 summarises the transaction data sets and extension methods that have been included in a large comparative study. The results of this study are presented and discussed in sections 5 and 6. Section 7 concludes.

# 2 The index extension problem and chain drift

## 2.1 Bilateral methods

This section serves as an introduction to the index extension problem and related issues, before giving a comprehensive characterisation and formalisation of the problem in the next section. In mathematical terms, a price index can be conceived of as a function of prices and quantities (if available) for a set of products in different periods ([3], [25]). A price index provides a measure for the composite or aggregate price change of a set of products between two periods. An index is a dimensionless ratio number, which is used to express aggregate price change between two periods as a percentage.

In practice, one of the two periods is the current or reporting period and the other period can be any period in the past, which is referred to as “reference period”. When data from only two periods are available, it is clear that any index formula can be applied in only one way and will give a unique result. This may change as soon as data from a third period becomes available. The new data must be used in order to compile an index for the next reporting period. One of the two periods in an index formula is the new reporting period, while different choices can now be made for the reference period.

The reference period also serves as the linking period on which the newly compiled index is linked. The linking results in an extension of an index series to the current reporting period. Although different choices for the reference period can be made, there are two methods for extending series of bilateral indices that price statisticians are familiar with:

- The *fixed base* or *direct* approach, which uses the same reference period in each reporting period. If $P_{0,t}$ denotes an index in period $t$ with 0 as reference period, then the direct or fixed base method yields a sequence $P_{0,0}, P_{0,1}, ..., P_{0,t}$ of index numbers, where $P_{0,0} = 1$, which denotes ‘no change’.
- *Period to period chaining* computes an index in period $t$ with respect to period 0 by combining all intermediate price comparisons for two successive periods as follows: $\prod_{z=1}^{t} P_{z-1,z}$.

An important question is whether the two linking or extension methods give the same result. Equality is achieved for index formulas that are transitive. An index formula is said to be *transitive* when each indirect comparison between any two periods $z$ and $t$ through some period $s$ yields the same result as the direct comparison between $z$ and $t$, which can be formalised as $P_{z,s}P_{s,t} = P_{z,t}$ for every $s$ and any pair of periods $z$ and $t$. Transitivity is equivalent with the property that fixed base and chained indices are equal. This means that transitivity can alternatively be formulated as the *Multi-Period Identity*

*Test*, which is satisfied when

$$P_{t,z} \prod_{s=z+1}^{t} P_{s-1,s} = 1 \tag{2.1}$$

for all $z$ and $t$. This test states that if we return to the initial point $z$, after moving from $z$ to $t$ through a sequence of intermediate points $s$ in time, then a price index is expected to return the value 1. If this does not happen, then an index formula is said to suffer from *chain drift* ([25], p. 283).

The class of bilateral index formulas can roughly be subdivided between formulas that only make use of prices and formulas that also include product quantities. This distinction is sometimes referred to as "unweighted" and "weighted" methods by price statisticians. The Jevons index, which calculates a price index as a geometric mean of the price ratios of each product between two periods, is a broadly supported example of an unweighted method. It is able to deal with sets of products that differ in quality and produces transitive index series, subject to the condition that prices must be available for each product in every period.

Although NSIs are showing greater interest in multilateral methods, the Jevons index formula is still used by numerous NSIs for processing supermarket transaction data. Eurostat set up guidelines on how the Jevons can be applied to transaction data [21], which is particularly useful for NSIs that want to start using transaction data in the production of their CPI. A downside of the Jevons is that each product has the same weight so that shifts in expenditures among products are not reflected in a Jevons index. For example, consumers may switch to products with smaller price changes, in which case Jevons indices will overstate aggregate price change. It may be possible to limit such distortions in an index series by setting a threshold on expenditure share ("low-sales filter"), which is used by NSIs to exclude products that remain below the threshold [21].

It is therefore important to include both prices and quantities in an index formula. Unfortunately, it will be harder to achieve transitivity when additional dynamics beside price changes over time are included (see also the discussion in CPI manual [25] on page 282). The bilateral index formulas known to date that allow variations in quantities over time are not transitive. Fixed base indices only depend on the prices and quantities of the current and fixed reference period, while chained indices also depend on the prices and quantities of intermediate periods.

The above discussion emphasises the problems with bilateral index formulas in reflecting consumers' changes in expenditure patterns and meeting the transitivity property at the same time. The importance and sensitivity of the choice of product weights and linking method is highlighted by several examples in [12], where different bilateral and multilateral methods are applied and compared on transaction data sets. Examples can also be found in [15] and [26].

## 2.2 Multilateral methods

Multilateral methods were originally developed for international and inter-regional price comparisons. An important feature of these methods is that the resulting Purchasing Power Parities are independent of the choice of reference country or region. Well-known examples of multilateral methods are the GEKS method ([19], [24], [32]), the Country Product Dummy method [31], and the Geary-Khamis method ([23], [27]). The reader may also consult Balk's book on details about these and other multilateral methods [3].

Although multilateral methods have a long history with regard to applications in a geographical setting, their potential for index calculation with transaction data was only studied at the beginning of the previous decade ([15], [26]) after a first application in the time domain in the 1980s [2]. The cited initial studies on transaction data were followed by numerous contributions from the academic world and statistical offices in recent years ([1], [4], [5], [6], [13], [17], [28], [29], [30], [34], [35], [36]).

Sets of countries or regions are replaced by months when applying multilateral methods in the CPI. A set of months or periods with different time units will be referred to as a "time window" in this paper. Multilateral methods generate a sequence of index numbers in each period of a time window after a single computational procedure. The independence of reference period in multilateral indices can be achieved in different ways depending on method. The GEKS method calculates an average of bilateral indices over a time window ([15], [26]), while the Geary-Khamis [6] and the Time Product Dummy method [28] compare product prices in different periods with average (deflated) prices over a time window.

The transitivity property resolves the problems and dilemma encountered with bilateral methods discussed at the end of the previous subsection. At least, this applies to each separate time window since index series are transitive on every window. Unfortunately, transitivity will also be lost for multilateral index series when successive transitive index series are combined, which has to be done in order to derive an index in each reporting period. This issue is discussed and illustrated in greater detail below.

Figure 2.1 gives an example of a typical application of multilateral methods in practice. A window length is first chosen, which is equal to 13 periods in the example. The initial time window starts in period 0 and ends in period 12, which in the CPI typically refer to December of the previous and current year, respectively. December acts as a base month in the CPI, the month in which a new index series is started for each calendar year. A window length of 13 months is recommended as a (minimum) length in order to allow prices of seasonal items to be compared between seasons of two successive years ([26], p. 33).

The first step is to compute an index series on the initial time window. Next, an index has to be compiled in period 13, which in practice would be January of the new year. In order to include the data of this month in the index calculations, the initial time window has to be adjusted in some way. One way of doing this is to shift the initial window by one month, so that the next time window will include the new month while the first month of the initial window is dropped. Applying a multilateral method to the data of periods

1, 2, ..., 13 results in an index series for time window 2.

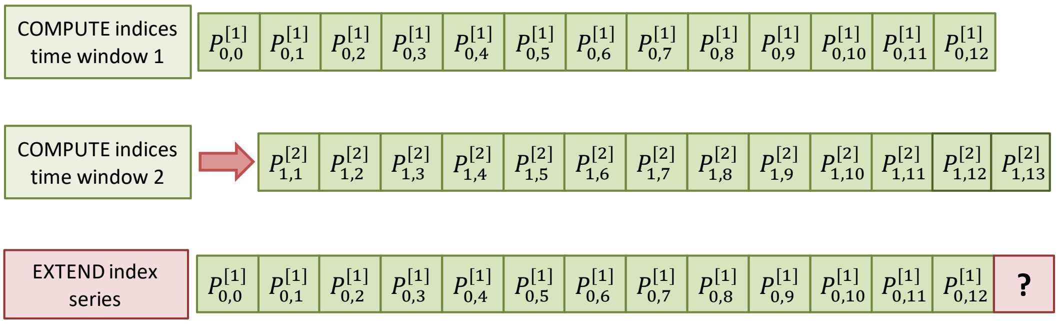


**Figure 2.1:** Illustration of the index extension problem for multilateral methods. The indices on time windows 1 and 2 are denoted by $P^{[1]}_{0,t}$ and $P^{[2]}_{1,t}$, with 0 and 1 as the respective initial periods of the two 13-period windows.

The question now is how the initial index series on time window 1 can be extended with an index in the new month. In order to achieve this, the index series obtained on the two time windows have to be linked. Multilateral methods offer different linking possibilities by making use of the indices generated in each period of the time window. For 13-month windows there are 12 months to choose from in order to link the two index series, which in the example of Figure 2.1 are months 1, 2, ..., 12.

The results for two linking methods are illustrated with a numerical example in Table 2.1. Series 1 and 2 are two transitive index series calculated on two time windows, which are defined as in Figure 2.1. The two index series "Link on 1" and "Link on 12" are equal to series 1 until period 12. "Link on 1" produces an index in period 13 by linking the change in period 13 with respect to period 1 according to series 2 on the index of series 1 in period 1. The index in period 13 for "Link on 12" is obtained by linking the change of series 2 in period 13 with respect to 12 on the index of series 1 in period 12.

| **Period** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Series 1 | 1 | 0.905 | 0.993 | 0.996 | 0.846 | 0.930 | 1.038 | 1.038 | 0.991 | 0.936 | 1.092 | 1.099 | 1.095 | |
| Series 2 | | 1 | 1.097 | 1.100 | 0.933 | 1.029 | 1.147 | 1.148 | 1.094 | 1.035 | 1.208 | 1.216 | 1.212 | 1.290 |
| Link on 1 | 1 | 0.905 | 0.993 | 0.996 | 0.846 | 0.930 | 1.038 | 1.038 | 0.991 | 0.936 | 1.092 | 1.099 | 1.095 | 1.168 |
| Link on 12 | 1 | 0.905 | 0.993 | 0.996 | 0.846 | 0.930 | 1.038 | 1.038 | 0.991 | 0.936 | 1.092 | 1.099 | 1.095 | 1.166 |

**Table 2.1:** Results for two methods that link two index series obtained on two 13-period time windows to produce an index in period 13. Link on 1 and 12 means linking on periods 1 and 12.

"Link on 12" is a method known as "movement splice" and was the first extension method to be proposed for multilateral methods [26]. "Link on 1" is known as "window splice" [28]. Table 2.1 shows that the two linking methods give different indices in period 13, with a rounded difference of 0.2 percentage point. Different linking methods thus yield different results in spite of linking two transitive index series. The difference between the two linking methods is caused by the inclusion of new data in period 13, which of course could not be included in the first time window. The inclusion of new data in subsequent time windows may lead to different index values in past periods compared to the indices calculated on preceding time windows.

Compared with bilateral methods, the problem of chain drift remains and is in fact postponed when using multilateral methods. A question that logically comes to mind is whether there is still anything to be gained with multilateral methods when trying to control chain drift. The answer is affirmative for a number of reasons:

- Transitivity of multilateral index series on subsequent time windows is still a very useful property in this respect, which is lacking in weighted bilateral methods.
- In contrast with bilateral methods, the length of the time window can be treated as a variable with different values to choose from.
- Multilateral methods have other important advantages compared with bilateral methods, such as the timely inclusion of new products without compromising transitivity on fixed time windows. The same holds for the possibility to specify varying product weights, both across products and over time.

The transitivity property on separate time windows can be used to control drift. This property is exploited in different ways in the remainder of this study. For instance, Section 3.3 presents a procedure for determining window length and linking period by minimising a specific objective function that is related to chain drift. The results of a comparative study in Section 5 offer extensive material that quantifies the extent of drift in different extension methods and provides insight into different choices that can be made to limit drift.

# 3 Multilateral index extension methods

## 3.1 Characterisation

The discussion and examples presented in the previous section serve as a starting point towards a comprehensive characterisation of the extension problem and for defining index extension methods. The present section starts with an identification of the variables that can be used to define extension methods. Examples of well-known methods are given in Section 3.2. Figure 2.1 and Table 2.1 already reveal several variables, which are completed and listed below:

- Extension methods are traditionally applied with choices that are kept fixed over time for each variable. It is also possible to allow variables to vary over time. This aspect is referred to as *splicing flexibility* in this paper, which can be either *static* or *adaptive*. The two extension methods in Table 2.1 can be considered as examples of static splicing, since window length and linking period are fixed over time.
- In order to calculate a multilateral index series, it is necessary to specify a *window length*. The windows in Section 2.2 consist of 13 periods.
- A second choice that concerns the time window is how to adapt a window over time in order to include data from successive periods, which is referred to as *window adaptation* in this paper. Two specific choices can be distinguished based on studies

on this topic: a fixed-length *rolling window*, which is shifted by one period when new data become available, and an *expanding window*. The examples in Section 2.2 make use of a rolling window.

- The example in Table 2.1 contains two methods that link on different periods. The choice of linking period thus constitutes the fourth variable. Instead of linking period it may be more convenient to use the length of the *linking interval* when formalising extension methods. The (length of the) linking interval is defined as the number of periods between the current or reporting period and the linking period. The linking interval is equal to 12 when linking on period 1 in the example of Table 2.1 and is equal to one period when linking on period 12.
- Different choices can also be made for the index in the linking period, which is referred to as the *linking index*. Suppose that period 14 were added to the example of Table 2.1. If we would link the index series obtained on the third time window to one of the periods between 2 to 12, then we can choose between the indices of the first two series. The indices of series 2 are *recalculated indices*, while the first calculated index series will result in a set of *published indices*. There is only one index to link on in previous period 13, which will always be a published index.

The variables listed above can be used to parametrise and organise extension methods according to a tree structure, which is shown in Figure 3.1. The red boxes at the lowest level of the tree contain extension methods. The static splicing methods that make use of a rolling window with a fixed length $W$ and linking interval $L$ are denoted by $S(W, L, r)$ and $S(W, L, p)$, where $r$ and $p$ indicate whether the linking index is a recalculated or a published index. The abbreviation MS stands for mean splice. Linking interval is omitted, since mean splice calculates an average index over all linking intervals.

