Extension of multilateral index series over time: Analysis and comparison of methods

Antonio G. Chessa

Department of Consumer Prices
Statistics Netherlands

Email: ag.chessa@cbs.nl

✧✧✧✧✧

Paper written for the 2021 Meeting of the
Group of Experts on Consumer Price Indices

7 May 2021

Summary

One of the most essential, but also less rigorously treated problems with regard to multilateral index methods, is how to compute and continue series of price indices over time when new data become available in the next period (e.g. month).

Objective: This paper characterises and formalises this "index extension problem" and extension methods. Different extension methods are compared on various analytical properties including drift and on their performance on a number of transaction data sets.

Data: Statistics Netherlands receives transaction data from various supermarket and other retail chains, which are used in the monthly compilation of its Consumer Price Index. Seasonal items pose various challenges to price index calculation. Transaction data of six product categories were selected in the present study: fresh fruit, fresh vegetables, clothing, footwear, plants and flowers, and garden furniture.

Results: The 25-month window half splice method (HASP), which links year on year indices on published indices of one year ago, gives the most accurate results when compared with the transitive full period index series. Most monthly year on year index differences lie within several tenths of a percentage point. Inflation figures produced by HASP are free of drift. Extension methods that make use of 13-month windows lead to considerable downward drift. Year on year indices differ by more than 1 percentage point from HASP in most cases, with the largest differences shown by footwear (2-5 pp per year) and garden furniture (mostly above 5 pp per year).

Conclusions: This study recommends using multilateral methods with a 25-month rolling window and the method HASP for extending index series over time. The average year correction method AYCO is recommended for adjusting the index level of new series when linking these series to previously published series in the CPI in situations where new data sources or new methods are introduced.

Keywords
Multilateral methods, index extension, transitivity, transaction data, seasonal items, CPI.

Contents

1 Introduction

The Consumer Price Index (CPI) is going through a series of rapid developments since the last decade. The number of National Statistical Institutes (NSIs) that are planning to acquire transaction data from retailers, or possess such large electronic data sets and are studying how to use transaction data to compile parts of the CPI is rapidly increasing. The number of NSIs in Europe that are currently using transaction data in their CPI has increased from 4 to more than 10 in five years' time. The acquisition of transaction data can be a lengthy process for different reasons, so that NSIs are also looking for alternative data sources to replace their traditional price collection in physical stores or via web sites. Web scraping has become increasingly popular among NSIs over the past few years.

The transition to processing large data sets like transaction and web scraped data gives rise to different challenges. Such data sets may contain tens of thousands of products at the barcode level, which we denote by the term "item", in accordance with the official term GTIN for barcode (Global Trade Item Number).[1] Tasks like classification, that is, the mapping of items to COICOP, can no longer be done manually.

Also the problem of defining samples of products within broader "product categories" (COICOP-5 or at a more detailed level) enters a different dimension when switching from traditional price collection to transaction or web scraped data. When dealing with very large sets of items there may not be only a single "representative item" for each product, but possibly a set of items with the same or comparable quality. These sets of items have to be identified and combined within the same strata, in particular to capture price changes of items that undergo a so-called "relaunch". Product strata cannot be defined manually, so that efficient methods are needed for this task as well. Statistics Netherlands has developed the method MARS for this purpose ([8], [9], [12]), which has already been used in CPI production [4].

The transition from traditional price collection to large electronic data sets has also increased the possibilities with regard to index methods that can be applied. This holds in particular for transaction data since such data sets contain both price and quantity information of items sold. Another important property of transaction data is the assortment dynamics, which contrasts with the static populations of items defined in traditional surveys. These properties have motivated a quest for methods that can deal with the greater dynamics in a better way than traditional bilateral methods.

Bilateral methods that use expenditure based product weights lead to drift. This holds both for chained indices and for fixed base indices, as these index methods are not transitive (apart from cases with simplified weighting schemes). The last decade has shown an increased focus on multilateral methods. These methods were originally developed for international price comparisons; well-known examples are the GEKS method ([15], [17], [27]), the Country Product Dummy method [26] and the Geary-Khamis method ([16], [23]). The first study that considers multilateral methods in the time domain probably goes back to 1981 (see [3]), but it is only since approximately 10 years that these methods

[1]The use of the terms "item" and "product" may differ across NSIs. For instance, the ONS (UK) uses "product" to denote "item" as used in this paper, while "item" is used by ONS to indicate a product category.

have been studied more systematically with regard to the CPI (since contributions [19] and [22]). The increased acquisition and use of transaction data by NSIs has been accompanied by an increased interest and research into multilateral methods, in particular over the past five years ([1], [5], [6], [7], [14], [20], [24], [25], [28], [31]).

Multilateral methods have many advantages over bilateral methods. Expenditure share based product weights can be defined, which are allowed to vary from period to period. This also implies that new products can be included in index calculations from the period of introduction, without having to impute prices in periods without sales. Price and quantity data of all items can be processed in every period. Transitive index series can be calculated with multilateral methods while including all these types of dynamics.

These new developments clearly show that the palette of methodological choices has considerably extended in different directions. With so many choices available, the problem of compiling accurate index numbers has become much more complex. An important question therefore is how insight could be gained into the choices that NSIs eventually have to make. A recent study carried out at Statistics Netherlands has quantified the impact of different choices with regard to classification, product stratification and index method on price indices [4].

One specific choice was left out of the cited study. When multilateral methods are used for international price comparisons, the set of countries or regions is usually fixed. Countries are replaced by periods (e.g. months) when transferring these methods to the time domain. The set of periods, which we will refer to as the "time window", obviously changes when data of the next period become available. A sequence of index series is thus generated for successive time windows, which have to be linked in order to produce the next index to be published. Methods that perform this linking are also known as "index extension methods".

Different index extension methods have been proposed over the past years. Although some comparative studies on multilateral methods also include index extension methods, it seems that this part of index calculation has not yet received the attention it deserves. The purpose of index extension methods is to continue an index series from one period to the next, which is key to any time series modelling.

The recent paper [10] presented a broad comparative study, which showed that so-called "splicing methods" can lead to severe drift. Methods that link index series to base month indices perform much better, since these methods are free of drift between successive base months by construction. The present paper builds on this study, but now focuses exclusively on seasonal items. These items have a significant weight in the CPI. To what extent index extension methods are able to deal with seasonal patterns is therefore a very relevant question.

This paper is organised as follows. Extension methods are applied to index series that are calculated with two multilateral methods. These two methods are described in Section 2: the Geary-Khamis method, which is also called "QU-GK method", and the GEKS-Walsh method. Section 3.1 gives a characterisation of index extension methods in terms of the choices that have to be made to specify these methods. An overview of extension methods is presented afterwards and the index formulas are given for each method (Section 3.2).

A fundamental basis for the index extension problem, the behaviour of extended index series and analyses of properties of extension methods seems to be an area that requires further development. Extended series of price indices will eventually be published, which means that inflation figures will be based on such series. It is therefore essential to gain a deeper insight into the aforementioned aspects. Expressions for the year on year indices are derived for six extension methods and statements are made with regard to the behaviour of extended index series based on an alternative definition of drift-free index series (Section 3.3).

The transaction data sets to which the index extension methods have been applied are described in Section 4. The data sets cover a broad range of product categories: fresh fruit and vegetables, clothing, footwear, plants and flowers, and garden furniture. The results for the index extension methods are presented in Section 5. Switching to a new data source and/or index method in the CPI may distort the new index series, the extent of which depends on the method used to link the new to the current series. Seasonal items in particular may be very sensitive to changes in seasonal patterns when new methods or data sources are used to compute index series. Section 6 discusses and compares different linking methods. Conclusions and recommendations are summarised in Section 7.

2 Two multilateral methods

2.1 The QU-GK method

2.1.1 The class of QU methods

In early 2014, Statistics Netherlands initiated a study as part of an innovation programme concerning its CPI, with the aim of finding a more generic method for processing transaction data and calculating price indices. This research has led to the development of a method, which we have called the "QU method", an abbreviation of Quality adjusted Unit value method. As the name suggests, this method is an extension of unit values as average prices for homogeneous products, that is, for products of equivalent quality.

The following notation is introduced in order to explain this idea. For convenience, months will be used as time units throughout this paper. Let G_t denote a set of items sold in month t, and let $p_{i,t}$ and $q_{i,t}$ be the price and quantity sold for item i in month t, respectively. As was already stated in the previous section, items that leave an assortment at clearance prices and their replacement items have to be linked in order to capture possible associated price increases. This linking of items under such item relaunches results in broader products or strata than at the individual item level.

The relaunch problem clearly shows the importance of product stratification, which aims at finding strata of items of the same or comparable quality (see [8], [9], [11] and [12] for the stratification method MARS developed at Statistics Netherlands). Product stratification falls outside the scope of the present paper from a methodological perspective. It was therefore decided to use the same notation for prices and quantities at the item and product level in this paper for the sake of clarity.

Let $s_{i,t}$ denote the expenditure share of item i in period t: $s_{i,t} = p_{i,t}q_{i,t}/\sum_{j\in G_t} p_{j,t}q_{j,t}$.

The unit value index $U_{0,t}$ in period t with respect to a comparison period 0 can alternatively be written as a ratio of weighted harmonic averages of prices in the two periods:

$$U_{0,t} = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_t} q_{i,t}}{\sum_{i \in G_0} p_{i,0} q_{i,0} / \sum_{i \in G_0} q_{i,0}} = \left(\frac{\sum_{i \in G_t} s_{i,t} (p_{i,t})^{-1}}{\sum_{i \in G_0} s_{i,0} (p_{i,0})^{-1}} \right)^{-1} \tag{2.1}$$

Quantities of sold products of different quality cannot be simply summed, since their quality differences would then be completely ignored. For instance, if an assortment improves in quality over time and if the quality improvements are reflected by higher prices, then these price increases would lead to a higher price index. This is incorrect, since price changes are related to quality changes. The price index will then be affected by what is known as "unit value bias".

In order to account for quality differences between items, we introduce "quality adjustment factors" $\nu_i > 0$ for each item i. We use these factors to adjust the item prices, so that the right-hand side of expression (2.1) can also be applied to the adjusted prices. We can thus write the resulting quality adjusted price index $P_{0,t}$ as follows:

$$P_{0,t} = \left(\frac{\sum_{i \in G_t} s_{i,t} \left(\frac{p_{i,t}}{\nu_i} \right)^{-1}}{\sum_{i \in G_0} s_{i,0} \left(\frac{p_{i,0}}{\nu_i} \right)^{-1}} \right)^{-1} \tag{2.2}$$

Note that the adjustment factors ν_i do not affect the expenditure shares. Expression (2.2) can be rewritten as follows:

$$P_{0,t} = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t}}{\sum_{i \in G_0} p_{i,0} q_{i,0}} \Big/ \frac{\sum_{i \in G_t} \nu_i q_{i,t}}{\sum_{i \in G_0} \nu_i q_{i,0}} \tag{2.3}$$

The first ratio in (2.3) is the change in total expenditure between 0 and t and the second, rightmost ratio is a weighted quantity index. This shows that the QU method automatically satisfies the "product test" in the axiomatic approach to index theory (the price index times the quantity index equals the value index). Index formulas (2.2) and (2.3) are also transitive when the values of the ν_i are fixed over time.

Another important property of index formula (2.3) is that this expression retains its form at different levels of aggregation. For instance, the expression at item level and at stratum or product level is the same. The ν_i have the same value for all items within the same stratum, so that (2.3) can be easily rewritten to an expression consisting of product prices, quantities and ν_i at stratum level. This implies that aggregation within strata of qualitatively homogeneous items is fully consistent with unit values.

Beside the aforementioned product dimension, index formula (2.3) also encapsulates aggregation formulas over time and geographical space. Changing the default spatio-temporal units used by many NSIs from monthly national to smaller units, such as by day or week and by outlet or region, will normally influence the ν_i but not the index formula itself. As in the aforementioned aggregation from items to products, the same index formula (2.3) can be used to aggregate expenditures and (weighted) quantities over time and space to yield monthly national indices or at any other aggregate spatio-temporal level. This can be achieved with the class of QU methods without having to calculate monthly national prices.

Applications and results for different transaction data sets can be found in the IMCOPS report written at Statistics Netherlands [11]. Aggregation over product, temporal and spatial dimensions is much more difficult to perform with other multilateral methods, like Time Product Dummy and GEKS. Their index formulas are not invariant to certain forms of aggregation (e.g. over time) and are not consistent with unit values when there are no quality differences involved in the aggregation process (see also the aforementioned IMCOPS report, Section 4.6.2, for a discussion).

An elegant feature of the QU family of index methods therefore is that the index formulas can be expressed at the most elementary level at which prices and quantities are available, irrespective of the level of spatio-temporal aggregation, product definition and quality differences. The only factors that are affected by choices concerning products or strata, and the temporal and geographical units are the adjustment factors ν_i. This is an essential detail of this class of methods. Intuitively, it makes sense to have this property in an index formula.

2.1.2 The GK method

Index formulas (2.2) and (2.3) represent a class of index formulas, since different choices with regard to the adjustment factors ν_i lead to different index formulas. The QU class of methods encompasses both bilateral and multilateral methods. For instance, setting the ν_i equal to current period or base period prices, or to their average prices, results in the Laspeyres, Paasche or Fisher price index formulas, respectively.

Note that the QU index formula is also essentially bilateral if we would know how to express quality differences explicitly among items through the ν_i. However, this is an impractical task since transaction data sets may contain tens of thousands of items at the GTIN level. In this respect, the conventional approach to index theory is also followed when applying multilateral methods. That is, quality differences between strata are in fact expressed as price ratios. The difference with bilateral methods is that these prices are defined over a time window that is longer than the one or two periods commonly used in bilateral methods.