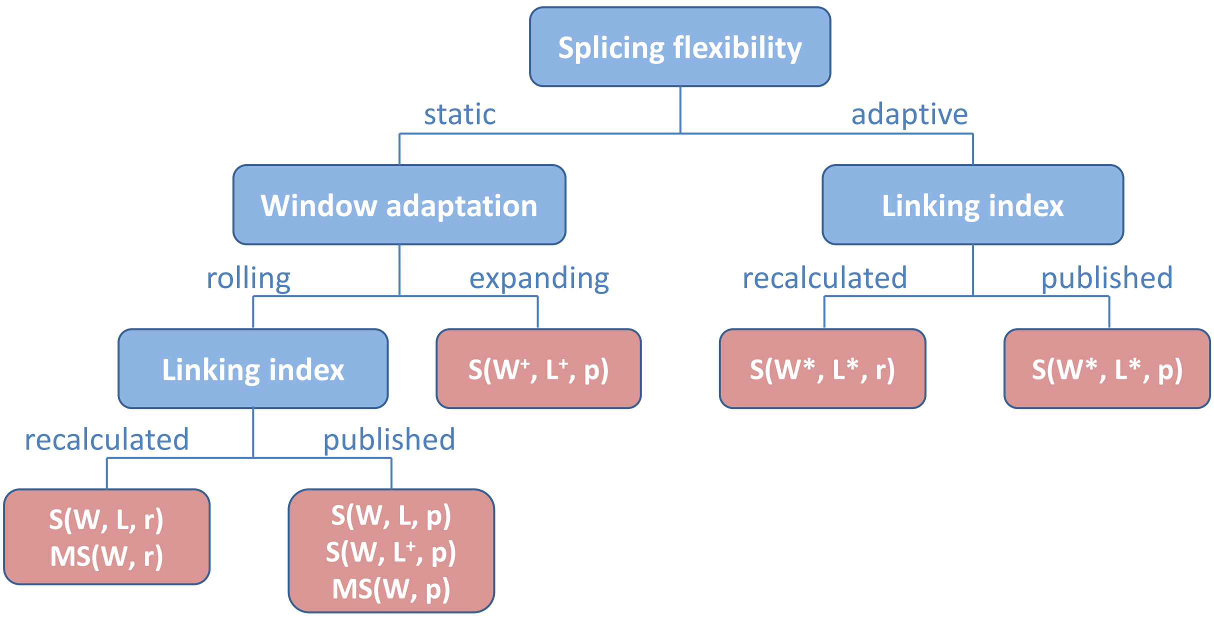


**Figure 3.1:** Characterisation of extension methods (red boxes) and variables in a tree structure.

An expanding window was originally designed with the CPI practice of compiling annual series in mind. December of the previous year is the first month of the time window, which contains two months in January, three in February and reaches its maximum length of 13 months in December of the running year. Window length is thus increased by one month when data of the next month become available, and is denoted as $W^+$ in Figure 3.1 for this type of method. The index published in December serves as the linking index in every reporting month. Index series are extended by linking fixed base indices calculated on each expanding window on the index in December of the previous year. This extension method is known as Fixed Base Expanding Window (FBEW) method [6] and is denoted in Figure 3.1 as $S(W^+, L^+, p)$. The linking interval $L^+$ also increases each month and is equal to $W^+ - 1$.

A method similar to FBEW is the Fixed Base Rolling Window (FBRW) method, with the only difference that FBRW uses a rolling window instead of an expanding window [29]. The FBRW method also links on the published December indices. Since the linking period is fixed, the linking interval increases each next period as in FBEW, so that the FBRW class of methods can be denoted as $S(W, L^+, p)$. This method can be found in the second red box from the left in Figure 3.1.

Although window length and/or linking interval are not constant over time in FBEW and FBRW, these methods are not considered forms of adaptive splicing since the linking period is fixed. Window length and linking interval are left free in adaptive splicing and may thus take any value, which are respectively denoted as $W^*$ and $L^*$ in Figure 3.1. The values of these variables are determined by solving an optimisation problem. More details will be given in Section 3.3. In theory, it is also possible to let the linking index vary across periods, but this option was not considered in this study. The adaptive splicing methods constitute a new class of methods on the topic of index extension. For this reason, it was decided to limit the complexity of these methods in this study.

## 3.2 Static splicing methods

The extension methods presented in Figure 3.1 will be formalised in the remainder of this section. The present subsection starts with the static splicing methods, which continues with adaptive splicing in Section 3.3. Well-known methods belonging to the class of static splicing methods emerge as special cases from the formal expressions and will be discussed separately.

The following notation is introduced. A published index in period $t$ is denoted by $P_{0,t}$, where 0 stands for the first period of an index series. The initial time window is denoted as $[0, T]$, which consists of $T + 1$ periods $0, 1, ..., T$. The number of periods $T + 1$ is the length $W$ of the time window, which is fixed for rolling windows.

Indices obtained on a specific time window are denoted in the same way as published indices, with an additional reference to the time window. For instance, $P_{t-1,t}^{[t-T,t]}$ denotes the index in period $t$ with respect to reference period $t - 1$ that results from the index series computed on time window $[t - T, t]$. An exception is made for indices calculated on the initial time window $[0, T]$, which are denoted by $P_{0,t}$ without referring to the time window.

This is done for reasons of convenience when formalising expressions for extended indices.

The description of the static splicing methods starts with methods that link on published indices, which have a simpler construction than methods that link on recalculated indices.

### 3.2.1 Linking on published indices

#### 3.2.1.1 Fixed linking interval

A number of well-known extension methods make use of a rolling window and a fixed linking interval. For windows with $T+1$ periods and linking interval $L$, extension methods that link on published indices can be denoted by the class $S(T+1, L, p)$ according to the notation in Figure 3.1.

The extension method for $L = 1$ is known as *movement splice*. This is basically the multilateral version of period to period chaining. It is the first extension method that was proposed, in studies where index series calculated with the GEKS method are extended by period to period chaining ([15], [26]). For 13-month windows ($T = 12$) the method is known as Rolling Year GEKS (RYGEKS).

Extension methods belonging to the class $S(T + 1, L, p)$ can be formalised as follows. Suppose that indices are published until period $t - 1$ and that an index $P_{0,t}$ has to be calculated and published in period $t$. The time window is shifted in order to include the data of period $t$, which results in the window $[t-T, t]$. The index $P_{t-L,t}^{[t-T,t]}$ in period $t$ with respect to $t - L$ of the current window is linked to the published index in period $t - L$. The published index in period $t > T$ can thus be calculated according to the following expression:

$$P_{0,t} = P_{0,t-L} P_{t-L,t}^{[t-T,t]} \tag{3.1}$$

Movement splice is sensitive to drift because of its short-term chaining nature ([7], [8]). The cited studies support linking on published indices, but with longer linking intervals in order to better control for drift over longer periods. Linking on published indices allows to preserve drift-free properties of the transitive index series calculated on subsequent time windows in the eventually published indices.

An extension method that uses a longer linking interval is the method known as *Half Splice on Published indices* (HASP). The method HASP takes the central period as linking period for windows with odd numbers of periods, which means that the linking interval is equal to $\frac{T}{2}$. The linking interval can be set at $\frac{T \pm 1}{2}$ for windows with even numbers of periods. Figure 3.2 gives an illustration of the method HASP.

A 25-month rolling window is recommended in practical applications, which is particularly convenient for products that may be temporarily unavailable, like seasonal items. Longer windows compensate for a higher sparseness in the data ([7], [8]).

Setting $L = T$ results in the method known as *Window Splice on Published indices* (WISP), in which the first period of each time window is the linking period. This method was proposed in the same two aforementioned studies in order to avoid or reduce drift in the original window splice method, which links on recalculated indices (see Section

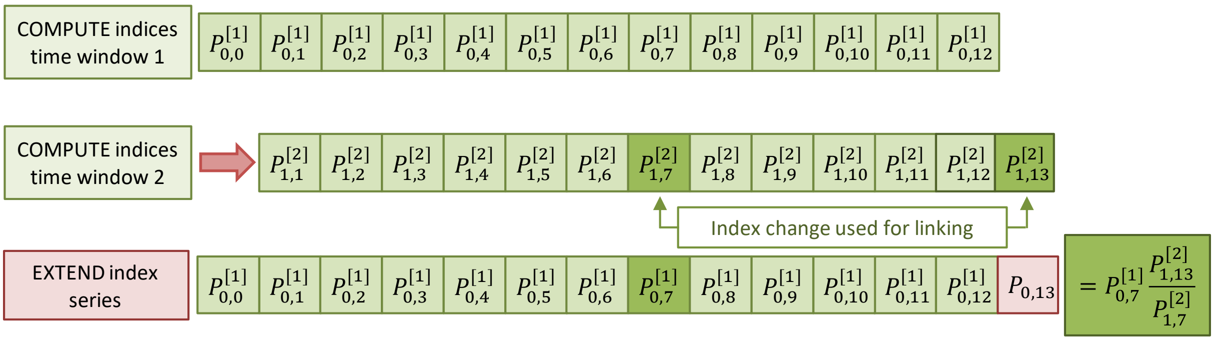


**Figure 3.2:** Illustration of the method HASP for one index extension using 13-period windows. The dark green coloured cell in the extended series denotes the linking period and index.

3.2.2.1).

#### 3.2.1.2 Fixed linking period

The linking periods in movement splice, HASP and WISP are shifted one period when a time window is shifted. Some extension methods make use of a fixed linking period. December of the previous year is used in the CPI as linking month in order to link annual index series and extend these series to longer term series of higher aggregates. The same choice is made in the fixed base window methods for multilateral methods, which, according to EU regulations, can be applied to the lowest aggregate of publication in the CPI (COICOP 5-digit level).

Two fixed base methods are known in the literature on index methodology: the Fixed Base Expanding Window (FBEW) method [6] and the Fixed Base Rolling Window (FBRW) method, which was proposed in response to the FBEW method [29]. Both methods link fixed base indices computed on each subsequent time window to the published index in a fixed reference or base period.

Some additional notation is required to formalise this type of methods because of the role of the base period. For notational convenience, let the base period be updated every $T$ periods, which means that 0 and $T$ represent the first two base periods. Let $bT$ denote the base period after $b$ updates, where $b$ is a non-negative integer. The published index in base period $bT$ is $P_{0,bT}$.

The FBRW method uses a rolling window and variable linking interval since the linking period is fixed. The linking interval in period $t \in [bT+1,(b+1)T]$ is equal to $t-bT$. The FBRW method can thus be denoted as a method of type $S(T+1, t-bT, p)$. An index to be published in period $t \in [bT+1,(b+1)T]$ can be calculated as follows:

$$P_{0,t} = P_{0,bT} P_{bT,t}^{[t-T,t]} \tag{3.2}$$

where $P_{bT,t}^{[t-T,t]}$ denotes the fixed base index in period $t$ with respect to the fixed reference period $bT$, which results from the index series calculated on time window $[t-T,t]$.

The FBEW method only differs from the FBRW method in the use of an expanding window instead of a rolling window. The base period $bT$ is the first period of the expanding window, which has length $t - bT + 1$ in period $t \in [bT + 1, (b + 1)T]$. The time window reaches its maximum length in the next base period $(b + 1)T$, in which case the length equals $T + 1$ and will thus coincide with the window length of the FBRW method.

#### 3.2.1.3 Mean splice

The methods described so far use one linking period or interval in order to extend index series from one period to the next. It is also possible to choose different linking periods and combine the results into one index. The method *mean splice* uses all possible linking periods in a time window and calculates an average from the resulting indices. This method was proposed by Diewert and Fox [17].

The version of mean splice that links on published indices can be formalised as follows. Since the method uses each linking interval, the indices that would be published in period $t$ for each choice $L = 1, 2, ..., T$ are given by expression (3.1). The index published in period $t$ according to mean splice is calculated as a geometric average of these indices:

$$P_{0,t} = \left( \prod_{L=1}^{T} P_{0,t-L} P_{t-L,t}^{[t-T,t]} \right)^{\frac{1}{T}} \tag{3.3}$$

Every linking interval is considered to be of equal importance in mean splice [17]. Mean spliced indices can therefore be interpreted as solutions of an ordinary least squares problem, where $\ln P_{0,t}$ minimises the sum of squared differences with $\ln \left( P_{0,t-L} P_{t-L,t}^{[t-T,t]} \right)$ over $L = 1, 2, ..., T$.

### 3.2.2 Linking on recalculated indices

This subsection continues with a formalisation of extension methods that link on recalculated indices. The term “recalculated” refers to the last recalculated index in a past period, which is obtained on the previous time window. Fixed base methods are not included in this description, since these methods link on published indices.

#### 3.2.2.1 Fixed linking interval

Linking on the last recalculated index of a past period is equivalent with a period to period extension for which only the index series computed on the last two time windows are used. Such methods can thus be implemented conveniently as a period to period index chaining method.

The period to period index used for extending an index series from period $t-1$ to $t$ is denoted by $P_{t-1,t}$. This index can be obtained as follows when using a linking interval $L$. First, the index in period $t$ with respect to linking period $t - L$ is derived from the index series on the current window $[t - T, t]$, which yields the index $P_{t-L,t}^{[t-T,t]}$. Next, the index in period $t - 1$ with respect to linking period $t - L$ is derived from the preceding window $[t - 1 - T, t - 1]$, which gives the index $P_{t-L,t-1}^{[t-1-T,t-1]}$. The period to period index that will be linked to the published index in period $t-1$ is equal to the ratio of these two indices:

$$P_{t-1,t} = \frac{P_{t-L,t}^{[t-T,t]}}{P_{t-L,t-1}^{[t-1-T,t-1]}} \tag{3.4}$$

The index to be published in period $t$ can thus be obtained as follows:

$$P_{0,t} = P_{0,t-1} P_{t-1,t} \tag{3.5}$$

with $P_{t-1,t}$ given by (3.4).

Index series constructed according to this expression belong to the class of methods $S(T+1, L, r)$, which appear in the leftmost branch of Figure 3.1. Well-known methods of this class are *window splice*, for which the linking interval $L = T$. In other words, window splice takes the first period of each time window as linking period. This method was proposed by Krsinich in order to overcome problems with movement splice. She argued that "the revised movement for back periods is not being incorporated into the longer-term index movement" when using movement splice ([28], p. 16). For the same reason, Krsinich states that movement splice does not properly handle contributions of new products to an index.

Revision of published indices is not allowed in the CPI. However, revised or recalculated indices in past periods can be used to link index changes, which will consequently have an effect on the index values that will be published in the present reporting period. Window splice takes the effects of these index revisions into account. However, de Haan notes that "a potential problem with Krsinich's (2014) method is that it does not revise for items that are only observed in the first month of the estimation window" and suggests to "extend the estimation window by 12 months prior to the linking month" in Krsinich's 13-month window ([14], p. 26).

Chessa found that window splice is very sensitive to downward index drift, which typically arises in situations where items leave an assortment at extremely reduced prices during clearance sales after the first period of a time window ([7], pp. 12-13). In a follow-up study, the same author states that this downward index behaviour can be resolved by shifting the linking period towards the middle of the time window. Items that leave an assortment under clearance prices soon after the linking period result in an index that will be less sensitive to downward drift since such items will contain regular prices before the linking period ([8], p. 29). This will be explained further in Section 6.1.

Chessa's observations and suggestions motivated the method HASP, while de Haan's adjustment of the window splice method led to the *half splice* method, which links on the last recalculated index of $L = \frac{T}{2}$ periods ago for windows with odd numbers of periods. This method is illustrated in Figure 3.3, which visualises expressions (3.4) and (3.5) that yield published indices.

Note that setting $L = 1$ results in movement splice, which is identical to the method mentioned in Section 3.2.1.1. There is only one index to link on in the previous period, which is always a published index.

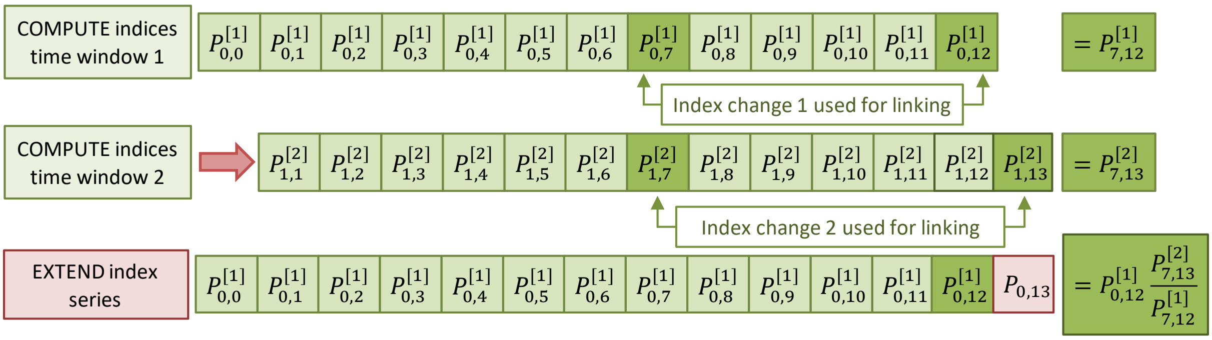


**Figure 3.3:** Illustration of the half splice method, which links on the last recalculated index, for one index extension using 13-period windows. The dark green coloured index of the extended series in previous period 12 is used to link the period to period index that is derived from index changes 1 and 2.

#### 3.2.2.2 Mean splice

Compared with the version of mean splice that links on published indices described in Section 3.2.1.3, it is easier to compute indices when linking on the last recalculated index. In this case it is not necessary to store the indices that would have been published for different linking intervals.

An index to be published according to mean splice when linking on the last recalculated index for each linking interval can be obtained by linking a geometric mean of the period to period indices given by expression (3.4) on the published index of the previous period:

$$P_{0,t} = P_{0,t-1} \left( \prod_{L=1}^{T} \frac{P_{t-L,t}^{[t-T,t]}}{P_{t-L,t-1}^{[t-1-T,t-1]}} \right)^{\frac{1}{T}} \tag{3.6}$$

## 3.3 Adaptive splicing and drift

The extension methods that are described in Section 3.2 use a set of fixed choices over time for the variables that characterise these methods, like the same window length, linking interval and linking index in each time period. The main question to be answered in this study is how the index series produced by different extension methods will behave, in particular with regard to chain drift. In Section 2.2 it was argued that multilateral index series will generally no longer be transitive as soon as index series that are transitive on an initial time window are extended after adapting the time window in order to include data from future periods (see also Table 2.1).

The Multi-Period Identity Test (MPIT), formalised as expression (2.1), can also be applied to multilateral methods. The question then is whether indices between two periods produced by an extension method are equal to the index that is obtained with the same multilateral method by directly comparing the prices and quantities between the same two periods.

For example, the linking of year on year indices on the published indices of 12 months ago with the method HASP (Section 3.2.1.1) results in drift-free year on year indices. However, the MPIT does not hold when linking on recalculated indices. Chaining the period to period indices given by expression (3.4) between periods $t-L$ and $t$ does generally not result in the direct index $P_{t-L,t}^{[t-T,t]}$, since the period to period indices are calculated on different time windows.

The aforementioned examples keep the window length fixed, as is normally done in classical applications of the MPIT. However, the length of the time window is unknown and should be treated as a variable, as was suggested in Section 3.1. Following the original formulation of the MPIT as in expression (2.1) is consequently not trivial; a relaxation of the strict equality in (2.1) makes more sense.

The above considerations motivate a new class of index extension methods, which are referred to as *adaptive splicing* in this paper (Figure 3.1). Window length and linking interval are not fixed in adaptive splicing, but may vary over time. Adaptive splicing methods solve an optimisation problem in order to find values for the window length and linking interval in each period. The indices calculated with mean splice can also be viewed as solutions to an optimisation problem (Section 3.2.1.3), but the length of the time window is fixed and also the set of linking intervals is the same in each period.

A suitable objective function has to be found for adaptive splicing. The objective function underlying mean splice is formulated in terms of the published indices $P_{0,t}$. The level of the index with respect to a fixed reference period is important, as it is used for indexation purposes, but other valid choices can be made as well. A choice that naturally comes to mind is the year on year index, that is, the index with respect to the same period in the previous year. This index is generally accepted as measure of inflation. Also inflation is used for indexation, but it also serves as a very important measure for policy making (e.g. monetary policy, price stability).

The optimisation procedure described below takes months as time units since the CPI is a monthly statistic in almost all countries, but the description can be applied to any choice of time unit. Window length is included in the procedure by comparing the year on year index to be published in a reporting period $t$ with the year on year index on a longer time window. Let $T_t+1$ denote the window length in $t$ for some extension method and let $T' > T_t$. Let $L_t$ denote the linking interval in period $t$ and $P_{t-12,t}^{(L_t)}$ the index in $t$ compared with 12 months ago that would be published using linking interval $L_t$ and window length $T_t+1$.

The idea behind adaptive splicing is to find a value $L_t^*$ for the linking interval that minimises the absolute difference between $P_{t-12,t}^{(L_t)}$ and the index $P_{t-12,t}^{[t-T',t]}$ that results from a longer window. The first index results from an extended index series, while the second index is the result of directly comparing the product prices and quantities in periods $t-12$ and $t$. An extended index is thus compared with a direct index, which is how a relaxed version of the MPIT is incorporated into adaptive splicing in order to control chain drift.

A relaxation of the MPIT is introduced in the following way. The optimal linking interval $L_t^*$ is determined in a stepwise manner by gradually increasing the values of $T'$ and $T_t$. It has to be decided whether a value of $T_t$ at any stage is satisfactory in a certain sense. This could be established by applying a threshold, say $\epsilon$, to the objective function which is added as a constraint to the procedure.

An adaptive splicing procedure can be set up according to the following steps for reporting periods $t$:

1. Choose a window with initial length $T_t + 1$ periods. Also set a maximum window length $W_{\max}$ and a step size $\tau$. Proceed with step 2.

2. Set a length $T' + 1$ for a second window such that $T' = T_t + \tau$. The procedure stops if $T' + 1 > W_{\max}$. Otherwise, proceed with step 3.

3. Determine a linking interval $L_t^*$ that minimises $\left|P_{t-12,t}^{[t-T',t]} - P_{t-12,t}^{(L_t)}\right|$ over $L_t \in \{1, 2, ..., T_t\}$. Proceed with step 4.

4. If $\left|P_{t-12,t}^{[t-T',t]} - P_{t-12,t}^{(L_t^*)}\right| < \epsilon$, then accept $T_t$ and $L_t^*$, in which case $P_{0,t}^{(L_t^*)}$ will be the published index in $t$. Otherwise, set $T_t = T'$ and return to step 2.

Below some remarks are made with regard to this procedure:

- The procedure needs to be initialised by setting values for $T_t$ and $T'$. A commonly accepted minimum window length is 13 months, as was also stated on page 9. This means that $T_t = 12$ could be taken as an initial value in step 1.
- The length of the time window could be increased in different ways, in case 13 months does not satisfy the constraint in step 4. Steps of $\tau = 1$ month could be used, but also more pragmatic choices like $\tau = 6$ or 12 months could be made. These considerations can be used to assign a value to $T'$ in step 2.
- Another question to be answered is how to assign a value to the threshold $\epsilon$ in step 4. Some guidance could be found in EU Regulation 2016/792. Margins are defined at different aggregate COICOP levels, which are used by statistical institutes to indicate whether a change in method has a 'significant impact' on the year on year index in any month compared with the previous method ([20], Article 2 (21)). For instance, the margin at all-items level is set at 0.1 percentage point, while 0.5 percentage point is used at the 4-digit COICOP level. The value of 0.5 was used for $\epsilon$ at the lowest COICOP level in this study (Section 4.3).
- It may happen that the constraint in step 4 is not necessarily satisfied after the second window length $T' + 1$ has reached the maximum length $W_{\max}$. The values found for $T_t$ and $L_t$ are then selected.

The next sections present an extensive comparative study with a summary of the data and methods used, followed by an overview and discussion of the results. Main findings about extension methods derived from the results and discussion are summarised in the final section.

# 4 Comparative study

## 4.1 Data sets

This section gives an overview of the data sets and the index extension methods that were included in a study that aims to compare the performances of different extension methods on key indicators for a wide range of product types. This comparative study has been extended considerably with respect to the two previous studies carried out at Statistics Netherlands ([7], [8]), in terms of data sets and extension methods.

Transaction data sets from 14 Dutch retail chains were selected, which contain sales data of products classified in 27 COICOPs at the 4-digit level (Table 4.1).[2] The selected data sets cover a broad range of products: food, non-alcoholic and alcoholic beverages, garments and footwear, products for maintenance and furnishing of the dwelling, equipment, household products and textiles, medical products, spare parts and accessories for personal transport, plants and flowers, products for pets and personal care products.

| **Store type** | **Nr of chains** | **COICOP (4-digit)** |
|---|---|---|
| Supermarkets | 3 | 0111 - 0119, 0121, 0122, 0212, 0213 |
| Clothing and footwear stores | 2 | 0312, 0321 |
| Department stores | 3 | 0312, 0511, 0520, 0540, 1213 |
| Do-it-yourself stores | 3 | 0431, 0511, 0520, 0551, 0552, 0721 |
| Furniture stores | 1 | 0511 |
| Garden centres | 1 | 0511, 0933, 0934 |
| Pharmacy stores | 1 | 0611, 0612, 1213 |

**Table 4.1:** Types of retail chain and products covered by the selected transaction data sets.

The data sets contain at least 5 years of data, most of which cover the period running from December 2017 until and including December 2022. One additional year of data was available for three retail chains, with December 2016 being the first month in these data sets. All data sets are used in the compilation of the Dutch CPI and cover more than 15 per cent of the CPI in terms of COICOP weights.

Statistics Netherlands makes use of transaction data in its CPI since 2002. Data of longer periods are available for a part of the selected retail chains. Parts of these data are archived, while a smaller part is available in the production environment. The data sets in Table 4.1 are selected from the production database, which only contains data from recent years that are classified according to the current COICOP system.

The time frames covered by the data sets contain some noteworthy events. The year 2022 was characterised by rapidly increasing prices, with the war in Ukraine as a main driver, which led to the highest inflation figures in the Netherlands in almost 50 years' time (10 per cent averaged over 2022). The COVID-19 pandemic also had a big impact on world economies between March 2020 and the first months of 2022, among others because governments imposed different measures like lockdowns. These events may constitute stress tests for index extension methods. The question is how different methods react to rapidly increasing prices and sudden changes in expenditures.

[2]For a description and details of these COICOPs, see `https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=HICP_2000&StrLanguageCode=EN&IntPcKey=&StrLayoutCode=HIERARCHIC`

## 4.2 Index method

As was stated at the beginning of Section 2.2, multilateral methods were originally developed for price comparisons between countries or regions. Multilateral methods were considered as potentially promising methods for price comparisons over time only since the beginning of the previous decade. These methods are now being studied on a large scale as more and more countries are trying to acquire transaction data.

In Section 2.2 reference was also made to the three most widely studied multilateral methods concerning the CPI: the GEKS method, the Geary-Khamis method and the Time Product Dummy method. Index extension methods can be applied in combination with any multilateral method. A multilateral method is used to generate index series on each time window, which can be linked with each of the extension methods described in Section 3.

The multilateral method used in the present study is the Geary-Khamis (GK) method, which was introduced in the Dutch CPI in January 2016 [6]. The GK method is used for processing transaction data, which also includes all data sets described in Section 4.1. The following notation in introduced to describe this method. Let $p_{i,t}$ and $q_{i,t}$ denote the price and quantity of product $i$ belonging to a set $G_t$ in period $t$. As was already mentioned in Section 2.2, the GK method compares the product prices $p_{i,t}$ in any period with some average price calculated over a full time window, say $[0, T]$. The average price of product $i$ is denoted as $\nu_i$ and will be defined below.

The aggregate price change $\tilde{p}_t$ for the set of products $G_t$ in period $t$ with respect to the average product prices $\nu_i$ is defined as

$$\tilde{p}_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t}}{\sum_{i \in G_t} \nu_i q_{i,t}} \tag{4.1}$$

This price change is not necessarily equal to 1 in a reference period. A price index $P_{0,t}$ in period $t$ with respect to a reference period 0 can be obtained by dividing $\tilde{p}_t$ by $\tilde{p}_0$:

$$P_{0,t} = \frac{\tilde{p}_t}{\tilde{p}_0} = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} \nu_i q_{i,t} / \sum_{i \in G_0} \nu_i q_{i,0}} \tag{4.2}$$

The ratio in the numerator of (4.2) measures the change in total expenditure between 0 and $t$. The ratio in the denominator is a quantity index in which products are scaled by their average prices $\nu_i$. Unlike expenditure, product quantities cannot be simply summed when products differ in terms of their quality characteristics. Some scaling or transformation of product quantities should then be applied in order to account for product heterogeneity.

In the GK method, the ratio in the denominator can also be interpreted as a change in 'real expenditure', that is, with prices held constant in some sense over time. Price changes should therefore be eliminated in the $\nu_i$, which is accomplished in the GK method by using the $\tilde{p}_t$ or the price indices $P_{0,t}$ in every period of a time window to deflate product prices. The $\nu_i$ are defined as follows:

$$\nu_i = \frac{\sum_{z=0}^{T} q_{i,z} p_{i,z} / P_{0,z}}{\sum_{z=0}^{T} q_{i,z}} \tag{4.3}$$

Price indices cannot be calculated directly with expressions (4.2) and (4.3), since the price indices also appear in the $\nu_i$. Different index calculation approaches exist for these expressions. A procedure that is easy to implement is an iterative algorithm that starts with a set of arbitrary index values on a time window, which are substituted in (4.3) to yield a set of initial values for the $\nu_i$. These values are substituted in (4.2) which yields updated price indices for each period of the time window. This procedure is repeated until the absolute differences between the last two index series remain below a threshold set by the user [6]. Price indices can also be found as solutions to an eigenvalue problem [18].

It will be clear that the choice of multilateral method influences the results obtained with index extension methods. However, different studies have shown that GEKS, GK and TPD give comparable results when applied to transaction data of different products ([8], [30]). Chessa found that the choice of extension method has a bigger impact than the choice of multilateral method [8].