There are different possibilities to define the ν_i in terms of multiple item prices and quantities over a time interval or window $[0, T]$. The paper by Chessa [6] describes three multilateral QU versions, which are applied and compared on transaction data. In the Geary-Khamis method, the ν_i are defined as weighted average product prices over the time window $[0, T]$. Price changes have to be excluded from the ν_i, since these factors are part of the quantity index. The indices $P_{0,t}$ therefore also act as deflators of the item prices in different periods of the time window. The deflated prices in the ν_i also ensure that the price indices are homogeneous in current and base period prices; this property is important for the constant tax CPI (e.g. VAT changes). The ν_i are defined as follows:

$$\nu_i = \frac{\sum_{z=0}^{T} q_{i,z} p_{i,z} / P_{0,z}}{\sum_{z=0}^{T} q_{i,z}} \tag{2.4}$$

The numerator of this expression is the sum of the deflated expenditures for item i over the time window $[0, T]$, while the denominator sums the number of sold items over the same time window. It should be noted that the two sums in (2.4) also extend over different items that belong to the same strata when products are defined at a broader level than the item level.

Expressions (2.2) or, equivalently (2.3), and (2.4) define the Geary-Khamis method in the time domain, which we abbreviate to QU-GK or simply GK. Since the price indices appear as deflators in (2.4), it is not possible to calculate price indices directly. Different methods exist for solving the resulting system of equations. One way to calculate price indices is to use an iterative algorithm, which starts with an arbitrary sequence of price indices as initial values. These values are used to compute initial ν_i values, which are substituted into either (2.2) or (2.3) to yield updated price indices. This procedure is repeated until the difference between the last two index sequences drops below a threshold set by the user. More details about this procedure and the QU-GK method can be found in [2], [6], [11], [16] and [23].

In addition to the properties and advantages of the QU class mentioned at the end of Section 2.1.1, an important reason for choosing the QU-GK method in particular at Statistics Netherlands is the robustness of price and quantity indices to clearance prices of items that are about to leave the stores. It follows directly from expressions (2.3) and (2.4) that clearance prices hardly affect GK indices. Not all multilateral methods have this property, such as the Törnqvist and Fisher variants of the GEKS method. This will be discussed in the next subsection. This means that the QU-GK method does not need price filters in order to exclude clearance prices from the index calculations. A very elegant property of this method therefore is that the data can be used without any form of manipulation of the original data, apart from general operations like removal of outliers.

2.2 The GEKS-Walsh method

A different class of multilateral methods is defined by the GEKS method ([15], [17], [27]). The GEKS method operates in a different way than the QU method and other methods, such as the Time Product Dummy and hedonic methods. It starts by computing bilateral indices $P_{z,t}$ for each pair (z, t) of time periods within a time window $[0, T]$. Bilateral indices that are not transitive are transformed by the GEKS method into a transitive index series. This is accomplished by computing an equally weighted geometric average of bilateral indices over a set of time paths:

$$P_{s,t} = \prod_{z=0}^{T} (P_{s,z} P_{z,t})^{\frac{1}{T+1}} \tag{2.5}$$

Note that a subset of all possible time paths is chosen to compute GEKS indices. This observation was used in [29] and [30] as a point of departure to develop a more generalised method (called the "cycle method" by the author).

In principle, any bilateral index formula could be used as input in expression (2.5). However, bilateral indices must satisfy the time reversibility property as a minimum requirement in order for GEKS indices to be transitive. Index formulas like Paasche and Laspeyres are not time reversible and can thus be discarded as candidates. The Fisher index, which is the geometric average of the Paasche and Laspeyres index, does satisfy time reversibility. It can be shown that expression (2.5) yields transitive index series [19]. Time reversibility is therefore not only a necessary condition, it is also sufficient in order to achieve transitivity within the GEKS framework.

The Fisher index belongs to a class of index formulas called "superlative indices" by followers of the economic approach to index theory. According to the axiomatic approach, the Fisher index satisfies all basic 20 properties, known as "tests" or "axioms", of bilateral indices (see [21], pp. 292-296). The Törnqvist and the Walsh index are also superlative indices, although they do not satisfy all 20 tests.

Index theorists claim that the three superlative index methods will give approximately the same results. However, applications to transaction data have shown that this is not true. In [7], the authors provide analytical arguments which show that the GEKS-Törnqvist is sensitive to downward trends when exiting products are sold under clearance prices. The same holds when using the Fisher index. The extent of downward deviations from multilateral methods that are practically insensitive to clearance prices can be substantial, also at retail chain level (see [11], Section 8.2). Downward trends can be suppressed by setting filters for clearance prices. But finding suitable threshold values is not straightforward and it may take quite some time to find suitable values, which also have to be monitored over time. What might work well for one product category may turn out be inadequate for other product categories.

Better options are available for the GEKS. The Walsh index is expected to be more robust to clearance prices. The Walsh price index in month t with respect to month z is given by the following expression:

$$P_{z,t} = \frac{\sum_{i \in G_{z,t}} p_{i,t} \sqrt{q_{i,z} q_{i,t}}}{\sum_{i \in G_{z,t}} p_{i,z} \sqrt{q_{i,z} q_{i,t}}} \tag{2.6}$$

where $G_{z,t}$ is the set of items that are sold in both months.

According to (2.6), the basket is defined as a geometric mean of the numbers of products sold in two months. Clearance prices are usually accompanied by very small numbers of products sold. As can be seen in (2.6), contributions of clearance prices to the index can be ignored, since the geometric means of the quantities sold of the corresponding terms will vanish in the numerator and the denominator. For this reason, we prefer the Walsh index over the Fisher and the Törnqvist index.

3 The index extension problem

3.1 Characterisation of extension methods

An important motivation for using multilateral methods is that transitive index series can be obtained. In other words, multilateral index series do not suffer from chain drift. Transitivity holds for any time window, also for windows that are adjusted in order to accommodate data of next months. A sequence of transitive index series is thus generated on subsequent time windows. These index series have to be linked in order to obtain price indices in every month. But here a problem arises: month on month indices will usually change for the same months in successive windows as new data is added.

Index series of subsequent time windows can be linked in different ways. As month on month indices are generally not the same for different windows, one can imagine that different linking or extension methods may produce different results. Questions that naturally arise are what linking strategies are appropriate, how this could be assessed and, hence, how 'well-behaving' extension methods could be distinguished from methods that perform less well.

Before addressing these questions it is important to give a characterisation of extension methods. The choices that have to be made in order to define an extension method are given below. Next, six extension methods are described, the index formulas are given and expressions for the year on year indices are derived for every extension method. Some remarks are made about the behaviour of the extension methods with regard to "drift", which will be defined more precisely in the successive subsection 3.3.2.

The first choice to be made is the length of the time window. Although window length is clearly an important element, it is not included in the characterisation of index extension methods, as was also decided in the comparative study [10] that preceded the present study. This decision is made in order to separate linking strategy, which is the essence of index extension, from other choices regarding multilateral methods. Window length can nevertheless affect the behaviour of index series, also with regard to drift. Window length will therefore return in discussions and analyses in later sections of this paper.

Once the length of the time window is set, choices are made on three elements that characterise index extension methods:

A. The adjustment of the time window from month to month;

B. The linking month;

C. The index in the linking month.

Different choices can be made with regard to each of the three elements, which are described below.

Window adjustment

Two methods have been proposed in the literature:

A.1 A "rolling" or "moving" window. The window keeps its full length when it is shifted each month to include data from the next month;

A.2 A monthly expanding window. The window is extended by one month when data of the next month become available. The first month of the window is fixed, for instance, December of the previous year (i.e. the base month). Expanding windows then consist of two months in January, three months in February, etc.

Linking month
The index series of the most recent adjusted window is used to calculate an index in the current month. For this purpose, each new index series is linked to an index in a past month, which we refer to as the "linking month". Any month between the first month and the penultimate month of the adjusted window can be chosen as linking month. The possible choices can be subdivided into two main types:

B.1 A moving linking month;

B.2 A fixed linking month.

Some examples may be useful in order to illustrate the meaning of these two choices. The first month of each rolling window results in a moving linking month, which is shifted by one month as the time window is moved one month. A rolling window may also be used in combination with a fixed linking month, for instance, when price indices are calculated with respect to the base month as comparison period. More details and examples will be given in the next subsection when the different extension methods are described.

Index in the linking month
There is one additional choice element, which seems to have been understated in the current literature. The choice of index in the linking month turned out to play a crucial role in the analysis of the behaviour and the differences between extension methods in the first study [10]. Linking index series of subsequent windows generates an index in the current month, which is the index that will be published. Subsequent window adjustments generate a sequence of additional indices after the first calculated (and published) index, which we call "recalculated indices". Some extension methods link onto the published index, while other methods take a recalculated index for linking a new series. This means that there are two main choices for the index in the linking month:

C.1 A recalculated index;

C.2 A published index.

Different decisions on each of the three choice elements lead to different extension methods. The next subsection describes and formalises the six extension methods that were also applied in the first comparative study [10].

3.2 Index extension methods

This subsection describes two classical splicing methods (window and movement splice), two fixed base extension methods and two variants of the window splice method which link onto published indices. The following notation is used. Let $P_{0,t}$ denote a price index to be published in month t, where 0 denotes the starting month of the index series. Next, $T + 1$ is used as the 'default length' of the time window, which is fixed in the case of rolling windows and is equal to the maximum window length for expanding windows. For different reasons (symmetry, practical reasons) we consider window lengths with an odd number of months, so that T is an even number. In practical applications, T is set equal to 12 months.

3.2.1 Window splice

This is an A.1–B.1–C.1 type of method and was proposed by Krsinich [24]. The method uses a fixed-length rolling window. The first month of the rolling window is used as linking month. Because of this property, the method is sometimes also called "full window splice". The most recently calculated index in the linking month is used to link the new index series.

We now explain this procedure in formal terms in order to gain a better understanding of the indices that are generated by window splice. The method starts with a sequence of index numbers on the initial time window $[0, T]$. The second window in the window splice method is obtained by shifting the initial window by one month, which yields the window $[1, T+1]$.

A new index series is calculated on this window. From this index series an index in the current month $T+1$ is calculated, which is the index that will eventually be published. We denote the time series on $[1, T+1]$ by $\{P_{1,z}^{[1,T+1]}\}_{z=1}^{T+1}$, where $P_{1,z}^{[1,T+1]}$ denotes the index in month z with respect to month 1 on $[1, T+1]$. By convention, the index $P_{1,1}^{[1,T+1]}$ for $z=1$ is set to 1. The index series is chained onto the index $P_{0,1}$ from the initial time window in month 1, so that we obtain the following indices in months $z = 2, ..., T+1$ with respect to starting month 0:

$$P_{0,z}^{[1,T+1]} = P_{0,1} \cdot P_{1,z}^{[1,T+1]} \tag{3.1}$$

The index in month $z = T+1$ is the index that will be published in that month. Since the published index in $T+1$ is denoted by $P_{0,T+1}$, the index in month $T+1$ according to window splice is equal to:

$$P_{0,T+1} = P_{0,T+1}^{[1,T+1]} = P_{0,1} \cdot P_{1,T+1}^{[1,T+1]} \tag{3.2}$$

where the right-hand side follows from expression (3.1). Window splice thus generates a published index in $T+1$ by linking the index calculated in this month on the window $[1, T+1]$ onto the index $P_{0,1}$ of T months ago obtained on the previous (initial) window. Figure 3.1 gives an illustration of the window splice method.

Expressions (3.1) and (3.2) can be generalised for any month $t > T$ as follows. The time window used to calculate a published index in month t is $[t-T, t]$. The index series calculated on this window is $\{P_{t-T,z}^{[t-T,t]}\}_{z=t-T}^{t}$. This index series is linked onto the latest calculated index in month $t-T$, which was obtained after linking an index series that was calculated on the preceding window $[t-1-T, t-1]$. The index of T months ago on which the current index series will be linked is therefore equal to $P_{0,t-T}^{[t-1-T,t-1]}$. The linking produces the following indices in months $z = t-T+1, ..., t$:

$$P_{0,z}^{[t-T,t]} = P_{0,t-T}^{[t-1-T,t-1]} \cdot P_{t-T,z}^{[t-T,t]} \tag{3.3}$$

The index for $z = t$ is the published index in month t, that is:

$$P_{0,t} = P_{0,t}^{[t-T,t]} = P_{0,t-T}^{[t-1-T,t-1]} \cdot P_{t-T,t}^{[t-T,t]} \tag{3.4}$$

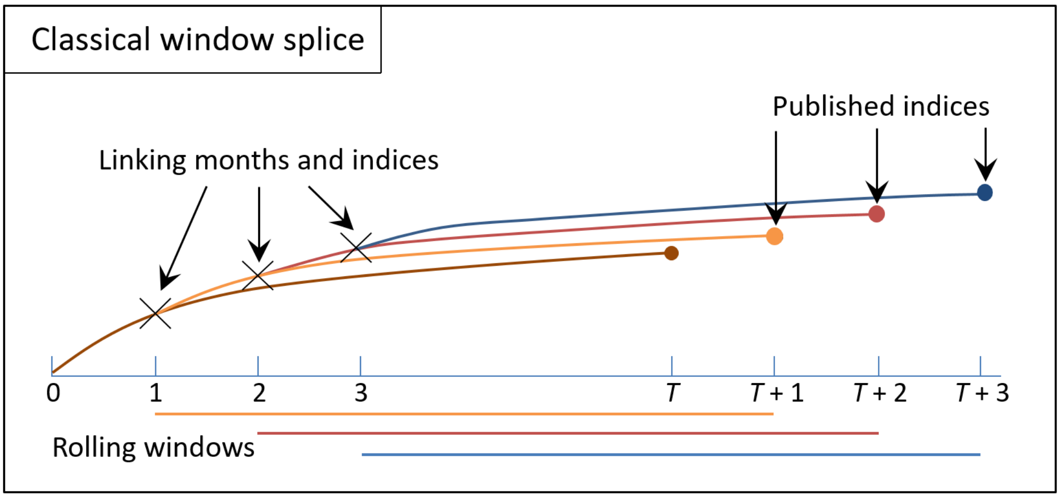

Figure 3.1: Illustration of the classical window splice method.

where the right-hand side of this expression follows by setting $z = t$ in (3.3). Note that expression (3.4) is equal to (3.2) when $t = T + 1$.