A specific remark about comparability between multilateral methods is worth making. Various statistical institutes are currently using the GEKS method with Törnqvist indices for bilateral price comparisons. Attention should be paid to clearance prices when applying the GEKS-Törnqvist method, which is sensitive to downward trends when products are sold at highly reduced prices during clearance sales. These risks can be limited by filtering out clearance prices. These comments also apply to GEKS-Fisher [11]. Another possibility is to use Walsh indices, which are robust against clearance prices and do not require any price filters ([8], [11]).

## 4.3 Procedure

### 4.3.1 Extension methods

All extension methods that appear in the tree of Figure 3.1 were included in the comparative study. Window length was restricted to the common choices of 13 and 25 months. All possible linking intervals were included for these window lengths. Also both fixed base methods described in Section 3.2.1.2 were included.

Price indices were also calculated for the adaptive splicing methods shown in Figure 3.1. Different versions of the splicing procedure described in Section 3.3 were run, with initial window lengths of 13 and 25 months, that is, by setting $T_t$ equal to 12 or 24 months in step 1 of the adaptive splicing procedure. The step size $\tau$ that relates this window length to the length $T' + 1$ of the second window was set equal to 12 months. In other words, the year on year indices of the extended series are compared with the direct indices calculated on windows that are 12 months longer in order to determine the optimal linking interval in each time period. The maximum window length $W_{\max}$ was set at 37 months, which means that the largest window size used in the extension is 25 months. The value of the threshold $\epsilon$ in step 4 was set at 0.5 percentage point at the lowest COICOP level for each retail chain.

### 4.3.2 Index compilation and evaluation

The GK method was used to calculate index series on subsequent time windows, which were linked by applying the extension methods included in the comparative study. Index series were calculated at different levels of product aggregation; a summary of the results is given in Section 5. A number of choices and conventions have to be implemented in order to compile index series in the CPI: (1) products have to be defined, and (2) the compilation of index series above the most detailed COICOP level follows a set of strict rules. Finally, the extended index series have to be evaluated in some way. These three aspects will be described in more detail below.

**Product definition**

Price indices in the CPI are compiled at various levels of product aggregation. Indices are published for each aggregate in the international COICOP classification. The most important aggregate is the overall or all-items aggregate, as this headline level yields the 12-month rates of change that are presented as "inflation". The next 2-digit level contains 12 "divisions". Each division is subdivided into more detailed aggregates down to the 5-digit level. For example, division 01 at the 2-digit level stands for Food and non-alcoholic beverages. Below this level we find, for instance: 011 Food, 0111 Bread and cereals and 01111 Rice. Statistical institutes are free to add more detailed product aggregates below the 5-digit level.

An important decision to be made is how to define individual "products" within the most detailed aggregates. The most detailed product level in transaction data sets is the bar code or GTIN level. As was stated in the introduction, bar codes may change after item re-introductions or relaunches. If prices of these re-introduced items are increased while quality remains the same, which is a common phenomenon, then the GTINs of such items have to be combined in order to capture the price increases. Relaunches typically occur with fashion goods like garments and footwear (division 03), and also with products for personal care (division 12).

GTINs are usually not suited as unique products for the product types mentioned, so that broader product definitions are needed in which different GTINs are combined. This clustering or stratification of GTINs can be achieved by identifying sets of common characteristics ([6], [9], [36]).

The product definitions used in the Dutch CPI were also chosen in this study. GTINs are used as unique products for the majority of the COICOPs in Table 4.1. Broader product definitions are used only for garments and footwear (0312, 0321), furniture and furnishings (0511), household textiles (0520), pharmaceutical and other medical products (0611, 0612) and personal care products (1213).

**Index compilation**

The index compilation process proceeds along two main stages. At the first stage, multilateral index series were computed at the most detailed COICOP level for each retail chain. This was done for every extension method included in this study (Section 4.3.1). At the second stage, the index series were aggregated to higher-level indices with the "Laspeyres-type" method described in Chapter 8 of the HICP Methodological Manual [22]. Index series for each COICOP were first derived by aggregating over retail chains. Next, the resulting index series were aggregated to higher-level COICOPs up to all-items level with the same procedure. The Laspeyres-type method makes use of fixed annual weights. The weights used in the Dutch CPI were also used in this study.

**Evaluation**

The index series obtained with each extension method were compared on two main indicators: (1) the 12-month rates of change or year on year indices, and (2) the yearly average index levels in 2020-2022, which are the years in which the extension methods take effect when using 25-month windows and five years of data.

The question is how the results for these two indicators could be assessed in a meaningful way. The usual approach followed in different studies is to compare extended index series with the transitive index series on the full data period. Although this sounds logical, this benchmark choice also has drawbacks. The full period index also contains the influence of future transactions on the indices in past periods.

A more realistic choice would be to compute index series based on all data up to the current period. This choice was made in this study. The year on year indices derived from these index series were used as benchmark, both for the 12-month rates of change and for the yearly average index levels. The yearly average indices were computed as a chained index of the geometric averages of the monthly year on year indices in 2020, 2021 and 2022, with 2019 as the reference year.

The margins given in Article 2 (21) of EU Regulation 2016/792 were partly used in order to establish whether index series are affected by drift [20]. A deviation in a year on year index in any month larger than 0.1 percentage point from the benchmark would be considered as drift at all-items level. The margins defined at lower levels of aggregation in the regulation seem more arbitrary. These margins are more difficult to define because of differences in the sizes of aggregates at the same level of aggregation. A more careful position was therefore taken in this study with regard to drift at lower levels of aggregation. The differences of extension methods with respect to the benchmark were compared instead of judging whether drift occurs below all-items level.

# 5 Results

This section gives a summary of the results of the comparative study, which are organised according to three levels of product aggregation: the all-items level, the 2-digit COICOP level and the 4-digit level. The results of all static splicing methods are shown in this section, with the exception of the FBEW method. The results for the FBRW method are comparable, which are the ones included in the overview. Results for adaptive splicing are only shown for methods that link on the last recalculated indices, since differences with the versions that link on published indices are small.

Index differences with respect to the benchmark index are expressed as percentage points (pp), with the benchmark index subtracted from the extended indices. A positive (negative) difference thus means that an extension method gives a higher (lower) index than the benchmark. Index values are scaled by 100, which is the value of an index series in the initial month (December 2017).

## 5.1 All-items level

First, an impression is given of the price development aggregated over all retail chains and COICOPs from December 2017 until December 2022. This is shown in Figure 5.1 for three extension methods: $S(13, 6, r)$, $S(25, 6, r)$ and the version of adaptive spicing on recalculated indices $S(25, L^*, r)$ that uses 25-month windows. The two methods $S(13, 6, r)$ and $S(25, 6, r)$ link on the last recalculated indices of 6 months ago. These two methods give the most accurate index levels in the final year 2022 for 13-month and 25-month windows, as will be shown afterwards. Figure 5.1 clearly shows the effects of the rapid increase in prices in 2022.

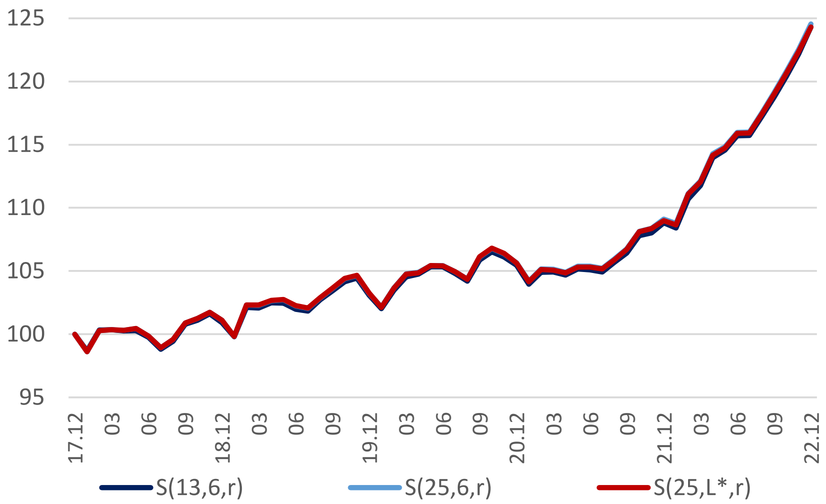


**Figure 5.1:** Index series at all-items level for three extension methods.

The differences between the three index series are very small, which is worth emphasising as one of the methods uses 13-month windows. Extension methods that use these shorter windows are apparently capable of producing accurate indices over periods of at least 5 years. This can also be noted in Figure 5.2, which shows the index differences from the benchmark in 2020-2022 for different extension methods including all linking intervals for the static methods. The results for linking intervals around 6 months when linking on recalculated indices are very accurate.

The yearly average index differences for adaptive splicing in the right-hand graph are

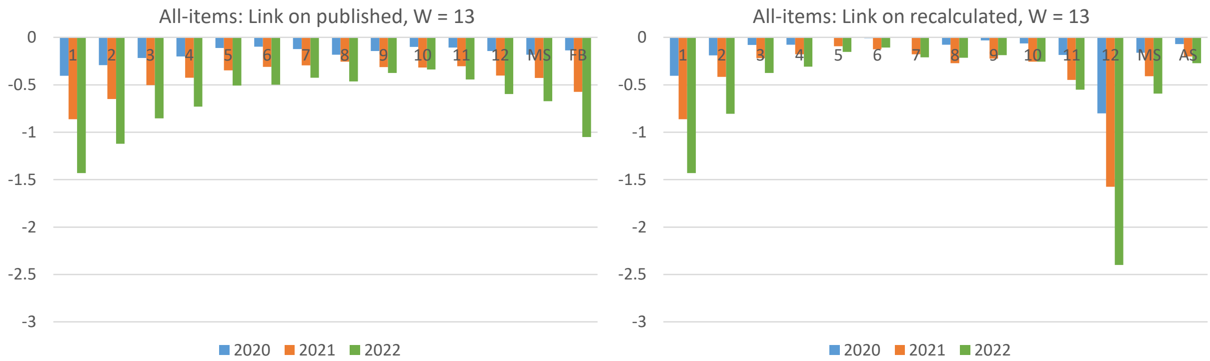


**Figure 5.2:** Yearly average index differences (pp) for extension methods with 13-month windows with respect to the benchmark. Horizontal axes show linking intervals, MS = mean splice, AS = adaptive splicing, FB = FBRW.

somewhat larger than for methods that link on the last recalculated indices of 5, 6 and 7 months ago. Although adaptive splicing finds the linking intervals that minimise the absolute differences with the year on year indices calculated on longer windows in each month, this method does not necessarily give smaller differences when aggregating indices over retail chains, COICOPs and, in the case of Figure 5.2, also over time. But in most cases adaptive splicing does yield the most accurate results, as will become clear from the other results that will be presented in this section.

Several other interesting features emerge from Figure 5.2:

- Linking on recalculated indices is more sensitive to the choice of linking interval than linking on published indices.
- Linking months near the edges of a time window produce the largest differences, in particular when linking on recalculated indices.
- Linking on the published index of 12 months ago produces much better results than linking on the last recalculated index of a year ago, which results in large deviations from the benchmark.
- Optimal linking intervals can be identified in both graphs. Methods that link on recalculated indices have a preference for shorter linking intervals compared with methods that link on published indices.
- The adverse effects of choosing linking months close to the edges of a time window also propagate to mean splice.
- The FBRW method shows rather poor performance, which suggests that December is generally not a suitable linking month.

The high sensitivity of linking on recalculated indices, in particular on 12 months ago for 13-month windows, was already highlighted in previous studies by the same author ([7], [8]). Arguments were also given for the downward drifting behaviour of this window splice method, which can also be noted in the right-hand graph of Figure 5.2.

Highly reduced prices during clearance sales constitute a potential source of drift. When using the GK method, the average deflated prices $\nu_i$ defined by expression (4.3) will increasingly depend on clearance prices as the window is shifted, while the $\nu_i$ should reflect regular prices in order to produce reliable indices. This is not a drawback of the GK method but is an implication of using a time interval smaller than the data period, which leads to a statistical phenomenon known as "censoring". In situations where exiting products still generate substantial expenditure in the first month of a window but quickly drops in subsequent months as prices decrease, the denominator $\tilde{p}_0$ of index formula (4.2), which corresponds with the linking month, will distort an index series and result in drift.

Linking on the published indices of 12 months ago produces much better results because the year on year indices that are computed on each rolling window will be preserved in the extended series, which is not the case when linking on recalculated indices. Deviations from the benchmark are therefore smaller in general when linking on published indices with long linking intervals.

On the other hand, Figure 5.1 also shows that certain choices for the linking interval produce very good results when linking on recalculated indices. For certain choices of the linking interval the index levels are more accurate than for methods that link on published indices. The index levels for a linking interval of 6 months deviate at most only 0.1 pp from the benchmark. This result is one of the main new findings compared with previous contributions on this subject.

Figure 5.3 shows the results for 25-month windows. Also in this case, linking on published indices reveals a preference for longer linking intervals. The results are more accurate for linking intervals longer than 11 months compared with linking on recalculated indices.

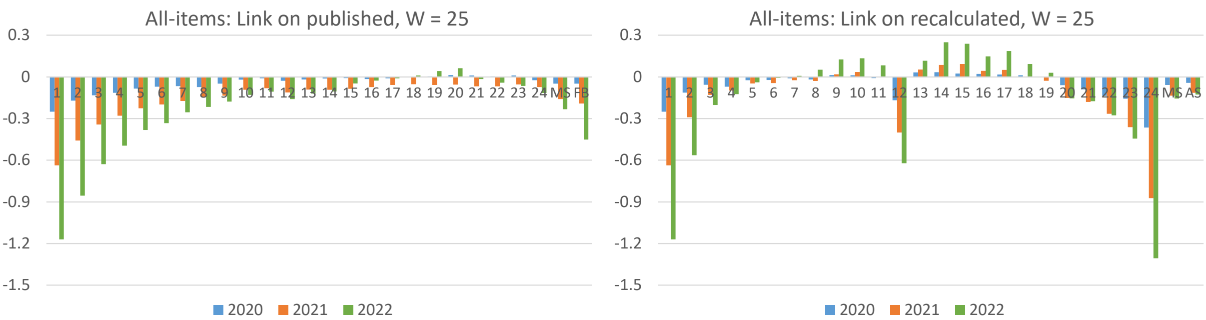


**Figure 5.3:** Yearly average index differences (pp) for extension methods with 25-month windows with respect to the benchmark. Horizontal axes show linking intervals, MS = mean splice, AS = adaptive splicing, FB = FBRW.

The results for methods that link on recalculated indices show a different pattern compared with the right-hand graph in Figure 5.2. Like for 13-month windows, the shortest and longest linking intervals show the poorest performance. A preference for shorter linking intervals also emerges for 25-month windows, with linking intervals around 6 months producing the best results again.

But the right-hand graph of Figure 5.3 also reveals some peculiar features. One of the most striking results is the fairly large downward deviation when linking on the last recalculated index of 12 months ago. A second interesting result is the accurate performance of linking intervals between 13 and 21-22 months. The improved results obtained with mean splice over the versions with 13-month windows are also worth noting, with mean splice on recalculated indices giving slightly more accurate results than the version that links on published indices (left graph). A more detailed look into the above results will be provided in next subsection, where the index differences at the 2-digit COICOP level will be presented.

This subsection closes with a summary of the year on year indices for a selection of static splicing methods. Figure 5.4 shows the differences for the 12-month rates of change with respect to the benchmark in each month of the last three years. The data of the first two years are used by methods with 25-month windows to compute an index series on the initial time window and are excluded in the presentation of the results since no extension is still carried out. The rightmost graph confirms the high accuracy of the method that links on the recalculated indices of 6 months ago. Of all methods shown in this figure, it is the only one that is practically free of drift. It is also worth noting that there is no trend in the differences for this method.

The adaptive splicing method, which is not shown, is slightly less accurate than $S(13, 6, r)$ and produces some drift. This result is in agreement with the differences between these two methods shown in the rightmost graph of Figure 5.2.

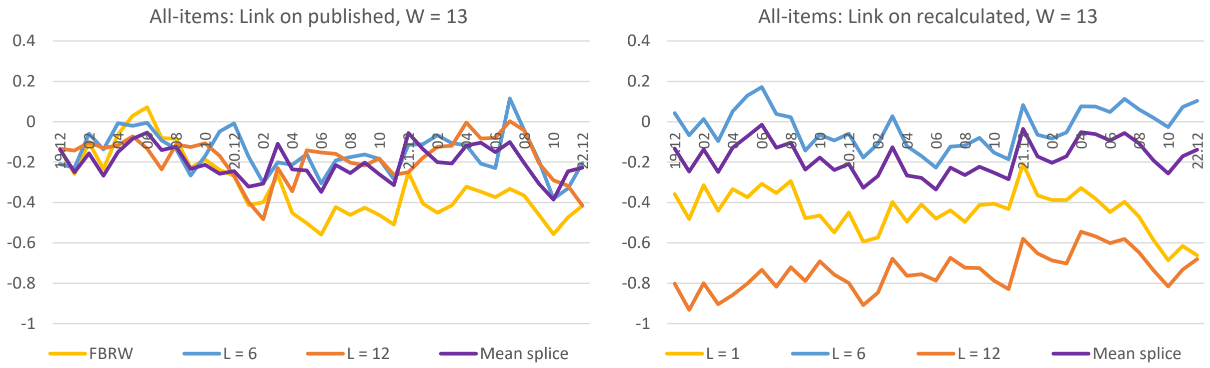


**Figure 5.4:** Year on year index differences (pp) for extension methods with 13-month windows with respect to the benchmark.

Figure 5.5 shows the year on year differences for a selection of methods with 25-month windows. The differences are smaller among the methods that link on published indices, which is in line with the results in Figure 5.3. The absolute differences for linking on recalculated values of 6 months ago and for mean splice hardly exceed 0.1 pp and perform better than most methods that link on published indices. Both methods can be said to be free of drift, which also holds for adaptive splicing (not shown).

## 5.2 COICOP-2 level

Accurate results at all-items level may be the result of well-behaving index series for lower product aggregates, but also of positive and negative differences that cancel each other. It is therefore important to quantify index differences also for more detailed aggregates. This was done in this study at different levels of aggregation.

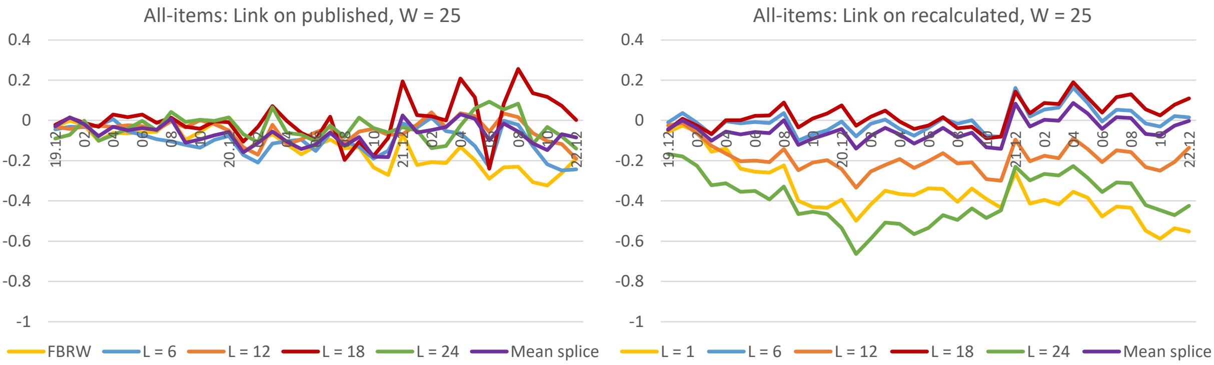


**Figure 5.5:** Year on year index differences (pp) for extension methods with 25-month windows with respect to the benchmark.

The 2-digit COICOP level contains a number of important aggregates with regard to transaction data. Statistical institutes typically start with supermarkets in their attempts of acquiring transaction data. Most products sold by supermarkets belong to COICOP division 01 Food and non-alcoholic beverages. This division has a high weight in the CPI of every country and has generated a lot of media exposure since food prices went up dramatically in 2022.

The results at 2-digit level are shown for the three divisions with the highest weight in this study: COICOP 01, COICOP 03 Clothing and footwear, and COICOP 05 Furnishings, household equipment and routine household maintenance. COICOP 0431, Materials for the maintenance and repair of the dwelling, was combined with COICOP 05, which is referred to as COICOP 04+05 in the results.

The overview starts again with the yearly average index differences for methods with 13-month and 25-month windows, which are shown in figures 5.6 and 5.7 respectively. The poor results for short and long linking intervals (for recalculated indices) at all-items level also emerge at the 2-digit level. Apart from these linking intervals, linking either on published or recalculated indices gives good results for COICOP 01 when using 13-month windows (Figure 5.6).

Quite interestingly, linking on recalculated indices also performs well on COICOP 03 when moving away from the edges of a time window. Overall it can be said that linking on the recalculated indices of around 6 months ago produces accurate indices, which perform better than linking on published indices. An exception is COICOP 04+05; the results in Figure 5.6 show that the index series suffer from downward drift. This also holds for linking on published indices.

The results for COICOP 04+05 show that 13-month windows are generally too short for these types of product. This holds for static splicing. It is interesting to see that index differences with respect to the benchmark are substantially reduced with the adaptive splicing method. This method gives very accurate results for COICOP 01 and can be said to perform quite well also for the other two COICOPs.

The results for 25-month windows are shown in Figure 5.7. There is hardly any improvement for COICOP 01 compared with the results for 13-month windows in Figure 5.6. The results for linking intervals of at least 12 months are even worse, in particular

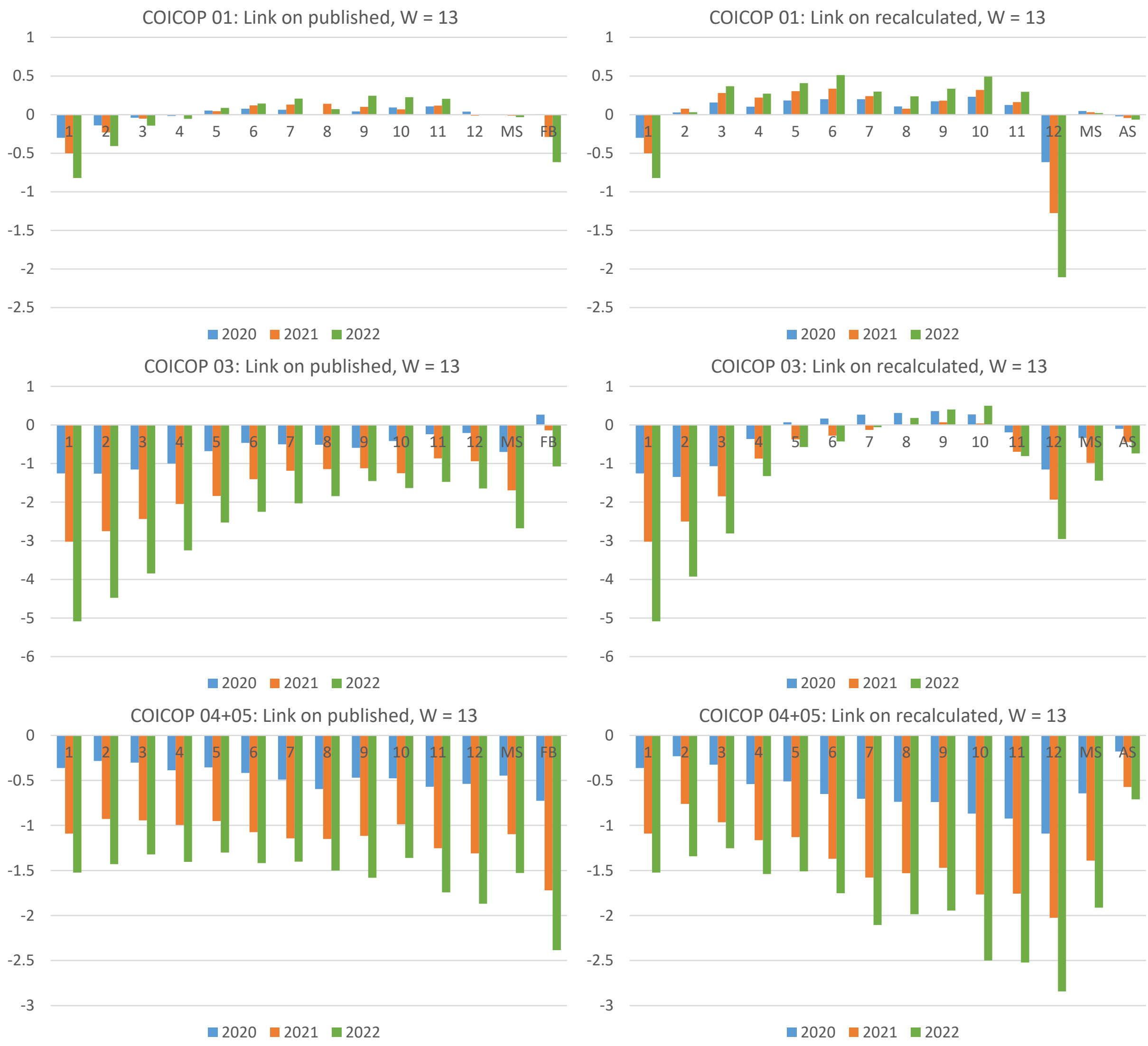


**Figure 5.6:** Yearly average index differences (pp) for extension methods with 13-month windows with respect to the benchmark for three aggregates at 2-digit level. Horizontal axes show linking intervals, MS = mean splice, AS = adaptive splicing, FB = FBRW.

when linking on recalculated indices. The results for COICOP 03 improve when linking on published indices, while the results were already accurate for 13-month windows when linking on the recalculated indices of around 6 months ago. Substantial gains in accuracy with a 25-month window are obtained for COICOP 04+05. However, long linking intervals should be avoided also in this case when linking on recalculated indices.

Returning to the results for 25-month windows at all-items level in Figure 5.3, it was noted that accurate indices are obtained with linking intervals between 13 and 21-22 months. The right-hand graphs of Figure 5.7 show fairly large positive differences for COICOP 01 and large negative differences for COICOP 04+05, which almost cancel out at all-items level.

On the other hand, the shorter linking intervals around six months not only produce accurate indices at all-items level, but also at 2-digit level. These differences in index behaviour between the shorter and longer linking intervals will be a topic of discussion


**Figure 5.7:** Yearly average index differences (pp) for extension methods with 25-month windows with respect to the benchmark for three aggregates at 2-digit level. Horizontal axes show linking intervals, MS = mean splice, AS = adaptive splicing, FB = FBRW.

in Section 6.1. The results for adaptive splicing are very accurate for each of the 2-digit COICOPs.

Also this subsection closes with the year on year differences measured with respect to the benchmark for 13-month and 25-month windows. The results are shown in figures 5.8 and 5.9. The extension methods with 13-month windows shown at all-items level are also included in Figure 5.8.

A number of extension methods with 25-month windows shown in Figure 5.5 are left out in Figure 5.9 in order to have a clearer overview of the differences among the remaining methods. Three methods were retained, which have the same linking intervals for linking on published and recalculated indices: linking on 6 months ago is one of the best methods when linking on recalculated indices, linking on 12 months ago is one of the best choices when using published indices, and the third method is mean splice.

Generally speaking, the year on year differences for the methods that make use of 25-month windows are smaller than for the methods with 13-month windows. The observations that were made based on the index differences for 13-month and 25-month windows are confirmed by the results for the year on year differences. Some methods with

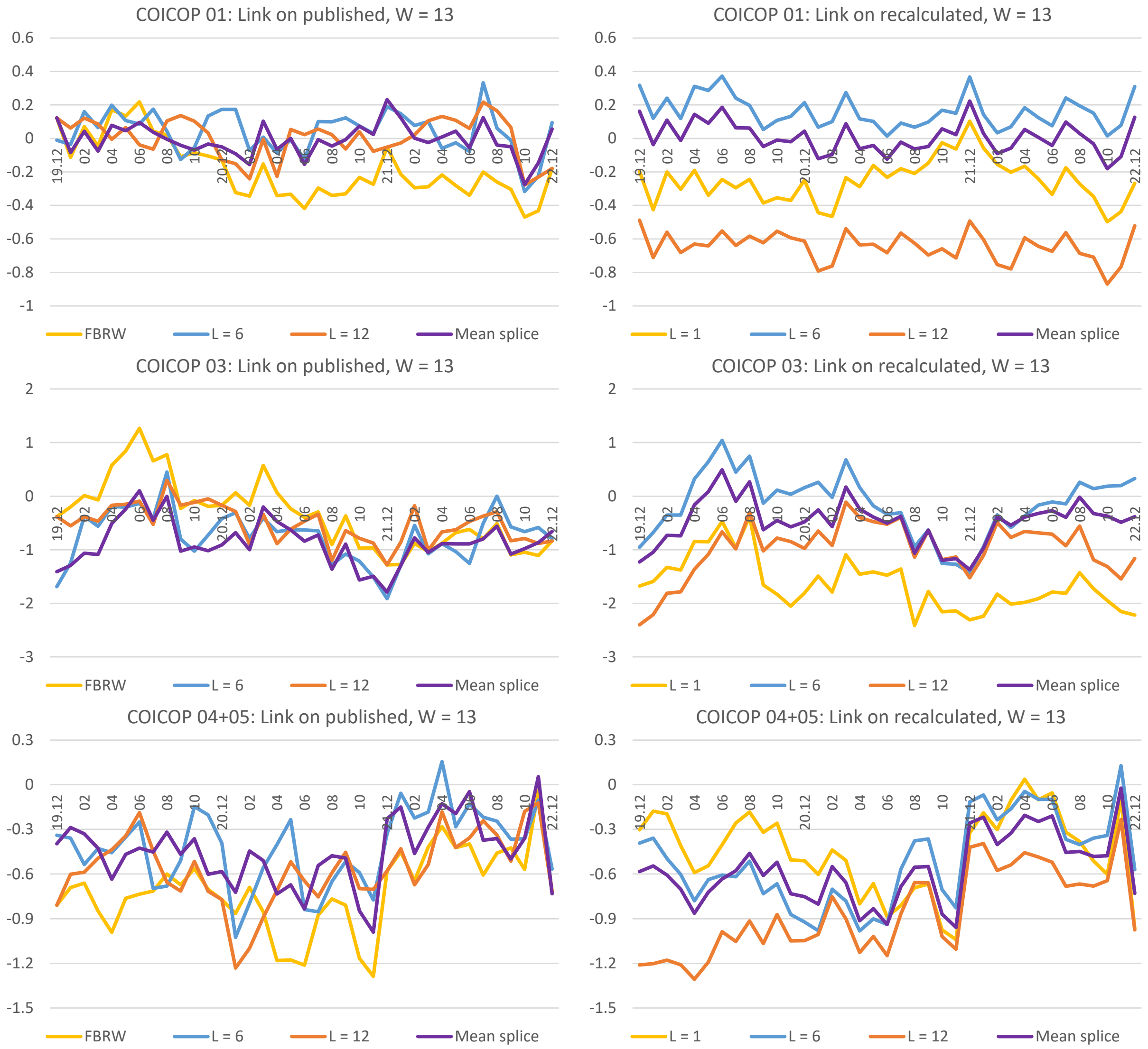


**Figure 5.8:** Year on year index differences (pp) for extension methods with 13-month windows with respect to the benchmark for the three aggregates at 2-digit level.