Window splice results in a sequence of published indices $\left\{P_{0,T+s}^{[s,T+s]}\right\}_{s\geq 1}$ where s denotes the number of window shifts. The aforementioned sequence thus extends the initial transitive index on the time window $[0, T]$ to successive months. It is also important to note that window splice results in a sequence of T indices in each month as a result of the window shifts. In an arbitrary month $t \geq T + 1$, the first calculated index is a published index $P_{0,t}^{[t-T,t]}$ on the window $[t - T, t]$. The subsequently generated sequence of $T - 1$ indices $\left\{P_{0,t}^{[t-T+s,t+s]}\right\}_{s=1}^{T-1}$ are recalculated. The most recently calculated index of this sequence, that is $P_{0,t}^{[t-1,t+T-1]}$, is used as a linking index in order to calculate the index to be published in month $t + T$.

3.2.2 Movement splice

This is an A.1–B.1–C.2 type of method and was proposed by Ivancic, Diewert and Fox [22] as part of the Rolling Year GEKS, which is usually shortened to RYGEKS or RGEKS (see also [19]). The penultimate month of the adjusted window is taken as linking month, and the month on month index of the rolling window is chained to the published index of the previous month.

Movement splice links onto published indices, which makes this method much simpler and easier to understand than window splice. Movement splice calculates the following index in every month $t > T$:

$$P_{0,t} = P_{0,t-1} \cdot P_{t-1,t}^{[t-T,t]} \tag{3.5}$$

Movement splice thus takes the month on month index $P_{t-1,t}^{[t-T,t]}$ in t with respect to $t - 1$ of the index series calculated on $[t - T, t]$, which is linked to the published index in $t - 1$. The result is a published index in month t.

3.2.3 Fixed base rolling window

This is an A.1–B.2–C.2 type of method. This method (FBRW), which was suggested by Lamboray [25], also uses a rolling window and links onto a published index, but takes a base month as linking month. The FBRW method calculates a sequence of fixed base indices from the rolling windows, which are used as the eventual published indices.

Also this method starts with an initial set of transitive indices $P_{0,0}, P_{0,1}, ..., P_{0,T}$ on a window $[0, T]$. Let 0 and T denote the first two base months and let bT denote the base month after b 'base shifts', for an integer $b \geq 1$. The published index in base month bT is $P_{0,bT}$. An index to be published in month $t \in [bT + 1, (b + 1)T]$ is calculated as follows:

$$P_{0,t} = P_{0,bT} \cdot P_{bT,t}^{[t-T,t]} \tag{3.6}$$

It can be easily verified that the published indices in each base month are equal to the indices without applying extension, that is, the indices according to the transitive index series on the intervals $[bT, (b + 1)T]$ for every $b \geq 0$.

3.2.4 Fixed base expanding window

This is an A.2–B.2–C.2 type of method and was suggested by Chessa [6] (shorthand: FBEW). The only difference with the FBRW method is that FBEW uses an expanding window instead of a rolling window.

Suppose that indices have been calculated until and including base month bT. For the next index in month $bT + 1$, a new expanding window is initialised with starting month bT and $bT+1$ as the second month. The window is expanded with each successive month until it reaches its maximum length $T + 1$ in month $(b + 1)T$. The FBEW method does not require an initial index series on $[0, T]$, since the method uses an expanding window with a base month as its first month.

Suppose we have an index $P_{0,bT}$ published in base month bT. An index to be published in month $t \in [bT + 1, (b + 1)T]$ is calculated as follows:

$$P_{0,t} = P_{0,bT} \cdot P_{bT,t}^{[bT,t]} \tag{3.7}$$

Note that the only difference with the FBRW method is the use of expanding windows $[bT, t]$ compared to the rolling windows $[t-T, t]$ in the FBRW method. Like in the FBRW method, also the published indices in the base months that are calculated with the FBEW method are equal to the indices according to the transitive index series computed on the windows of full length $T + 1$. The base month indices are equal to the indices calculated with the FBRW method.

An illustration of the FBEW method is given in Figure 3.2. Note that the index notation in this figure is simplified. Time windows are sequentially numbered within brackets, and the comparison period is omitted for simplicity.

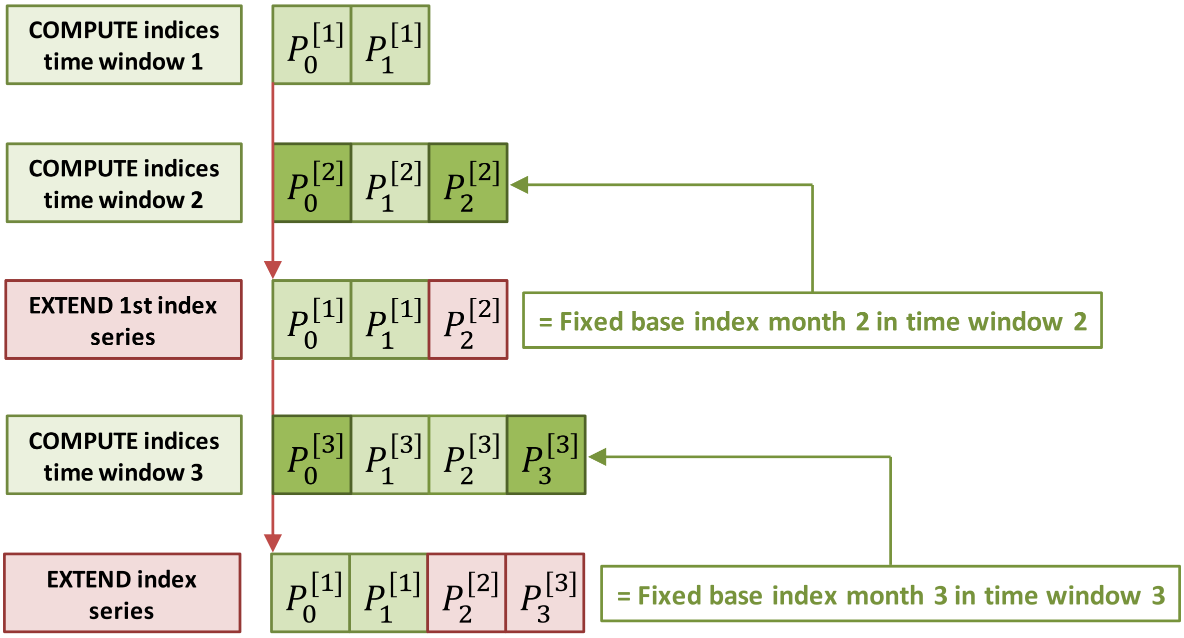

Figure 3.2: Illustration of the method FBEW for two extensions after the bilateral index calculation on the initial 2-month window. $P_0^{[\cdot]} = 1$ for every time window.

3.2.5 Window splice on published indices

The first comparative study on index extension methods revealed that window splice leads to drift, which can often be severe [10]. The results of the cited study also suggested a downward nature of this drift. A high impact of the downward drift was even found at chain level; the year on year indices for a large supermarket chain were almost 1 percentage point below the full period transitive indices (i.e. without index extension).

Window splice produces a published index in month t according to expression (3.4). The last recalculated index in month t, that is $P_{0,t}^{[t-1,t+T-1]}$, is used as linking index in order to calculate a published index T months later. The published and recalculated indices are usually different, which implies that the month t on month $t+T$ indices for the published series are not equal to the transitive indices that are linked onto recalculated indices. This implies that circularity over periods of length T is lost, which also holds for any other lag. More details will be given in Section 3.3.

Full transitivity is difficult to achieve, since new data enter the index calculations once a time window is shifted. But the behaviour of indices generated with window splice can nevertheless be improved by performing the linking in a simpler way. In fact, the idea is similar to the linking performed in the FBRW and FBEW methods. These fixed base methods control drift over periods of length $T+1$ by linking fixed base indices of subsequent windows on the published indices in the base months.

In the previous comparative study, Chessa [10] suggested to link the index $P_{t-T,t}^{[t-T,t]}$ in month t with respect to $t-T$, which is obtained on the window $[t-T,t]$, to the published index in month $t-T$ instead of the last recalculated index. This technique yields a published index in months $t > T$, which takes the following form:

$$P_{0,t} = P_{0,t-T} \cdot P_{t-T,t}^{[t-T,t]} \tag{3.8}$$

This method will be called "window splice on published indices", which, in shorthand, will be referred to as WISP.

3.2.6 Half splice on published indices

In principle, any month can be chosen as linking month. De Haan [18] suggested to take the central month of a rolling window (for windows with an odd number of months). This results in a method known as "half splice". Also half splice leads to drift when linking on recalculated indices, for the same reasons as window splice. In the first comparative study it was already suggested to perform the linking on published indices for the half splice method as well [10].

The cited study showed that a "half splice on published indices" (HASP) with a window of 25 months gave better results than WISP with 13 months. The WISP method produced indices with erratic behaviour in some cases, which disappeared with the longer window used in the HASP method. In practical applications, HASP links year on year indices to published indices of 12 months ago like WISP, but uses an additional 12 months before the central linking month. One of the questions in the present study is whether a longer window will produce better results for seasonal items.

Generally speaking, we consider a version of the half splice method with a window length of $2T + 1$ instead of $T + 1$. The choice for a longer window in the half splice method implies that HASP starts with a transitive index series $P_{0,0}, P_{0,1}, ..., P_{0,2T}$ on the window $[0, 2T]$. A published index is calculated in months $t > 2T$ as follows:

$$P_{0,t} = P_{0,t-T} \cdot P_{t-T,t}^{[t-2T,t]} \tag{3.9}$$

The linking is carried out in the same way as in expression (3.8) for the window splice variant WISP, with the longer window used in HASP being the only difference. An illustration of the method HASP is given in Figure 3.3.

3.3 Analysis of extension methods

This section presents an analysis of the six extension methods described in Section 3.2. The analysis focuses on two sets of properties of index series generated by the extension methods: expressions for the year on year indices of each extension method are derived in Section 3.3.1, while Section 3.3.2 gives an assessment of the behaviour of the six extension methods with regard to drift and a set of other index-theoretical axioms.

3.3.1 Year on year index formulas

In this subsection, expressions are derived for the year on year indices for each extension method, as the annual rate of inflation is a statistic of primary interest. Time windows of

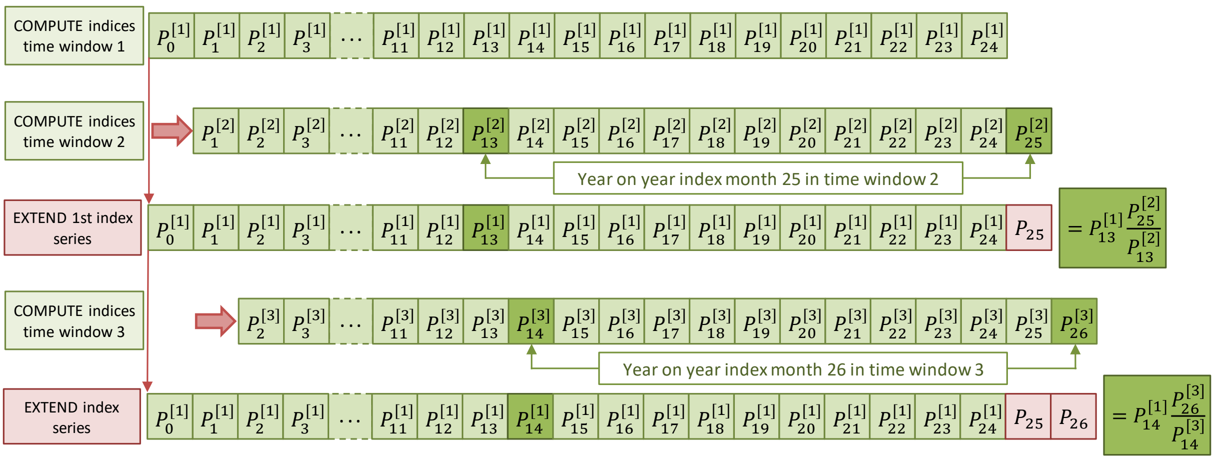

Figure 3.3: Illustration of the method HASP for two extensions of the initial 25-month index series.

13 and 25 months are considered in the applications of this study, so that $T = 12$ months, but we continue to use the notation T for this time lag in this section. The published year on year index in month t will thus be denoted as $P_{t-T,t}$.

Window splice

The expression for the index in month t with respect to $t - T$ follows from expression (3.4) and can be written as follows:

$$P_{t-T,t} = \prod_{z=t-2T}^{t-T-1} \frac{P_{z,z+1}^{[z,z+T]}}{P_{z,z+1}^{[t-2T,t-T]}} \cdot P_{t-T,t}^{[t-T,t]} \tag{3.10}$$

The derivation of this expression is not straightforward, but it was decided to leave out the steps leading to the above expression to enhance readability.

It is interesting to note that the year on year index that results from the published indices is not equal to the year on year index obtained from the multilateral index on the window $[t - T, t]$. The difference between these two year on year indices is equal to the product term in (3.10). The time windows in the numerator and the denominator of this term are different, so that the product term will usually differ from 1.