13-month windows perform well on COICOP 01 and to some degree also on COICOP 03. Larger differences result for COICOP 04+05, which are reduced with 25-month windows.

It is interesting to note the differences between mean splice on recalculated indices and the method that links on the last recalculated indices of 6 months ago for COICOP 04+05 (lower right graph in Figure 5.9). The poor results for long linking intervals that were shown in Figure 5.7 have a big influence on mean splice. The best results for this product aggregate are obtained with adaptive splicing (not shown), which also gave the most accurate yearly index levels (Figure 5.7).

## 5.3 COICOP-4 level

The results presented in the two previous subsections contain a lot of information about index behaviour as a function of different extension variables. The summary overview makes clear that it is extremely challenging to identify choices that produce accurate

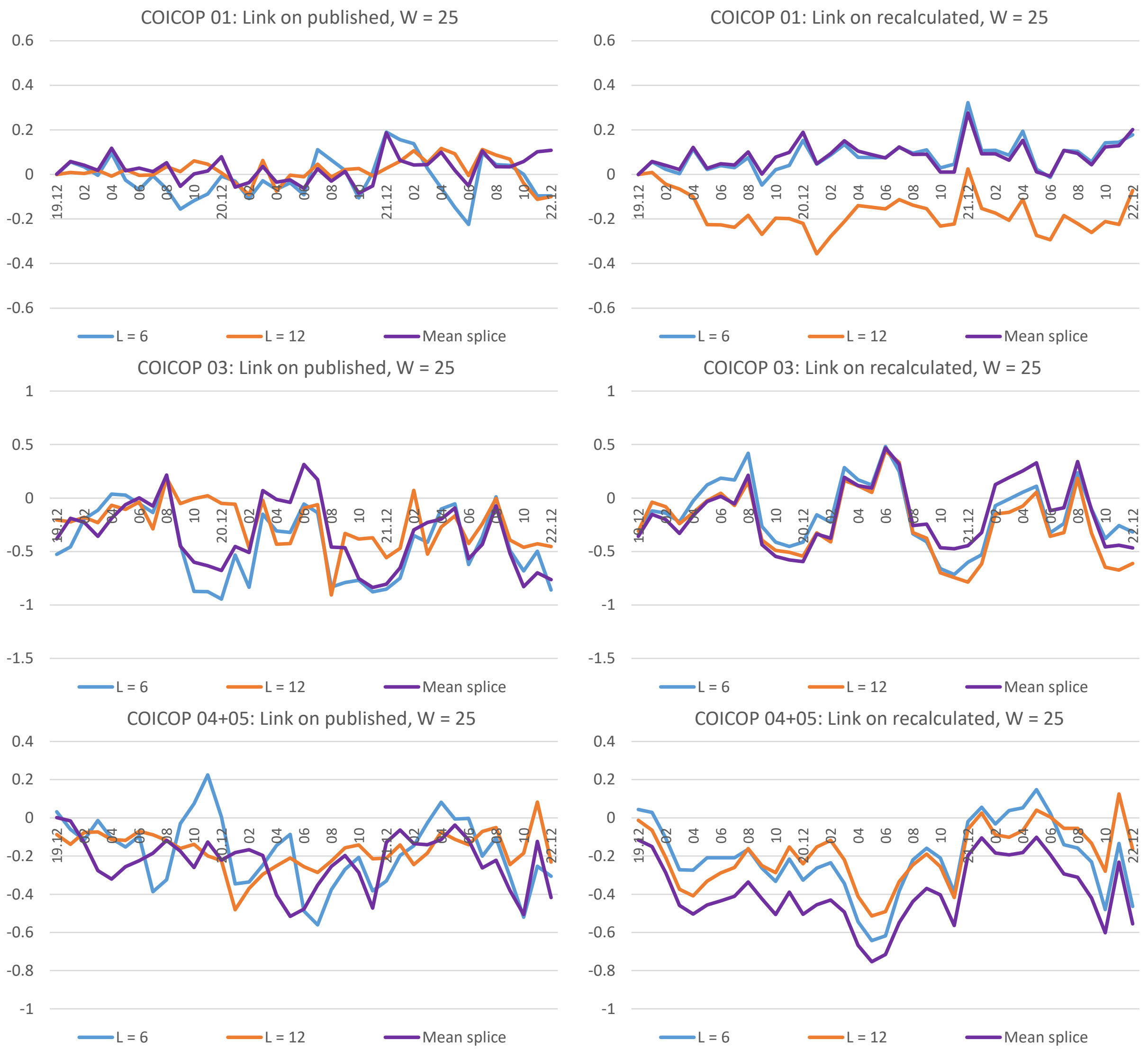


**Figure 5.9:** Year on year index differences (pp) for extension methods with 25-month windows with respect to the benchmark for the three aggregates at 2-digit level.

indices across a broad range of product types. An extension method may perform very well on one aggregate, but poorly on others. The results have nevertheless provided insight into promising methods, which will be explored further in this subsection and in Section 6.

This subsection presents the results at 4-digit level for a small selection of index extension methods since 27 COICOPs were included at this level of aggregation (Table 4.1). Results for each COICOP are shown for a selection of promising index extension methods.

The index differences at all-items level and at 2-digit level have shown that linking on the recalculated indices between about 5 and 8 months ago produces the most accurate indices for this class of static splicing methods, which can be said to hold for both 13-month and 25-month windows. Given the marginal differences between these choices, linking on the index of 6 months ago is a representative choice.

Figures 5.3 and 5.7 show that linking intervals between 9 and 14 months perform consistently well when linking on published indices with 25-month windows. A linking interval of 12 months can be considered representative for this range. The selection of methods is complemented with mean splice for both 13-month and 25-month windows and for linking on published and recalculated indices, and with adaptive spicing for both (initial) window sizes.

Differences with respect to the benchmark indices are shown again for year on year indices and yearly average index levels. In order to present the results in a concise way with such a large number of COICOPs, the median of the absolute differences in the year on year indices over the last three years and the differences between the average indices in the final year 2022 are shown.

Figure 5.10 shows the index differences in 2022 for the selected extension methods. The two methods with 25-month windows that link on published indices give very accurate results at COICOP-4 level, with $S(25, 12, p)$ performing slightly better than mean splice. The method $S(25, 6, r)$ generally performs better than mean splice on recalculated indices, in particular for the 05 aggregates, which is in agreement with the results at 2-digit level in the lower right graphs of figures 5.7 and 5.9. The method gives accurate results for all COICOPs, with larger differences only for COICOPs 0540 and 0933.

The method $S(13, 6, r)$ with a 13-month window also performs well at the 4-digit level for most COICOPs. The method shows fairly large deviations from the benchmark indices for two COICOPs in division 01, for most COICOPs in the 05 division and for 0721.

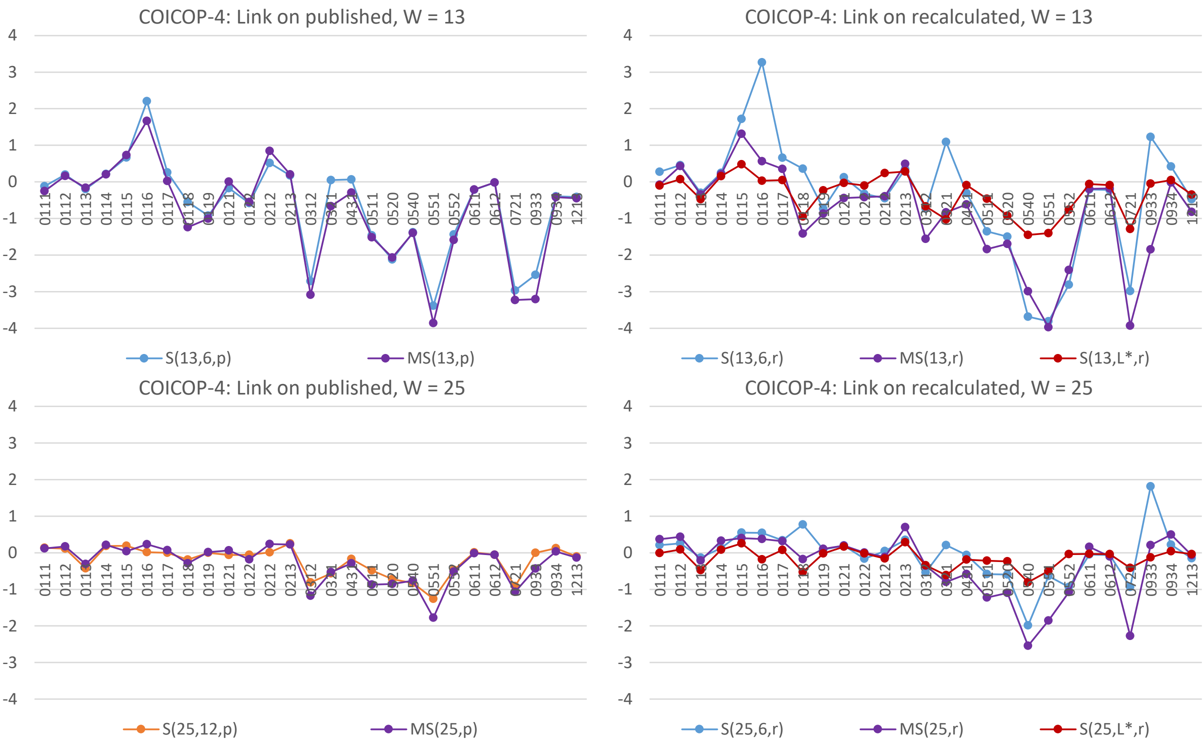


**Figure 5.10:** Average index differences (pp) in 2022 for several extension methods with respect to the benchmark for the 27 COICOPs at 4-digit level (horizontal axis).

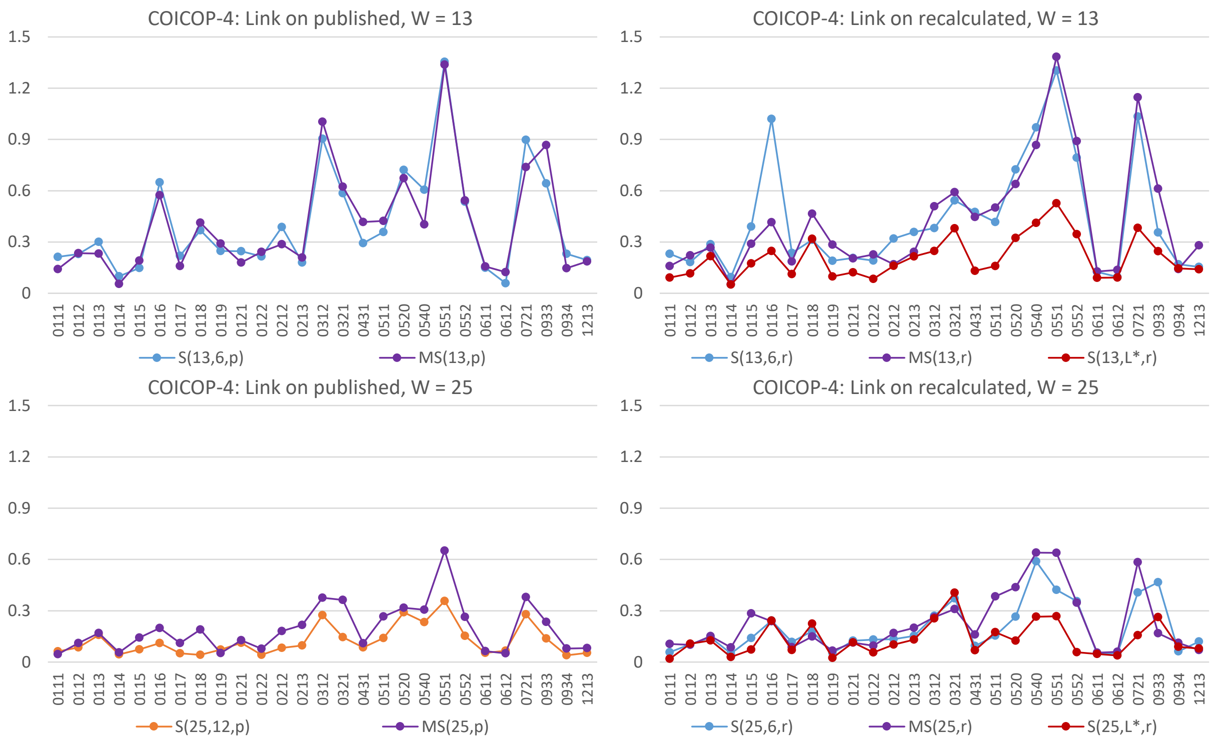


**Figure 5.11:** Median values of the absolute differences (pp) between the year on year indices of several extension methods and the benchmark in the last three years for the 27 COICOPs at 4-digit level (horizontal axes).

The two adaptive splicing methods produce very accurate results. The 25-month version has a very high accuracy at the 4-digit level, but the 13-month version comes very close, showing comparable deviations, except for some of the 05 aggregates and 0721 (Spare parts and accessories for personal transport equipment). It is worth emphasising that index series were extended with 13-month windows in almost 75 per cent of the cases (retail chain $\times$ COICOP $\times$ month) with this adaptive splicing version. This can be considered a high percentage for a tightly set threshold $\epsilon = 0.5$ percentage point applied at the lowest COICOP level for each retail chain.