This observation has an important implication for the behaviour of indices computed and published according to the classical window splice method. If the product term in (3.10) is not equal to 1, then combining month on month indices within time intervals of length $T + 1$ will give a result that differs from the index $P_{t-T,t}^{[t-T,t]}$ that is obtained by directly comparing prices and quantities from two periods that are separated by T months. This holds for any time lag T, which means that published index series do not possess any form of transitivity (also known as "circularity").

Classical window splice is therefore sensitive to drift. In view of the discussion of the results in Section 5, it may be instructive to give an illustration of the possible nature of the drift. Note that the time windows in the numerator of (3.10) are shifted forward in time with respect to the window in the denominator, for all $z > t - 2T$. Suppose that products will leave an assortment at clearance prices after $t-2T$. The shifting windows $[z, z+T]$ in the numerator of the product term will then be increasingly dominated by clearance prices, which has a downward effect on the price indices $P_{z,z+1}^{[z,z+T]}$ in the numerator of (3.10), as is argued in [10]. The cited study clearly shows that window splice may lead to severe drift, which may persist in severity even at retail chain level, and also that the drift is predominantly of a downward nature.

Movement splice

The expression for the index in month t with respect to $t-T$ follows from expression (3.5) and can be written as follows:

$$P_{t-T,t} = \prod_{z=t-T}^{t-1} \frac{P_{z,z+1}^{[z-T+1,z+1]}}{P_{z,z+1}^{[t-T,t]}} \cdot P_{t-T,t}^{[t-T,t]} \tag{3.11}$$

This expression is very similar to (3.10) for classical window splice. Also expression (3.11) shows that the year on year index that results from the published indices is not equal to the year on year index $P_{t-T,t}^{[t-T,t]}$ that is obtained on the window $[t-T, t]$. This means that also movement splice does not possess any form of circularity and is therefore not free of drift. This is not a surprising conclusion, since movement splice is basically a monthly chained index method.

In contrast with window splice, note that the windows $[z-T+1, z+1]$ in the numerator of (3.11) are shifted backwards with respect to the window $[t-T, t]$ in the denominator. As a consequence, the occurrence of clearance prices may have an opposite, upward effect on price indices calculated with movement splice. A part of the results in [10] seems to suggest this behaviour.

Fixed base rolling window

Consider $t \in [bT+1, (b+1)T]$, for an integer $b > 0$, which was previously referred to as the number of shifts of the base month. From expression (3.6) it follows that the year on year index for this extension method can be written as:

$$P_{t-T,t} = \frac{P_{(b-1)T,bT}^{[(b-1)T,bT]}}{P_{(b-1)T,t-T}^{[t-2T,t-T]} P_{t-T,bT}^{[t-T,t]}} \cdot P_{t-T,t}^{[t-T,t]} \tag{3.12}$$

Like for window and movement splice, also the year on year index for the FBRW method differs by a factor from the year on year index $P_{t-T,t}^{[t-T,t]}$ that is obtained on the window $[t-T, t]$. However, it can be easily verified that the two year on year indices are equal for $t = (b+1)T$. This means that pairs of periods can be found, namely bT and $(b+1)T$, for which the published year on year indices equal the year on year indices of transitive index series on the rolling windows. This makes it possible to control FBRW indices for drift.

This was exactly the thought behind the construction of the fixed base extension method in [6]: index series generated in this way will be equal in each base month to an index that is obtained from a transitive series on windows of length $T + 1$, which will prevent published indices to run away from a transitive index on such time windows. However, absence or occurrence of drift not only depends on whether indices of an extended series can be matched in some way to transitive indices; the behaviour of extended index series with regard to drift also depends on window length. This will be treated in greater detail in the next subsection and in the presentation and discussion of the results in Section 5.

The window $[(b-1)T, bT]$ in the numerator of (3.12) is centred around the union of the two intervals in the denominator, as t runs through $[bT+1, (b+1)T]$. It can therefore be expected that price indices generated by the FBRW method will also behave well in months t that are not equal to the base month.

Fixed base expanding window

Consider again months $t \in [bT + 1, (b + 1)T]$, for an integer $b > 0$. The two fixed base extension methods are very similar and only differ by window length, which is only the same for indices calculated in the base months. It may therefore be instructive to write the year on year index for the FBEW method in terms of the year on year index for the FBRW method. The two year on year indices are denoted by $P^{FBEW}_{t-T,t}$ and $P^{FBRW}_{t-T,t}$. These two indices are related as follows:

$$P^{FBEW}_{t-T,t} = \frac{P^{[t-2T,t-T]}_{(b-1)T,t-T}}{P^{[(b-1)T,t-T]}_{(b-1)T,t-T}} \frac{P^{[bT,t]}_{bT,t}}{P^{[t-T,t]}_{bT,t}} \cdot P^{FBRW}_{t-T,t} \tag{3.13}$$

Expression (3.13) shows that the differences between the two fixed base methods can be attributed to differences in window length. The two year on year indices move towards each other when t increases and are equal when $t = (b + 1)T$. Both fixed base extension methods can therefore be said to be free of drift, under the condition that time windows of length $T+1$ are an appropriate choice. The largest differences can be expected for months t that are close to $bT + 1$, that is, when the number of months used in the expanding window method is small. Expression (3.13) does not indicate that systematic differences can be expected between the two fixed base extension methods. Both methods performed very well in the first study [10].

Window splice on published indices

The version of window splice that links on the published index of $T = 12$ months ago obviously yields year on year indices that are equal to those obtained on the window $[t - T, t]$. That is, the published year on year indices will be the same as those calculated on the rolling window. By expression (3.8) we have:

$$P_{t-T,t} = P^{[t-T,t]}_{t-T,t} \tag{3.14}$$

Index series generated by the WISP method are therefore, by construction, able to control drift. Note that the year on year indices obtained with this method are equal to the year on year indices for the two fixed base methods in the base months. As the WISP method links year on year indices in each month, it can be expected that it will generally give more accurate year on year indices than the two fixed base methods.

Half splice on published indices

This is in fact a copy of the WISP method, the only difference being that HASP includes an additional T months before the (central) linking month of the time window, which has a length of $2T + 1$ months. By expression (3.9) we have:

$$P_{t-T,t} = P_{t-T,t}^{[t-2T,t]} \tag{3.15}$$

The year on year indices published according to HASP are equal to the year on year indices calculated on the rolling windows $[t - 2T, t]$, which means that also HASP is capable of controlling extended index series for drift. WISP and HASP will normally give different results, which can be ascribed to the different window lengths. The use of a longer window in the HASP method, and the symmetrical use of the window in its linking strategy, may lead to more accurate and stable results, as was already noted in the first comparative study [10].

3.3.2 Axiomatic properties

The exposition given so far may leave the reader with a paradoxical feeling. On the one hand, deciding to work with multilateral methods has clear advantages compared to bilateral methods. Transitivity is one of the most important properties of multilateral index series and has been a main driver for focusing on multilateral methods. On the other hand, transitivity is lost when the game is actually played as we move from one period to the next. This raises the important and inevitable question how linked index series of subsequent time windows could be assessed. A yardstick is needed that allows us to establish whether an index series produced by an index extension method behaves well or not.

This subsection presents an analytic comparison of the six index extension methods described in Section 3.2. The comparison builds on the previous subsection where expressions for the year on year indices of published series were derived for each extension method. Statements with regard to drift were made based on the resulting year on year expressions, which are incorporated within a broader axiomatic framework below. First, some additional remarks are made with regard to drift, which is a fundamental property for assessing the behaviour of extended index series.

The loss of transitivity in extended index series forces us to think about a different definition of drift-free index series. An indication was given in the previous subsection by comparing the year on year indices of extended index series with those of the transitive index series on successive time windows. This comparison is included as one of the main elements of the definition of a drift-free index series, which is formulated below. The time lag T mentioned in the definition is the same time variable used to characterise time windows.

Definition. *An extended index series* $\{P_{0,z}\}_{z=0}^{t}$ *is* free of drift *if, for given time lags* T*, pairs of periods* $(\tau+(i-1)T, \tau+iT)$*, with* $i = 1, ..., b := \lfloor \frac{t}{T} \rfloor$, $0 \leq \tau < T$, $\tau + bT \leq t$*, and intervals* $W_i \subset \mathbb{R}$ *of equal length exist, such that* $P_{\tau+(i-1)T,\tau+iT} = P^{W_i}_{\tau+(i-1)T,\tau+iT}$ *for all* $i = 1, ..., b$*. Index series on* W_i *are transitive.*

This definition says that an extended index series is free of drift when a set of time intervals $[\tau, \tau+T], [\tau+T, \tau+2T], ..., [\tau+(b-1)T, \tau+bT]$ can be found, with b equal to the smallest integer greater than or equal to $\frac{t}{T}$, such that the indices on each of these intervals are equal to indices according to a transitive index series, which are computed on sets W_i.

This gives a necessary condition for an extended index series to be free of drift. If the stated condition in the above definition cannot be satisfied, then an extended index series cannot be said to be free of drift. This is the case for classical window splice and movement splice, which followed from the year on year index formulas that were derived in Section 3.3.1.

The condition stated in the definition is not sufficient in order to have a drift-free index series. That is, an extended index series that satisfies the condition in the above definition can still lead to drift. This depends on the choice of the intervals or time windows W_i which, for example, could be chosen too small for seasonal items. The question is how we could choose these windows.

This could be formalised by introducing sequences $\{W_i^{(n)}\}_{n\geq 1}$ such that $W_i^{(k)} \subset W_i^{(m)}$ for all $k, m \geq 1$ with $k < m$. A value of n and corresponding time intervals $W_i^{(n)}$ could be established such that $|P^{W_i^{(n)}}_{\tau+(i-1)T,\tau+iT} - P^{W_i^{(n-1)}}_{\tau+(i-1)T,\tau+iT}| < \epsilon$ for some $\epsilon > 0$. This would also turn the condition stated in the definition into a sufficient condition for a drift-free index series.

The condition on the absolute difference tells us that we can set $W_i = W_i^{(n)}$ if the absolute difference between the two indices drops below a value ϵ. This condition should not be too difficult to operationalise in practice. This was not done in this study because of the novelty of this idea. Instead, the index series of different extension methods were compared with full period transitive indices. These comparisons also give very good indications on the choice of window length, with the main question in the present study being whether 13-month windows are sufficient or whether longer time windows are needed for seasonal items.

The CPI manual lists 20 tests or axioms in the axiomatic approach to bilateral indices (see [21], pp. 292-296). A selection of these axioms was made in the present study: the identity test, the fixed basket or constant quantities test, homogeneity or proportionality in both current and base period prices, invariance to proportional changes in current and base period quantities, commensurability and time reversal. The drift-free property defined above is added and completes the list.

The identity test is often a topic of debate during international meetings with other NSIs and is therefore included. However, the identity test as intended in this paper is the version of the axiom in "the narrow sense", which means that both prices and quantities have the same values in the current and comparison period. The identity test in its most well-known standard form, in which only product prices stay at the same values in two periods, regardless of the quantities, is not uncontroversial (see the CPI manual [21] on this test and footnote 21 on page 293).[2] Homogeneity in prices is relevant with regard to the constant tax CPI, in particular when VAT rates change.

The QU-GK method satisfies the aforementioned axioms, except for the invariance to proportional changes in quantities. The reason behind the violation of this test is that the adjustment factors ν_i are affected by changes in the product quantities. Alternative definitions of the ν_i exist that resolve this problem, for instance, by defining the ν_i as equally weighted deflated product prices over a time window. The problem with this version of the QU method is that clearance prices will have too high a weight compared to the very low sales, so that price filters would be needed. Comparisons between this method, after setting price filters, and the QU-GK method resulted in small differences, so that the violation of the invariance to proportional changes in quantities by the QU-GK method seems to be of minor practical relevance [6].

The GEKS-Walsh method satisfies all selected axioms, except for the fixed basket test. Although the bilateral Walsh price index does satisfy this test, this property is lost after applying the GEKS formula (2.5).

The question is what happens to these properties for QU-GK and GEKS-Walsh after applying extension methods. The adjustment factors ν_i of the products will normally change after each adjustment of the time window, so that it can be imagined that the axioms will be harder to satisfy. The behaviour of the six extension methods is summarised in Table 3.1.

Axiom	**WS**	**MS**	**FBRW**	**FBEW**	**WISP**	**HASP**
Drift-free property	No	No	Yes	Yes	Yes	Yes
Time reversibility	No	Lag 1	Fixed base	Fixed base	Lag T	Lag T
Identity (narrow sense)	No	Lag 1	Fixed base	Fixed base	Lag T	Lag T
Fixed basket						
QU-GK	No	Lag 1	Fixed base	Fixed base	Lag T	Lag T
GEKS-Walsh	No	No	No	No	No	No
Homogeneity in prices	Yes	Yes	Yes	Yes	Yes	Yes
Invariance to prop. quantity changes						
QU-GK	No	No	No	No	No	No
GEKS-Walsh	Yes	Yes	Yes	Yes	Yes	Yes
Commensurability	Yes	Yes	Yes	Yes	Yes	Yes

Table 3.1: Axioms satisfied and violated by the six index extension methods.

The derivation of the year on year index formulas of the six extension methods in Section 3.3.1 already pointed out that classical window and movement splice are not free of drift. The other four extension methods are indicated as being free of drift, which means under the condition that the time windows used in these methods are of appropriate length as intended according to the previously given explanation in this subsection.

[2] The controversy can also be illustrated by constructing a counterexample with a fixed set of products that do not vary in price over time but do differ in price across products with marginal differences in quality, so that the price index is practically identical to the unit value index. The price index is not necessarily equal to 1 if the sold product quantities change over time.