The indices for $S(25, 6, r)$ show small deviations compared with the 25-month adaptive splicing method, with notable differences only for aggregates 0540 (Glassware, tableware and household utensils) and 0933 (Gardens, plants and flowers), and to a smaller extent for 0552 (Small tools and miscellaneous accessories). A similar picture emerges from the year on year differences in Figure 5.11.

The differences between the results for methods that make use of different window sizes can partly be explained from the product dynamics of the different types of product. Figures 5.10 and 5.11 show that the differences are larger for a number of COICOPs, and also that larger deviations remain for certain COICOPs when using 25-month windows.

One of the components in the method MARS for product stratification is the so-called degree of *product match*, which is defined as the share of the number of products sold in a period $t$ that were also sold in a reference period, expressed as a proportion between 0 and 1 of the total number of products sold in period $t$ [9]. This notion is closely related to the complement of product churn.

Figure 5.12 shows the minimum and maximum values of product match for each of the COICOPs at 4-digit level. The reference period is chosen 12 months before the period in which product match is calculated. As can be noted, product match takes very low values for certain COICOPs, which includes most aggregates from division 05. Large differences were found for these aggregates between methods that make use of different window sizes. In these cases, it is very important to use longer windows in order to avoid or mitigate drift as a result of high price reductions for disappearing products. As can be seen in Figure 5.10, the drift has almost always a downward nature.

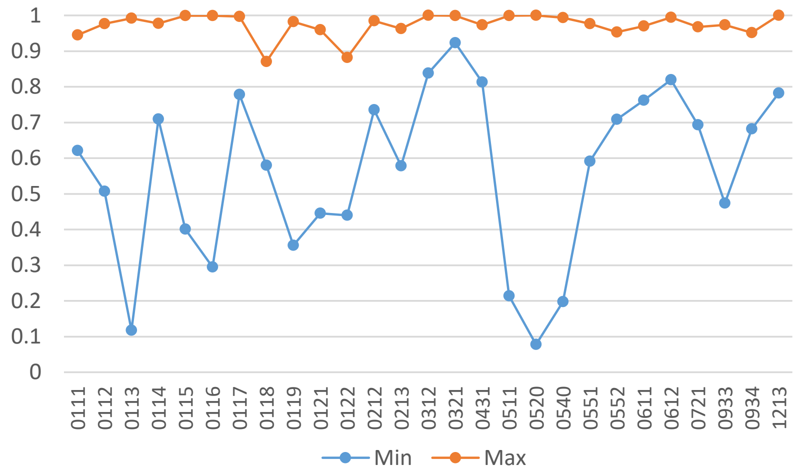


**Figure 5.12:** Minimum and maximum values of the degree of product match for the 27 COICOPs at 4-digit level (horizontal axis).

Very low values for product match are also found for food products, such as 0113 (Fish and seafood), 0115 (Oils and fats) and 0116 (Fruit). Figures 5.10 and 5.11 also show differences between methods using different windows sizes for aggregates 0115 and 0116. Some methods with 13-month windows show upward drift in these cases. The results presented in this section will be discussed further in Section 6.1.

# 6 Discussion

## 6.1 Analysis

The broad comparative study has shown that various extension methods exist that are capable of controlling drift in time series of inflation. This can be achieved in different ways, since the results have shown that different appropriate choices for the extension variables presented in Section 3.1 can be made:

- Adaptive splicing provides control over the year on year indices by finding the linking interval that minimises the absolute difference with respect to a longer time window in each period.

- Linking on published indices also provides direct control over year on year indices in published index series, in particular when linking year on year indices on the published indices of one year ago.
- The results have also shown that accurate results can be obtained with methods that link on the last recalculated indices by using a fixed time window and linking interval, provided that the linking interval is shorter than 12 months and not close to the reporting period.

The accuracy in the year on year indices obtained with adaptive splicing and with methods that link year on year indices on published indices is understandable, but the performance of methods that link on recalculated indices using a fixed window length and linking interval is more difficult to understand. The question is why differences between methods with different linking intervals can be large and why certain choices for the linking interval give accurate results. This subsection provides arguments for these observations.

In order to identify possible causes for the different results of extension methods that link on recalculated indices, the key question is which components of the index formulas differ between methods that use different linking intervals. An index series that is extended by linking on the last recalculated index can be written as a monthly chained index series according to expression (3.5) on page 17 with month on month price index given by (3.4). An explicit expression for this month on month index can be derived from price index formula (4.2) of the GK method.

The numerator of this formula measures the change in expenditure between two periods. It can be left out from the analysis since it is transitive and therefore free of chain drift. The interesting part of (4.2) is the denominator, which is a quantity index. This index contains the average product prices $\nu_i$, which will normally change when shifting a time window. As a consequence, drift in an extended index series is caused by these changes in the $\nu_i$ in combination with the changes in the product quantities.

Let $Q_{t-1,t}$ denote the published month on month quantity index in period $t$ with respect to $t-1$. The shorthand notation $\nu_i^t$ is used for $\nu_i$ calculated on a time window $[t-T, t]$. The month on month quantity index that corresponds with price index expression (3.4) can be written as follows for the GK method when linking on the last recalculated index in period $t-L$ with linking interval $L$:

$$Q_{t-1,t} = \frac{\sum_{i \in G_t} \nu_i^t q_{i,t} / \sum_{i \in G_{t-L}} \nu_i^t q_{i,t-L}}{\sum_{i \in G_{t-1}} \nu_i^{t-1} q_{i,t-1} / \sum_{i \in G_{t-L}} \nu_i^{t-1} q_{i,t-L}} \tag{6.1}$$

Expressions $\sum_{i \in G_t} \nu_i^t q_{i,t}$ in the numerator and $\sum_{i \in G_{t-1}} \nu_i^{t-1} q_{i,t-1}$ in the ratio of the denominator are the same for every linking interval. So are the $\nu_i^t$ and $\nu_i^{t-1}$ computed on the last two time windows. The differences between extension methods are fully determined by the sets of products $G_{t-L}$ when choosing a specific linking interval $L$.

As was already mentioned in the discussion of the results at all-items level in Section 5.1, window splice is sensitive to downward index deviations when exiting items are sold at clearance prices. The linking period in window splice is the first period of a time window.

Subsequent windows will be increasingly dominated by clearance prices as a window is shifted. This is a form of window or interval censoring, which leads to decreasing values $\nu_i^t$ as a function of $t$ for products that are about to be removed from the stores. This implies that the sum $\sum_{i\in G_{t-L}} \nu_i^t q_{i,t-L}$ in (6.1) tends to be smaller than the term $\sum_{i\in G_{t-L}} \nu_i^{t-1} q_{i,t-L}$ in the ratio of the denominator. This has an upward effect on the quantity index and hence a downward effect on the price index.

The downward effects are felt most for linking periods close to or equal to the first period of a time window. The choice of linking period determines which products enter a price index. Moving the linking period towards the current period will select products of which the $\nu_i$ will increasingly depend on regular prices when items leave the stores after the linking period. The quantity and price indices will then be less sensitive to clearance prices, which will limit downward index deviations. These effects are clearly visible in different graphs shown in Section 5.

On the other hand, the linking period should also not be chosen close or equal to the final period of a time window. Setting $L = 1$ yields the value 1 for the denominator in (6.1), which simplifies to the following expression:

$$Q_{t-1,t} = \frac{\sum_{i\in G_t} \nu_i^t q_{i,t}}{\sum_{i\in G_{t-1}} \nu_i^t q_{i,t-1}} \tag{6.2}$$

This is the expression for the quantity index of the method known as movement splice (Section 3.2.1.1, page 14). Expression (6.2) leads to a classical case of period to period chaining, which is known to be sensitive to chain drift.

The above arguments provide explanations for drifting behaviour when choosing linking periods close to the edges of a time window and offer support for the results presented in Section 5 for different product aggregates and for division 05 in particular. This aggregate is characterised by high rates of product churn and clearance prices.

An interesting question is why linking intervals around $L = 6$ months consistently show accurate results across different product types and levels of aggregation for methods that link on recalculated indices. The year on year index can be written as a chained index of 12 month on month indices, with quantity indices given by expression (6.1). The monthly chained index has the following 12 linking months: $t-11-L, t-10-L, ..., t-L$. For $L = 6$ or 7, this set of linking months is centred around the reference month $t-12$ of the 12-month rate of change.

The accurate results obtained for $L = 6$ and 7 months are therefore not surprising. These linking intervals are apparently natural choices that lead to accurate year on year indices when linking on the last recalculated indices. The transitivity property of the index series calculated on each time window allows to control for drift in the year on year indices. Given the choice of linking interval, the question that remains is how many periods should be included before the linking period. This number of periods is important for controlling previously described censoring effects caused by the edges of a time window. Given the accurate results obtained with the method $S(25, 6, r)$, choosing 18 months before the linking month seems to be sufficient.

The above considerations offer guidance for choosing a suitable extension method. The centrality of the set of 12 linking months around the reference month in year on year indices also explains why shorter linking intervals are preferred when linking on the last recalculated index compared with methods that link on published indices. A linking interval of 12 months is a natural choice for methods that directly link year on year indices on published indices.

## 6.2 Short-term price changes

Year on year indices and index levels with respect to some reference period are the two most important indicators for policymakers and other stakeholders of the CPI. The rapid price changes that have characterised the food and energy market in recent times may also shift the attention to price changes on a shorter term, such as month on month. It is therefore important to care about different aspects of an index series than to focus only on the aforementioned two indicators.

This subsection is dedicated to a short investigation of the differences between the month on month indices of different extension methods. An important extension variable in this respect is the linking index. Linking on recalculated or published indices may have different effects on period to period indices. This can be understood by comparing the linking in figures 3.2 and 3.3.

Both figures show how an index series can be extended to the next period for the half splice method on published and recalculated indices. Linking on the last recalculated index (Figure 3.3) produces a period to period index that only depends on the index series of the last two time windows. The index changes that are derived from the last two index series share the same linking period. These properties appear in formalised form in expression (6.1). The resulting period to period indices can therefore be expected to be timely and accurate.

The period to period indices that follow from index series that are extended by linking on published indices have different properties. The period to period indices that follow from the index series of the last two time windows do not have a common linking period, but differ by one period when shifting a time window. This implies that the period to period indices will depend on the period to period indices from the initial time window.

Year on year and month on month index differences with respect to benchmark indices are compared in Figure 6.1 for extension methods $S(25, 6, r)$ and $S(25, 12, p)$. Two product aggregates are included in the comparison: Alcoholic beverages in COICOP 02 and 054020 Cutlery, flatware and silverware of a specific retail chain.

The year on year indices of the two extension methods are largely comparable. It can even be noted that method $S(25, 6, r)$, which links on recalculated indices, produces less extreme deviations from the benchmark year on year indices than the method that links year on year indices on published indices. The results in sections 5.2 and 5.3 pointed out that product types belonging to division 05 are hard to deal with. Product aggregates are characterised by high rates of churn, including COICOP 054020. Method $S(25, 6, r)$ performs better than $S(25, 12, p)$, in spite of the fact that the latter directly links year on

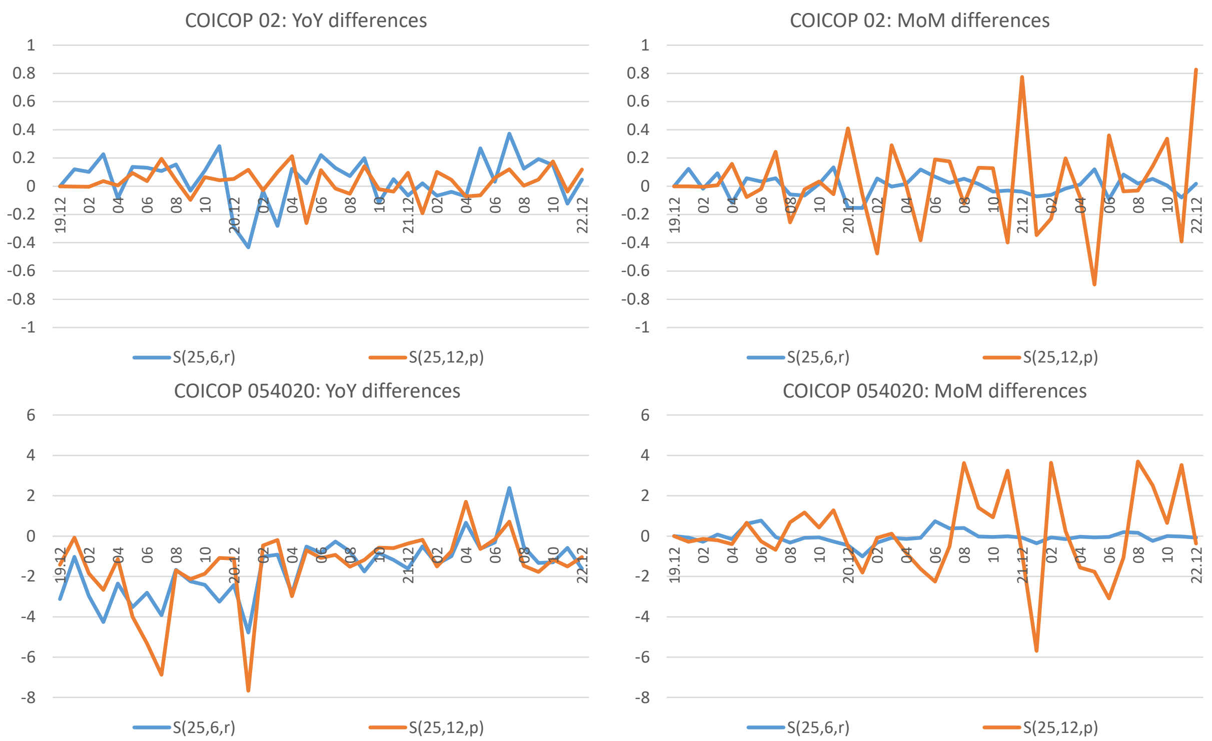


**Figure 6.1:** Year on year (YoY) and month on month (MoM) index differences (pp) with respect to benchmark indices of two product aggregates for two extension methods.

year indices on published indices.

The largest differences between the two methods are found in the month on month indices. The results for method $S(25, 6, r)$ are clearly more accurate than for method $S(25, 12, p)$ and are also visible for the higher aggregate Alcoholic beverages. The results confirm the aforementioned expectations about the possible effects on period to period indices when linking on recalculated and published indices.

The month on month differences for $S(25, 12, p)$ also indicate that inaccuracies may amplify over time, since there is no direct control over the month on month indices when linking on published indices. Methods that link on the last recalculated indices do not suffer from this drawback.

Finally, it is also interesting to note that index series obtained by linking on the last recalculated index are independent of the choice of initial period of an index series. This property holds because index series are extended by linking the period to period indices derived from the last two index series on the index published in the previous period. This also means that index series constructed with this linking approach are uniquely determined, irrespective of the amount of historical data used. Methods that link on published indices do not have this property, except for the method that links on the previous period.

## 6.3 Practical considerations

The choice of index method does not only have theoretical implications. Also practical aspects have to be considered, in particular concerning activities in production that involve index calculations. This subsection gives several examples of such activities and tries to give insight into the implications of choosing different extension methods.

An important stage between index compilation and publication of inflation figures is often referred to as "validation of indices". Statistical institutes may receive questions about rates of change that draw the attention from users of official statistics. It is therefore important to analyse and understand rates of change prior to their publication. There are different ways of analysing index figures; one approach is to calculate contributions to rates of change from product aggregates and individual products. Such methods enable staff members to identify those products or groups with the largest (positive or negative) influence on rates of change at different levels of aggregation. Such analyses provide a better understanding about the possible causes behind rates of change that are about to be published at different levels of aggregation (down to 5-digit level).

Understanding and explaining inflation figures is obviously very important, but a recently held survey by Statistics Netherlands revealed that statistical institutes also very much focus on month on month changes [10]. The complexity of an index method and the choice of index extension method has implications for the possibilities of finding an analytically tractable method for calculating product contributions.

A factor that simplifies the development of a method for this purpose is whether a rate of change to be examined is obtained by linking that change in the index extension to the index published in the reference period. For example, for the method $S(25, 12, p)$, or HASP-25, finding a method for calculating product contributions to year on year indices is easier than for month on month indices, which is very complicated. On the other hand, contributions to month on month indices are relatively easy to derive for methods that link on the last recalculated index. Deriving a method for contributions to year on year indices is more complicated in this case but should still be feasible.

Another frequently occurring problem in CPI production is changes in product classification. The composition of product aggregates may change over time, for instance because the composition of lower aggregates has changed and now better fits other higher aggregates. This means that products will be shifted from one aggregate to another at a certain point in time.

The question is how this problem affects the application of index extension methods. Methods that link on published indices are expected to produce index series that depend on the index series of the old classification, except when linking on the index of the previous period. A workaround may be needed for these methods in order to generate an index series according to the new classification.

Methods that link on the last recalculated index should resolve these classification problems on their own, because of the month on month linking that characterises these methods. Since the index series of the last two time windows are used in the extension, it is expected that the index series of an aggregate will reflect the price development that corresponds with the new classification.

Developing methods for product contributions and dealing with changes in product classification are topics that deserve further study, motivated by the new insights on the problem of index extension that result from the present study.

# 7 Conclusions

Studies that have been carried out so far on the problem of extending index series over time are limited to a number of classical extension methods. The study presented in this paper has considerably been expanded in scope, which includes all possible linking intervals for different lengths of the time window and also different options of linking on past indices. The broader scope has resulted in new insights on this topic, in particular on the extent to which different methods are able to control chain drift.

Some findings of previous studies have also been confirmed, such as the accurate year on year indices produced by the method HASP-25, which links year on year indices on published indices of 12 months ago, and the high sensitivity to drift of movement splice and window splice. The latter method links on the last recalculated index in the first period of each rolling time window. Those results initially motivated the author, and possibly also other researchers and statistical institutes, to avoid, or at least be very careful with methods that link on recalculated indices.

As a result, statistical institutes are heading towards HASP or mean splice, with 25 months almost becoming an accepted standard for window length. The present study has shown that other suitable choices exist, which are found to perform even better than the ones mentioned. The main findings that result from this study can be summarised as follows:

- While it's true that linking on recalculated indices is more sensitive to drift than linking on published indices, one of the key findings of this study is that it is possible to make appropriate choices for the linking interval in order to control drift.
- Generally speaking, the best results are obtained with methods that link on the last recalculated index, in particular for windows of 25 months. Linking on the indices of 6 months ago produces good to excellent results across all product types. Index series are free of drift at all-items level and accurate at lower levels of aggregation. The same can be said about adaptive splicing. Period to period indices are very accurate for both methods.
- Mean splice gives somewhat less accurate year on year indices, as it suffers more from the poorer results of linking on periods near the edges of a time window.
- The method HASP-25 yields very accurate year on year indices, which are comparable to adaptive splicing. But linking on published indices may give inaccurate period to period indices.
- Very good results are also obtained with methods that use 13-month windows and link on the last recalculated indices of 6 months ago. But these methods should not

be used for product types with high churn and frequency of clearance prices, which characterise COICOP division 05.

A very nice and elegant property of methods that link on the last recalculated index is that the index series are independent of the choice of starting period. This implies that the index series produced by this class of methods are unique. This does not hold for methods that link on published indices.

An advantage of the latter methods is that axioms like time reversibility and the identity test in a narrow sense (with both prices and quantities equal in two periods) are satisfied between the linking and reporting period. These properties are lost when linking on recalculated indices. However, given the accuracy of the methods listed above, this can hardly be considered a drawback.

This study has shown that there are different suitable methods to choose from. As was mentioned in Section 6.3, beside index theoretical considerations it is also important to value practical aspects when choosing an index method. Once a decision has been made, a method takes its course in production and will be used for different purposes, such as index validation. Choosing for a simpler method with a fixed linking interval has clear benefits over more refined methods like adaptive splicing. This leads to a trade-off between methodological complexity and accuracy of index figures. In this respect, the method denoted as $S(25, 6, r)$ is a recommended option, also for the arguments expressed in Section 6.1.

One of the insights gained from this study is that methods with 13-month windows that link on the last recalculated indices can produce very good results. The method that links on the indices of 6 months ago produces the most accurate index levels and year on year indices at all-items level, which hardly exhibit drift. Most countries have chosen or are seriously considering 25-month windows, but windows of 13 months should not be excluded. The use of 13-month windows is an interesting option for statistical institutes that have acquired their first transaction data, which may contain limited historical data (e.g. see a recent study by the statistical institute of Austria [33]).

The consistently good results with linking intervals of 6 months for different window lengths when linking on the last recalculated index is not a coincidence. The linking months used for deriving monthly chained year on year indices form a set that is centred around the reference month of one year ago, as was mentioned in Section 6.1. It is therefore expected that this choice will also produce accurate results for other multilateral methods like GEKS and TPD.

The promising potential of methods that link on recalculated indices motivates research on different topics. Some of these topics were mentioned in Section 6.3.

# Acknowledgements

Part of this research was funded from Eurostat grant 101034180-2020-NL-IMREVA, for which the author expresses his gratitude. The author also wants to thank colleague

Florin Barb for the discussions this paper benefited from. Correspondence can be sent to: ag.chessa@cbs.nl

# References


[1] ABS. Making greater use of transactions data to compile the Consumer Price Index. In *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, May 2017. https://www.bundesbank.de/en/bundesbank/research/making-greater-use-of-transactions-data-to-compile-the-consumer-price-index-australia-635722.

[2] Bert M. Balk. A simple method for constructing price indices for seasonal commodities. *Statistische Hefte*, 22:72–78, 1981. https://doi.org/10.1007/bf02932820.

[3] Bert M. Balk. *Price and quantity index numbers: Models for measuring aggregate change and difference*. Cambridge University Press, 2008.

[4] Jacek Białek. The general class of multilateral indices and its two special cases. In *17th Meeting of the Ottawa Group on Price Indices*, Rome, Italy, June 2022. https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/90eefd961765454eca25886a00042203/$FILE/The%20general%20class%20of%20multilateral%20indices%20and%20its%20two%20special%20cases_Paper.pdf.

[5] Jacek Białek and Anna Bobel. Comparison of price index methods for the CPI measurement using scanner data. In *16th Meeting of the Ottawa Group on Price Indices*, Rio de Janeiro, Brazil, May 2019. https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/bialek_bobel_paper_2.pdf.

[6] Antonio Gavino Chessa. A new methodology for processing scanner data in the Dutch CPI. *Eurostat Review on National Accounts and Macroeconomic Indicators*, 2016(1):49–69, 2016. https://ec.europa.eu/eurostat/cros/content/new-methodology-processing-scanner-data-dutch-cpi-antonio-g-chessa_en.

[7] Antonio Gavino Chessa. A comparison of index extension methods for multilateral methods. In *16th Meeting of the Ottawa Group on Price Indices*, Rio de Janeiro, Brazil, May 2019. https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/index_extension_methods_compared_chessa_og19_1.pdf.

[8] Antonio Gavino Chessa. A comparison of multilateral index extension methods for seasonal items. In *Meeting of the Group of Experts on Consumer Price Indices*, online meeting, June 2021. https://unece.org/sites/default/files/2021-05/Session_1_Netherlands_Paper.pdf.

[9] Antonio Gavino Chessa. A product match adjusted R squared method for defining products with transaction data. *Journal of Official Statistics*, 37(2):411–432, 2021. http://dx.doi.org/10.2478/JOS-2021-0018.

[10] Antonio Gavino Chessa, Florin Barb, Lisa Ooms, and Stefan Boumans. Further developments in multilateral index calculation and validation for the HICP. Technical report, Statistics Netherlands, The Hague, December 2022. Report written for Eurostat grant project 2020-NL-IMREVA.

[11] Antonio Gavino Chessa, Lucie Nan, Florin Barb, Koen Link, and Kim Ketelaar. Towards a comprehensive methodology for index calculation in the CPI with transaction data. Technical report, Statistics Netherlands, The Hague, December 2020. Report written for Eurostat grant project 2018-NL-IMCOPS.

[12] Antonio Gavino Chessa, Johan Verburg, and Leon Willenborg. A comparison of price index methods for scanner data. In *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, May 2017. https://www.bundesbank.de/en/bundesbank/research/a-comparison-of-price-index-methods-for-scanner-data-635922.

[13] Jörgen Dalén. Unit values and aggregation in scanner data - towards a best practice. In *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, May 2017. https://www.bundesbank.de/en/homepage/unit-values-and-aggregation-in-scanner-data-which-way-forward--636104.

[14] Jan de Haan. A framework for large scale use of scanner data in the Dutch CPI. In *14th Meeting of the Ottawa Group on Price Indices*, Urayasu City, Japan, May 2015. https://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf.

[15] Jan de Haan and Heymerik A. van der Grient. Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics*, 161:36–46, 2011. https://doi.org/10.1016/j.jeconom.2010.09.004.

[16] Jan de Haan, Leon Willenborg, and Antonio Gavino Chessa. An overview of price index methods for scanner data. In *Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, May 2016. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_1_room_doc_Netherlands_an_overview_of_price_index_methods.pdf.

[17] W. Erwin Diewert and Kevin J. Fox. Substitution bias in multilateral methods for CPI construction using scanner data. In *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, May 2017. https://www.bundesbank.de/en/bundesbank/research/substitution-bias-in-multilateral-methods-for-cpi-construction-using-scanner-data-635970.

[18] Walter Erwin Diewert. Axiomatic and economic approaches to international comparisons. In A. Heston and R.E. Lipsey, editors, *International and Interarea Comparisons of Income, Output and Prices*, volume 61, pages 13–87. University of Chicago Press, 1999.

[19] Ö. Eltetö and P. Köves. On an index computation problem in international comparisons. *Statiztikai Szemle*, 42:507–518, 1964. In Hungarian.

[20] Eurostat. Regulation (EU) 2016/792 of the European Parliament and of the Council, 2016. Official Journal of the European Union, 11 May 2016.

[21] Eurostat. Practical guide for processing supermarket scanner data. Technical report, Eurostat, Luxembourg, September 2017. https://circabc.europa.eu/sd/a/8e1333df-ca16-40fc-bc6a-1ce1be37247c/Practical-Guide-Supermarket-Scanner-Data-September-2017.pdf.

[22] Eurostat. Harmonised index of consumer prices (hicp): Methodological manual. Technical report, European Union, Luxembourg, November 2018. https://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/ks-gq-17-015.

[23] Roy C. Geary. A note on the comparison of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society A*, 121:97–99, 1958. https://doi.org/10.2307/2342991.

[24] C. Gini. On the circular test of index numbers. *Metron*, 2:3–24, 1931.

[25] ILO, IMF, OECD, UNECE, Eurostat, and World Bank. *Consumer Price Index Manual: Theory and Practice*. Geneva, 2004. https://doi.org/10.5089/9787509510148.069.

[26] Lorraine Ivancic, W. Erwin Diewert, and Kevin J. Fox. Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics*, 161:24–35, 2011. https://doi.org/10.1016/j.jeconom.2010.09.003.

[27] Salem H. Khamis. A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society A*, 135:96–121, 1972. https://doi.org/10.2307/2345041.

[28] Frances Krsinich. The FEWS index: Fixed Effects with a Window Splice – non-revisable quality-adjusted price indices with no characteristic information. In *Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, May 2014. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/New_Zealand_-_FEWS.pdf.

[29] Claude Lamboray. The Geary Khamis index and the Lehr index: How much do they differ? In *15th Meeting of the Ottawa Group on Price Indices*, Eltville am Rhein, Germany, May 2017. https://www.bundesbank.de/en/bundesbank/research/the-geary-khamis-index-and-the-lehr-index-how-much-do-they-differ--636014.

[30] Ken Van Loon and Dorien Roels. Integrating big data in the Belgian CPI. In *Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, May 2018. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Belgium.pdf.

[31] Robert Summers. International price comparisons based upon incomplete data. *Review of Income and Wealth*, 19(1):1–16, 1973. https://doi.org/10.1111/j.1475-4991.1973.tb00870.x.

[32] B. Szulc. Index numbers of multilateral regional comparisons. *Przeglad Statysticzny*, 3:239–254, 1964. In Polish.

[33] Adam Tardos. Introduction of scanner data into austrian CPI and HICP: Practical implementation experience, with a focus on window length options. In *Meeting of the Group of Experts on Consumer Price Indices*, Geneva, Switzerland, June 2023. https://unece.org/sites/default/files/2023-06/1a.2%20Austrian_CPI_and_HICP_practical_implementation_Experience_Adam_Tardos.pdf.

[34] Ludwig von Auer. The Generalized Unit Value index family. *Review of Income and Wealth*, 60:843–861, 2014. https://doi.org/10.1111/roiw.12042.

[35] Ludwig von Auer. Processing scanner data by an augmented GUV index. *Eurostat Review on National Accounts and Macroeconomic Indicators*, 2017(1):73–91, 2017. https://ec.europa.eu/eurostat/cros/system/files/euronaissue1-2017-art3.pdf.

[36] Li-Chun Zhang, Ingvild Johansen, and Ragnhild Nygaard. Evaluating unit-value price indices in a dynamic item universe. In *16th Meeting of the Ottawa Group on Price Indices*, Rio de Janeiro, Brazil, May 2019. https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/statistics_norway_paper.pdf.

# List of Figures



# List of Tables