Time reversibility, the narrow identity test and the fixed basket test (for QU-GK) hold for the extension methods indicated in the table when comparing prices and quantities in the current and in the linking period, but not for other pairs of periods by virtue of the moving time windows and the resulting time varying adjustment factors ν_i in case of the QU-GK method.

Homogeneity and commensurability are scalar multiplications applied to prices and/or quantities, which are preserved when extending index series. Overall, we can say that classical window splice and movement splice show the poorest performance on the tests in Table 3.1, while WISP and HASP yield the best performance.

4 Transaction data sets

Statistics Netherlands is using transaction data in its CPI since 2002. Transaction data from nearly 40 retail chains are currently used in the Dutch CPI, which account for more than 31 per cent of the CPI in terms of COICOP weights. Retail chains send the data files each week to Statistics Netherlands, most of which contain weekly expenditures and quantities sold per GTIN. In some cases, the data sets contain transaction information at a more fine-grained time scale. An example are fuel prices and litres, which are specified by the time of fueling for every single transaction.

In this study, transaction data sets are selected that cover six COICOPs of seasonal items: 01161 Fresh or chilled fruit, 01171 Fresh or chilled vegetables, 0312 Garments, 0321 Shoes and other footwear, 05112 Garden furniture and 09332 Plants and flowers. Shorter or more "common language" descriptions are used in this paper for some of the COICOPs, as is indicated in Table 4.1. The data sets are selected from five retails chains; the data on fresh fruit and vegetables are from the same supermarket chain. These data were also used in the first study [10], but separate results for fruit and vegetables were not reported because a much broader coverage of product types was used in that study. Also the data on clothing were used in [10]; results at the main 0312 COICOP level and also for lower aggregates were already reported in this preceding study.

COICOP	**Retail chain**	**Months**
Fresh fruit	Supermarket chain	48
Fresh vegetables	Supermarket chain (same)	48
Clothing	Department store	47
Footwear	Shoe shop	49
Garden furniture	Do it yourself store	49
Plants and flowers	Garden centre	35

Table 4.1: Transaction data sets and COICOPs selected for this study.

The data sets on footwear, garden furniture and plants and flowers are added to this study and were not used previously. The data on plants and flowers were recently acquired by Statistics Netherlands at the time this study was carried out, which explains the shorter period for which data are available.

Each data set in Table 4.1 represents a vast assortment of products, as all the corresponding retail chains have a significant market share in the Netherlands. The data on clothing and footwear contain expenditures and quantities of sold items for men, women and children. Taken over all retailers, the selected COICOPs contribute about 7 per cent to the total Dutch CPI. Other seasonal COICOPs with substantial weights are passenger transport by air, package holidays and accommodation services (although their weights have been substantially reduced in 2021 as a result of the Covid-19 crisis and restrictions). Nevertheless, this illustrates the importance of having critically assessed index extension methods for calculating inflation figures of seasonal items.

5 Results

5.1 Results for the six extension methods

Price indices can be calculated once we have defined the sets of products in each COICOP. Product categories with high rates of churn, such as clothing, pharmacy products and consumer electronics, require that items that leave the stores and re-enter with different GTINs and possibly at higher prices are linked in order to capture price increases (the so-called relaunches). The linking can be achieved in an automated fashion with the method MARS, a variance-based method that is adjusted by the degree to which products can be matched over time with respect to a comparison period, so that MARS accounts for product churn (see [8], [9], [12]).

For stable assortments, that is, with low rates of churn, GTINs can be chosen as unique products. Broader product definitions result when MARS is applied to product categories with higher rates of churn. GTINs can be combined into broader products or strata by selecting product attributes with MARS, so that GTINs with the same characteristics are assigned to the same product.

The method MARS was applied to the data sets used in this study. This led to the following product definitions for each of the six COICOPs:

- GTINs are chosen as unique products for fresh fruit, vegetables, garden furniture and plants and flowers;
- Products are defined in terms of sets of characteristics for clothing and footwear. For example, shoes are characterised by brand and type of shoes, while a distinction is made between shoes for boys and girls in the COICOP Footwear for infants and children.

Prices (unit values) and total quantities were calculated for each product or stratum, which were subsequently used in expressions (2.3) and (2.4) for the QU-GK method and in (2.5) and (2.6) for the GEKS-Walsh method. Price indices for the six COICOPs were calculated with these two methods, and index extensions were carried out on the calculated index series by applying the six extension methods described in Section 3.2. Time windows of 13 months were used except for the half splice method, which uses a window of 25 months.

The price indices obtained with the extension methods are shown in figures 5.1 and 5.2. Figure 5.1 contains the indices of four extension methods; the indices for the two methods that splice on published indices are separately shown in Figure 5.2. The indices are compared with the price indices that are simultaneously calculated on the full period for which data are available. These transitive indices are used as benchmarks (indicated by the fatter green lines in the two figures). The price indices in figures 5.1 and 5.2 are calculated by applying the QU-GK method on each time window. Comparisons with the GEKS-Walsh are shown in Section 5.3.

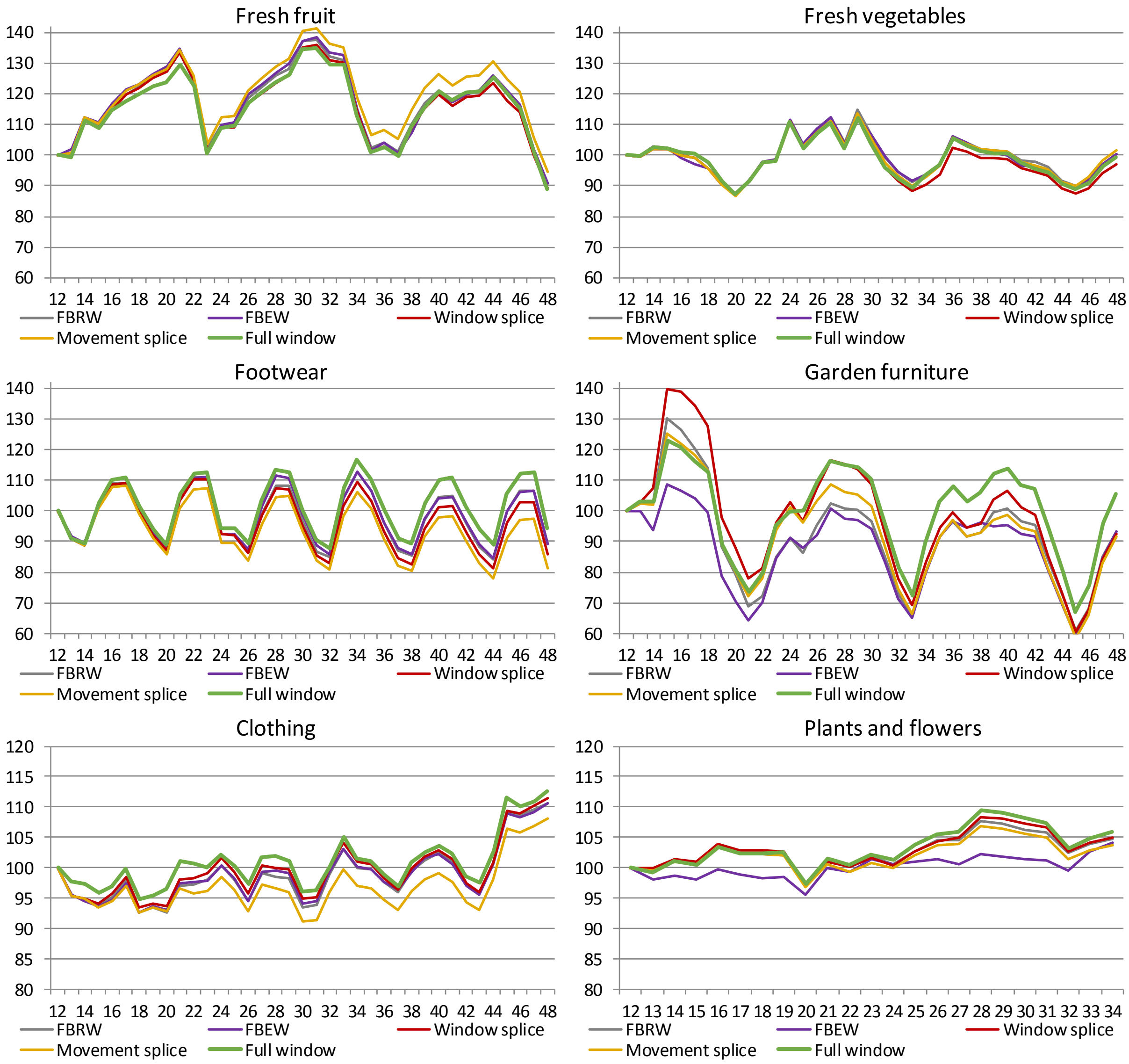

Figure 5.1: Price indices (QU-GK) for four extension methods compared with the full period indices.

Price indices that are calculated with extension methods that make use of 13-month windows take effect from the second year, except for the FBEW method which uses an expanding window. Index extensions with the FBEW method are carried out from the first year. The indices for the first year are excluded from all graphs, since these results do not contribute much to the comparisons. The price indices in each graph have month

12 as starting month, which is December of the first full calendar year for which data are available.

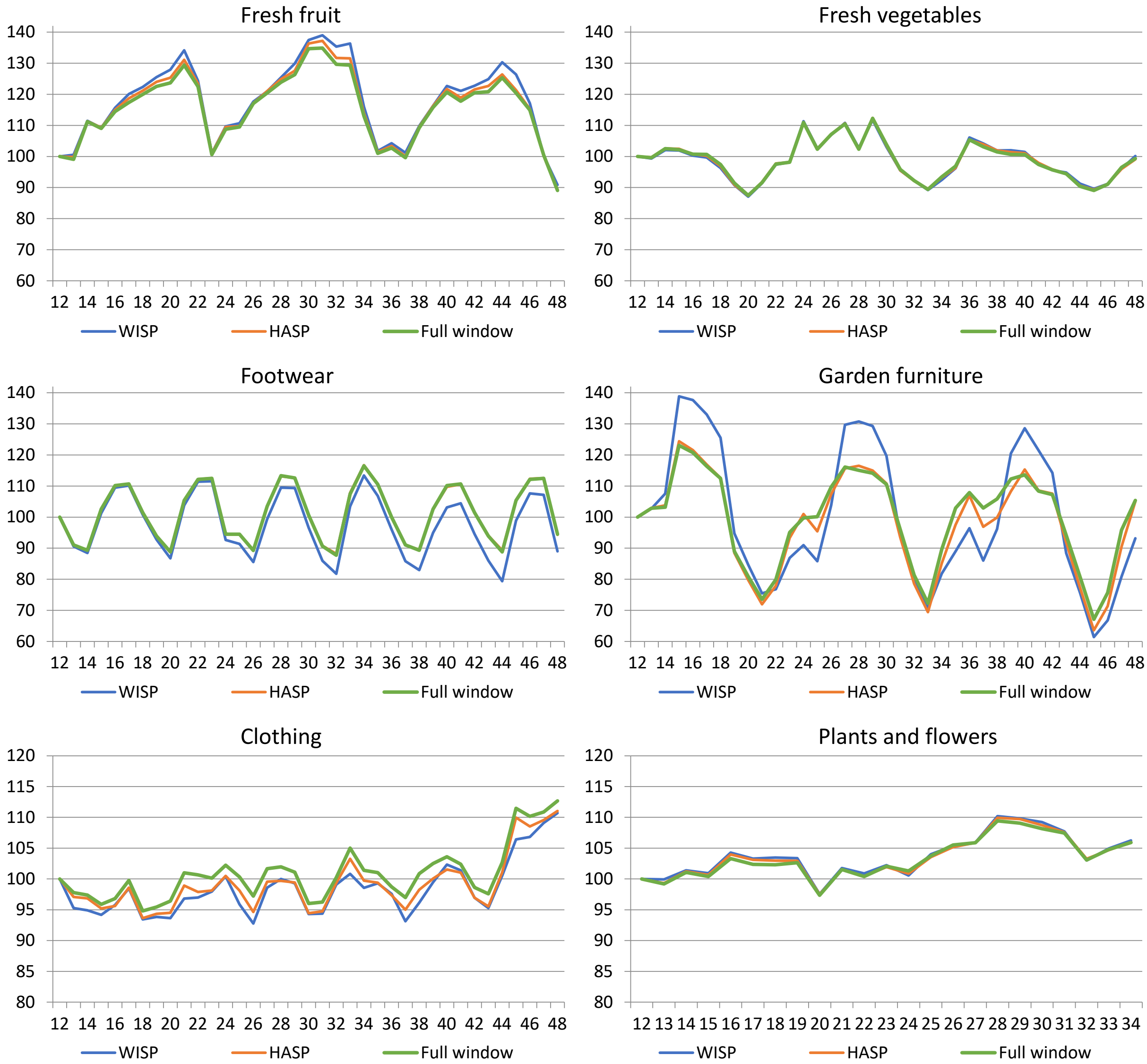

Figure 5.2: Price indices (QU-GK) for WISP and HASP compared with the full period indices.

Figure 5.1 shows that movement splice is the method that gives the poorest overall performance. Based on the discussion in Section 3.3.1 we would expect systematic upward effects with movement splice due to the lower out-of-season prices. A clear upward effect only occurs for fresh fruit, while a small upward effect can also be noted for fresh vegetables. The indices for the other four COICOPs lie below the full window transitive indices. One of the reasons for this behaviour could be that 13-month windows are generally too short for seasonal items.

Classical window splice performs slightly better than movement splice, but gives rather bad results for footwear and garden furniture. The fixed base methods perform well for fresh fruit, fresh vegetables and quite well also for clothing. The indices for footwear obtained with the two fixed base methods are better than for the two splicing methods, but still differ considerably from the full period transitive index. The indices for garden furniture also exhibit considerable differences with respect to the transitive benchmark index.

Plants and flowers show a strikingly poor performance of the FBEW expanding window method, which also produces large deviations for garden furniture. These two product categories build up their full assortments during the spring and summer months. The expanding window method does not capture the prices and quantities of items launched during the previous spring and summer, which offers an explanation for the poor performances shown in Figure 5.1. The use of a fixed-length rolling window gives much better results for the two COICOPs.

The differences with the full period indices become smaller when splicing on published indices (Figure 5.2). The window splice variant WISP still deviates from the benchmark indices, in particular for footwear and garden furniture. Enlarging the window to 25 months as in the half splice method HASP substantially increases the accuracy of the indices. The HASP method shows very small deviations in most periods for the majority of the COICOPs. The precision of the HASP indices is particularly impressive for footwear and garden furniture, for which none of the other extension methods gives satisfactory results. Also the results in Figure 5.2 make clear that a 13-month window is generally too short for seasonal items.

A better impression of the differences can be obtained by calculating the year on year indices for each extension method and by comparing these with the year on year indices for the full period indices. The differences in each month are shown in Figure 5.3, while Figure 5.4 shows the differences averaged over all months per calendar year.

Differences of at least two percentage points are frequently found in every COICOP for all extension methods that use a 13-month window, in particular when considering the differences per month (Figure 5.3). Especially the results for footwear and garden furniture are very poor, both in terms of the size and the direction of the differences. The year on year indices are systematically smaller than the full period year on year indices. The same holds for plants and flowers, and to a large extent also for clothing, although the differences are smaller than for footwear and garden furniture.

Because the 13-month window extension methods yield systematically lower year on year indices than the full period indices, these differences will not average out over a year. This can be seen in Figure 5.4. The differences on a yearly basis are still very large. The window splice on published indices turns out to be the best method that uses a 13-month window, but also the average differences can still be quite large for this method as well.

The half splice on published indices is the only method that performs well for each COICOP. The average differences in the year on year indices compared with the full period indices are in the order of several tenths of a percentage point at most, with the exception of a single year for garden furniture, in which case the differences are somewhat larger. Two main reasons can be advanced for the good performance of the HASP method:

- The accuracy of the indices clearly benefits from a 25-month window. The results show that windows of 13 months are generally too short for seasonal items;
- Price indices calculated with the half splice method also benefit from the symmetry

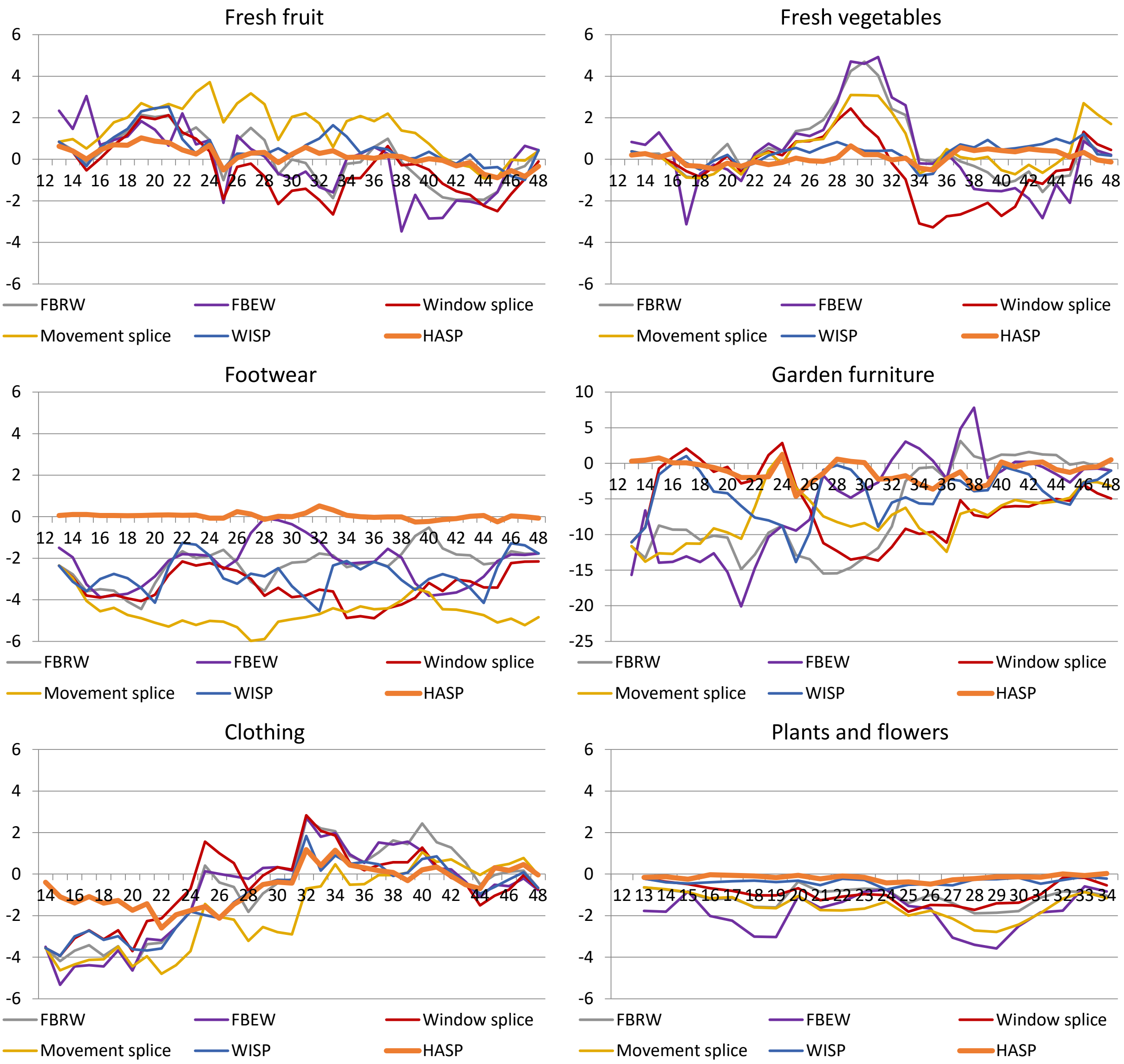

Figure 5.3: Differences between the year on year indices for each extension method and the full period indices, with the latter indices subtracted from the indices obtained with the extension methods. The differences are expressed in percentage points.

in the 25-month window, in which month 13 acts as the linking month. Using 12 months before the linking month has the great advantage of including a fairly large number of regular prices when products are about to leave an assortment shortly after the linking month. In the case of the QU-GK method, the regular prices ensure that the adjustment factors ν_i will be hardly influenced by clearance prices. This will prevent price indices from downward drift, as was also argued in the previous study [10].

5.2 Results for other methods

This section considers a number of variants of the six extension methods. First, we discuss the mean splice method proposed by Diewert and Fox. Next, we investigate the behaviour

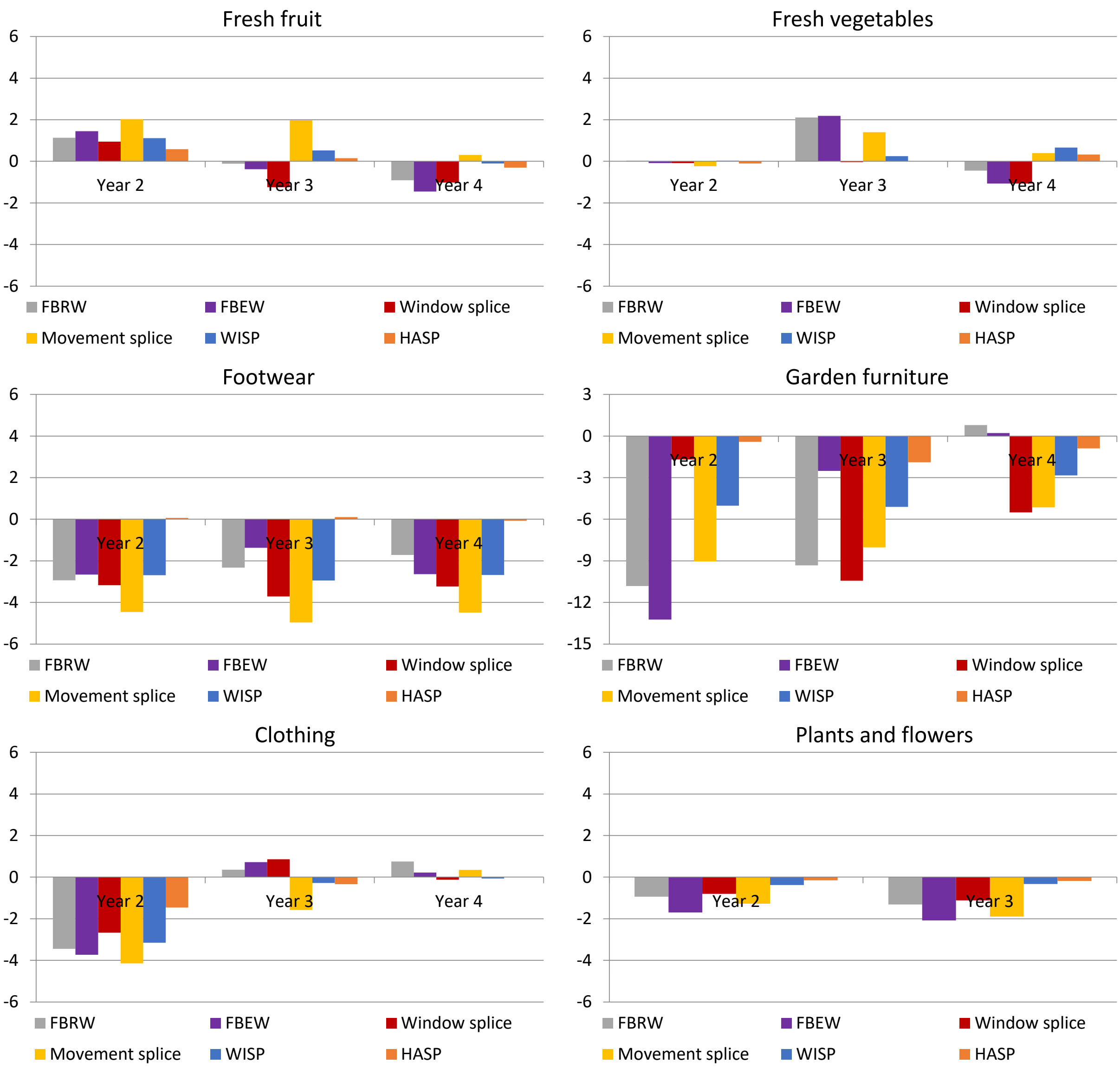

Figure 5.4: Average year on year index differences with the full period indices, in percentage points, per calendar year.

of movement splice and the FBRW method with a 25-month window.

Mean splice

Diewert and Fox [14] proposed the so-called "mean splice" method, which computes an equally weighted geometric mean of the indices obtained by linking on indices in each of the past T months for time windows of length $T + 1$. Also mean splice produces index series that are not free of drift. The classical version of this method links on recalculated indices, which leads to drift for the same reason as for classical window splice. One could then consider a version of mean splice with the linking done on published indices. But also this method is not free of drift, which can be easily seen as the year on year indices will never equal the year on year indices of the transitive indices on the rolling windows.

The mean splice version with the linking on published indices was applied to plants and flowers using a 13-month window. Plants and flowers were chosen because WISP performs very well in this case and therefore serves as a benchmark for mean splice. The extent of drift in mean splice could be alleviated by linking onto published indices instead of the classic linking on recalculated indices. The results are shown in Figure 5.5.

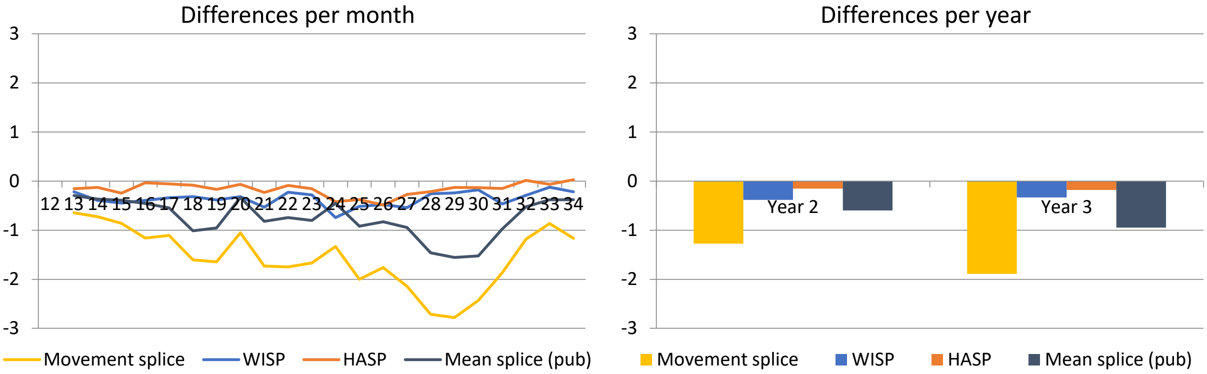

Figure 5.5: Year on year index differences, in percentage points, between mean splice and the full period index for plants and flowers, compared with three other extension methods.

In spite of linking index series of successive windows on published indices, the results in Figure 5.5 show that mean splice performs less well than WISP and HASP. This is in line with our expectations of this method with regard to drift. Moreover, mean splice is a function of the index obtained with movement splice. As we have argued and demonstrated with empirical results that movement splice may lead to considerable drift, part of this behaviour will obviously also affect the indices calculated with mean splice. This can be clearly noted from the results in Figure 5.5, which show that the year on year index differences for mean splice lie between the differences for window splice on published indices and movement splice. In addition, the mean splice method is much more time consuming than other extension methods, approximately by a factor 10 according to our calculations.

Using a longer window in other methods

At this stage in this study, a 25-month window has only been used in the method HASP. It is interesting to use this longer window also in extension methods for which a 13-month window is used. Movement splice and the FBRW method were also applied with a 25-month rolling window on a part of the data sets. The obvious question is how the use of this longer window in these two methods influences the results. The impact on the year on year indices of the two extension methods are compared with HASP in Figure 5.6.

The results are shown from the third year, which is the year in which index extensions take effect for each of the three methods. The results for movement splice (MS-25) and for the FBRW method clearly improve when using a 25-month window over a 13-month window, which can be noted by comparing the results in Figure 5.6 with the results in Figure 5.3 on page 29.

However, movement splice still shows drift, which is still very large for garden furniture. This drift is not surprising, since the analysis of the year on year indices in Section 3.3.1

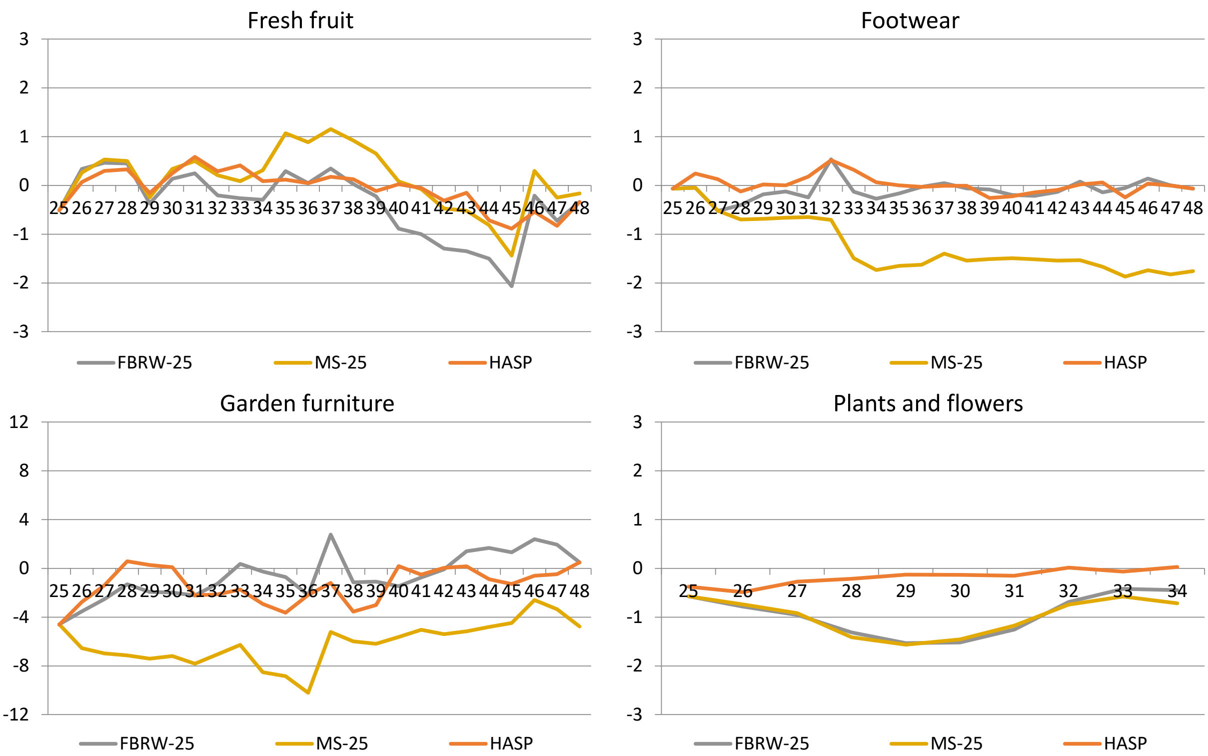

Figure 5.6: Year on year index differences, in percentage points, between three extension methods that use a 25-month window and the full period indices for four COICOPs, with the full period indices subtracted from the indices of the extension methods.

points out that index series produced with movement splice lead to drift irrespective of the length of the time window (see expression (3.11) and the corresponding discussion).

The results for the FBRW method show substantial improvements over the results with a 13-month window shown previously in Figure 5.3. In particular the improvements achieved for footwear and garden furniture are impressive. The FBRW-25 method is the only method that comes close to HASP, which still gives the overall best results.

The results in this study clearly indicate that time windows of 25 months are needed in order to obtain systematically good results for seasonal items. The necessary condition for a drift-free index series as stated in the definition on page 22 is met when linking on published indices as in the fixed base method FBRW and in HASP. The sets W_i mentioned in this definition should be larger than 13-month intervals, and this study suggests that 25-month intervals seem to be an appropriate choice.

5.3 Comparison of QU-GK and GEKS-Walsh

Index extension methods were also applied to index series calculated with the GEKS-Walsh method. Also the results for this multilateral method show that the HASP method gives very accurate indices. At this point, the question is how the results for the two multilateral methods compare when using the HASP method. The price indices for the QU-GK and GEKS-Walsh method are shown in Figure 5.7.

The graphs show that the indices for the two methods follow each other quite closely, but less well for garden furniture. The GEKS method applies equal weights to the time paths that are used to calculate an index between two periods. This means that all time paths are considered equally important, regardless of the expenditures that are carried by each time path. This may have a levelling effect on an index. The comparisons in Figure 5.7 show that the GEKS indices have somewhat lower peaks and higher troughs than the QU-GK indices, apart from certain periods for some COICOPs. It is nevertheless interesting to point out that this version of the GEKS did not make use of price filters for clearance prices, which confirms our claim made at the end of Section 2.2 that the GEKS-Walsh is a better choice than the GEKS-Törnqvist and the GEKS-Fisher.

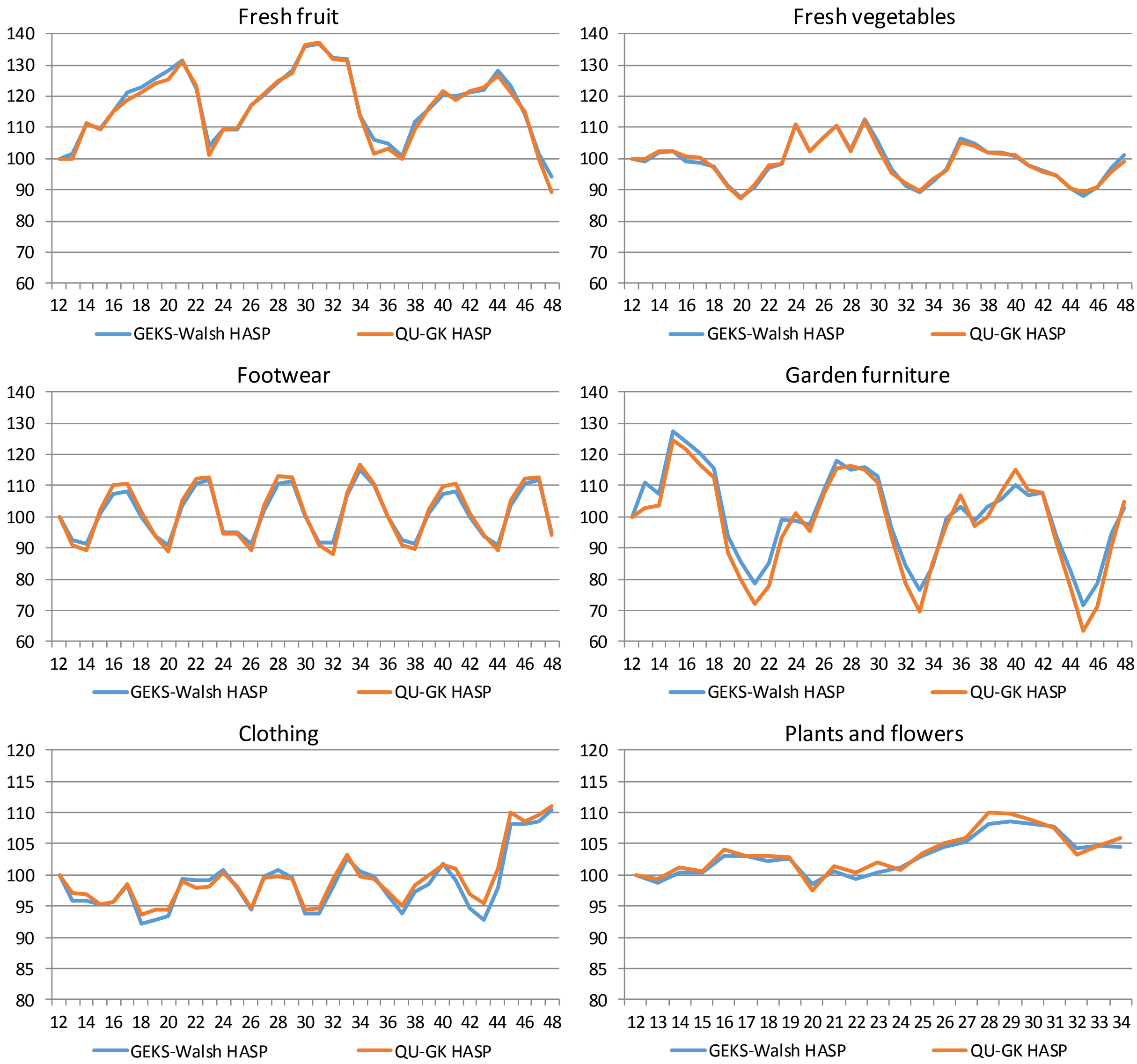

Figure 5.7: Price indices for QU-GK and GEKS-Walsh with HASP as extension method.

The differences between the year on year indices for the QU-GK and GEKS-Walsh method are shown in Figure 5.8. The yearly average differences are several tenths of a percentage point at most; the difference is only larger for garden furniture in year 3. The results also show that there is no systematic deviation between the two methods, which becomes clearer when inspecting the differences at a monthly level. The differences in the latter case mostly lie within a band between −2 and +2 percentage points. The differences are larger for garden furniture, with spikes in late autumn. The spikes, which coincide with the troughs in Figure 5.7, could be ascribed to the equal weighting of the time paths in the GEKS method. The equal weighting makes the method less sensitive to large price changes than other methods like the QU-GK method and Time Product Dummy.

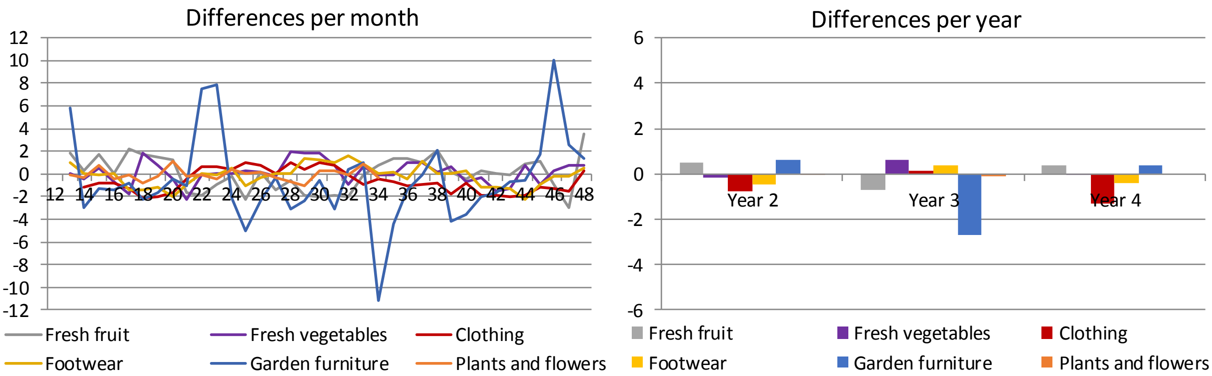

Figure 5.8: Differences between the year on year indices, in percentage points, for QU-GK and GEKS-Walsh with HASP as extension method. The year on year indices for the QU-GK method are subtracted from the GEKS-Walsh indices.

If we compare the differences in Figure 5.8 with the differences in figures 5.3 and 5.4, then, based on this study, we can conclude that the impact of index extension method on a price index is larger than the choice of multilateral method. The present study compared the QU-GK and the GEKS-Walsh method, but recent results showed that also the TPD method gives results that are very similar to the QU-GK method (see Section 9.4 in [11]). This is an important conclusion, since international discussions concerning multilateral methods still seem to focus primarily on the choice of index method.

6 Linking new to old index series in the CPI

6.1 Three linking methods

National statistical institutes are investing significant amounts of resources into innovative research concerning new large data sources and index methods. More than 10 NSIs in Europe have already introduced transaction data in their CPI. Some countries, like Norway and the Netherlands, have a long tradition as they have built up considerable experience with transaction data in their CPI over a period of about 20 years.

The introduction of multilateral methods in the CPI is of more recent date. Statistics Netherlands introduced the GK method in January 2016 for mobile phones and has now expanded the use of this method to more than 30 retail chains in the Dutch CPI and HICP, including all supermarket chains three years ago. Also Belgium (2020) and Norway (2021) are using multilateral methods in their CPI, while other NSIs have concrete plans to follow suit (UK in January 2023 and Luxembourg probably in 2022).

These new developments lead to drastic changes in the CPI and HICP, which may well include differences in the behaviour of price indices compared to current practices. This raises the important and increasingly relevant question how new index series that result from the introduction of new data sources and/or new methods could be linked to the old series. This linking may have considerable effects on year on year and month on month indices of the new series, in particular in the year of introduction.

The index extension problem is directly related to this linking problem. Linking new to old series in the CPI could be achieved by the linking method according to which extension methods operate. For instance, fixed base extension methods link on the indices published in the base month (December of the previous year in practice), while WISP and HASP link year on year indices to published indices of one year ago. This section considers these methods also as potential candidates for linking new to old series in the CPI, and analyses and compares the methods on a number of data sets used in the present study. Three linking methods are examined, which are described below.

We introduce the following notation to formalise the three methods, to describe their differences and to derive their implications with regard to index behaviour after linking. We consider some time interval $[0, T]$, the elements of which denote months, with T representing December of some year. Suppose that a switch to a new method or data source is planned for January of the next year, that is, in month $T+1$. Let $P_{0,t}^{(o)}$ denote the published price indices according to current/old practice in the CPI or HICP and let $P_{0,t}^{(n)}$ represent price indices according to a new method, for months $t = 0, 1, ..., T$. We assume that both series are available for periods of at least 13 months. If the extension method HASP would be introduced in month $T + 1$, then the interval $[0, T]$ typically covers 25 months. The price index that results after linking a new to an old series is denoted by $P_{0,t}$.

Base month linking

The first method links a new series directly to the index published in the base month. The resulting index $P_{0,t}$ can be expressed as follows, which holds for all $t \geq T + 1$:

$$P_{0,t} = P_{0,T}^{(o)} \cdot \frac{P_{0,t}^{(n)}}{P_{0,T}^{(n)}} \tag{6.1}$$

It is easily verified that the month on month indices that result from this method are equal to the month on month indices according to the new index series for all months $t \geq T+1$. This does not hold for the year on year indices in the year in which a new series is introduced. This linking method does not provide any mechanism for controlling the differences between the old and new series in the year before a new series is introduced, which may affect the year on year indices in the introduction year. The year on year indices in all subsequent years will obviously be equal to the indices of the new series.

Year on year linking

The second method produces price indices by linking year on year indices according to the new method to indices published 12 months ago. The resulting price indices can be expressed as follows for all $t \in [T+1, T+12]$:

$$P_{0,t} = P^{(o)}_{0,t-12} \cdot \frac{P^{(n)}_{0,t}}{P^{(n)}_{0,t-12}} \tag{6.2}$$

The same formula applies to each subsequent year, in which case the index $P^{(o)}_{0,t-12}$ is replaced by $P_{0,t-12}$.

It will be clear that year on year indices calculated with this method will be equal to the year on year indices for the new method in each month. This is a big advantage of the method. On the other hand, this method will generally not produce the same month on month indices as the new method. Large differences can be expected in particular for seasonal items, where this linking method is likely to produce short term index behaviour that is more similar to the old series.

Average year correction

The third method can be viewed as an improved version of the method that links onto the index in the base month. Expression (6.1) can be read as the index according to the new method in month t multiplied by a factor that is equal to the ratio of the indices in the base month for the old and new method. In the third method this ratio is replaced by the ratio of the sums (or averages) of the indices for the old and new method over the months in the year before a new method is introduced.

The index published in month $t \geq T+1$ can thus be written as follows:

$$P_{0,t} = \frac{\sum_{\tau=T+1}^{T+12} P^{(o)}_{0,\tau-12}}{\sum_{\tau=T+1}^{T+12} P^{(n)}_{0,\tau-12}} \cdot P^{(n)}_{0,t} \tag{6.3}$$

Price indices according to a new method or data source are thus adjusted by the extent to which indices for the old and new method differ on average in the year before introducing a new method. We therefore call the method that makes use of expression (6.3) an "average year correction method", which we abbreviate to AYCO.

Like the base month linking method, also AYCO produces year on year indices that will differ from the indices of the new series in the introduction year. But the average year correction ratio in (6.3) is able to control the average year on year index in the introduction year. The yearly average year on year index with respect to the old method can be shown to be equal to an expression for the average year on year index that results from the new series (see expression (11.10) on page 141 in [11]). The monthly year on year indices for the resulting extended series can also be expected to be smaller than the indices produced by the base month linking method. The year on year indices in all subsequent years will of course be equal to the indices of the new series.

The month on month indices are equal to the indices for the new series from the year of introduction, except for the first month. This follows directly from (6.3), as the index $P_{0,T+1}/P^{(o)}_{0,T}$ is generally not equal to $P^{(n)}_{0,T+1}/P^{(n)}_{0,T}$. The month on month indices for the base month linking method are equal to the indices for the new series from the first month. The method AYCO in fact sacrifices the month on month index in the first month in order to apply a correction factor that allows to control the average year on year index in the introduction year. This is a big advantage and added value of AYCO, which makes this method superior to base month linking. This will be further examined by applying the three linking methods to a number of transaction data sets.

6.2 Results

The three linking methods were applied to the transaction data sets on fresh fruit and fresh vegetables. The indices for GK-HASP in the third year of the data period were linked to the CPI of the two COICOPs in the preceding year. The data sets are from one of the largest supermarket chains in the Netherlands, while the CPI applies to all chains. Although the results may change when restricting the CPI to the same supermarket chain, the examples are equally instructive, to which we add that the data sets are from a major supermarket chain that has a high weight in the Dutch CPI.

The CPI was computed with the chained Jevons method in the first two years of the data period considered in this study, which also made use of transaction data. The price indices for the three linking methods are shown in Figure 6.1. The indices for base month linking and AYCO exactly follow the GK-HASP indices from the year of introduction (third year), except for the first month of the third year for the method AYCO, as was explained in the previous section. The behaviour of the two linked series exactly matches the expectations stated above. Year on year linking produces different patterns. Although the resulting series for vegetables follows the GK-HASP series quite nicely, year on year linking produces big differences in month on month indices.

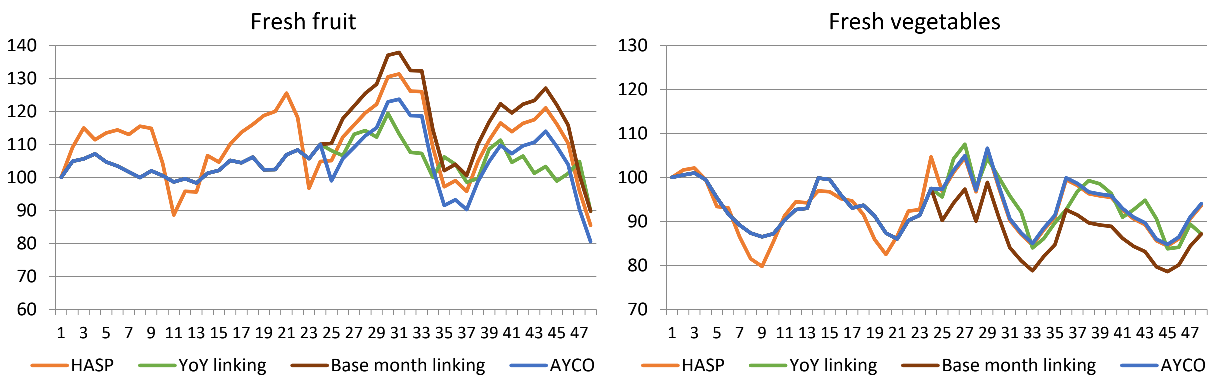

Figure 6.1: Price indices for the three linking methods compared to GK-HASP.

Although the index series obtained with base month linking and AYCO have the same shape from the second month onward after linking, the index values for two series differ considerably, which may result in big differences between their year on year indices. These differences are shown in Figure 6.2. The results for year on year linking are not shown, since this method trivially yields zero differences.

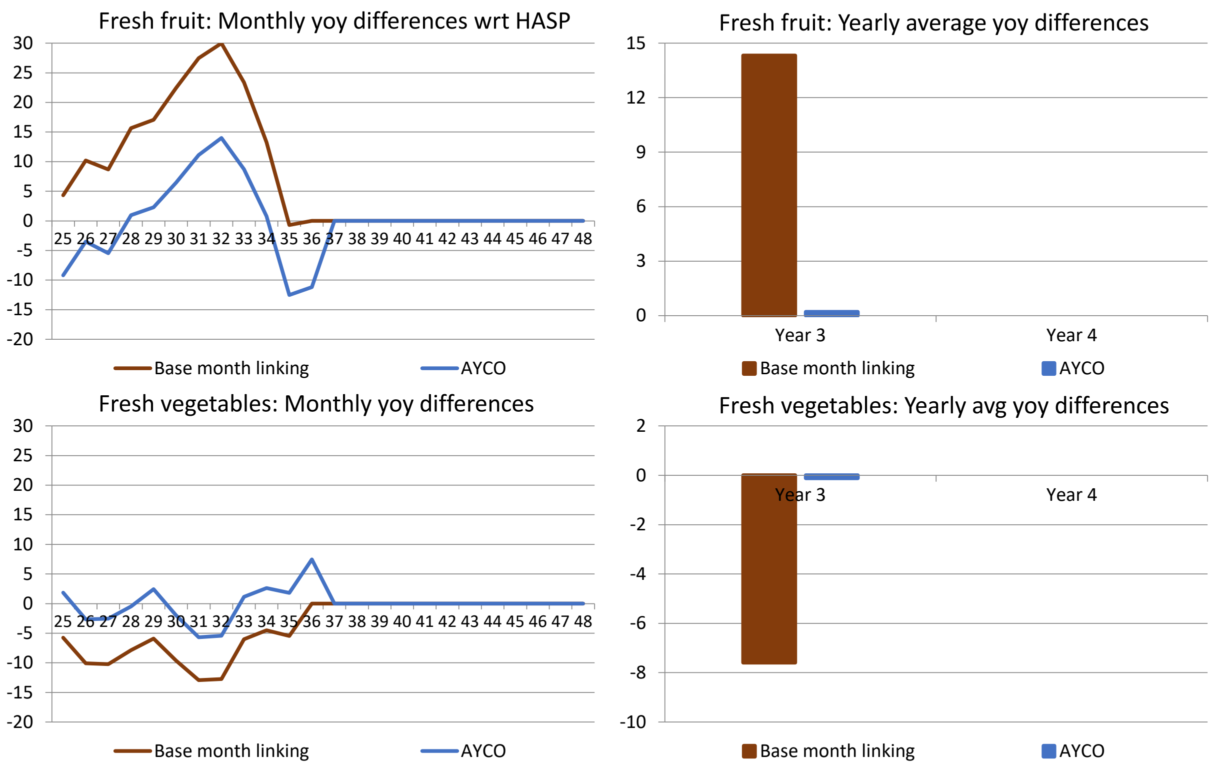

Figure 6.2: Differences (in pp) between the year on year indices of the three linking methods and GK-HASP, with the latter subtracted from the indices of the three linking methods.

The results in these graphs indeed show that base month linking deviates considerably from the year on year indices of GK-HASP. The yearly average differences for the two COICOPs are still very large, the sizes of which could even affect overall CPI. The method AYCO performs much better, as the correction method tends towards evenly distributing positive and negative year on year deviations over the months in the third year. As a result, the average differences over the same year are very close to zero (0.1 percentage point for both fruit and vegetables). The yearly average differences are not equal to zero because the averages in the graphs are calculated with equal weights in each month. The aforementioned identity of AYCO with regard to average year on year indices includes weights which may differ over months. The differences in the successive fourth year are zero for the two linking methods, as would also be expected.

Both the analytical arguments in Section 6.1 and the results in the present section point towards a clear preference of the average year correction method AYCO over base month linking. The question is whether year on year linking or AYCO should be selected. This could be decided on a case by case basis. For instance, for non-seasonal items one could decide for year on year linking, while AYCO would be the preferred method for seasonal items because of possible large differences between the short term index changes in the old and new series.

However, the year on year differences for non-seasonal items can be expected to be relatively small in the introduction year when applying AYCO. If we add to this the elegance of reproducing the new series from the second month, the accuracy of AYCO in the average year on year indices and the transparency of a uniform choice, it is advisable to use AYCO for level corrections in general in the CPI and HICP.

6.3 Other cases of level correction

The method AYCO can be applied to other situations where index level corrections may be desirable. Situations may occur where prices of new products cannot be dealt with appropriately in the first months after their appearance on the market. This may be the case with high introduction prices, which are very hard to conceive as such when calculating indices in real time. It is only when prices and quantities from additional months are available that we can say with more certainty whether or not products were introduced at higher prices than in successive months.

A simple way to establish this is by comparing the index series of the most recent time window with the published index series. Notable differences between the two series can be taken as a clear indicator of impactful events. The method HASP links year on year indices to published indices of one year ago. This linking may cause anomalies in an index series in successive years when singular drastic events took place in a previous year.

This may not necessarily affect year on year indices, but it may be desirable to apply an index level correction in order to remove possible anomalies in the year following an influential event. This can be done by applying AYCO, in which case the correction term in expression (6.3) can be derived from the published indices and the indices in the second year of the most recent 25-month window. In this application of AYCO a shift of the initial 25-month window is in fact applied.

On the other hand, the numerous applications of index extension methods to tens of transaction data sets that can be found in the previous studies [10] and [11], and in the present study, show small differences between indices obtained with HASP and the full period indices. This suggests that the occurrence of impactful events and the possibility of applying level corrections may be quite rare.

7 Conclusions

The rapid developments that have taken place in the CPI on a global scale over the past five years call for a greater need of guidance to inform NSIs about methodological choices for processing transaction data and web scraped data. The transition towards these data sources has considerably expanded the range of possible choices about index methods and related aspects. It is therefore important to set up criteria for comparing and assessing methods, but also to gain greater insight into the effect of different choices on price indices.

Research carried out to date has shown that the majority of the studies and discussions with regard to index methods still primarily focus on index formula. The previous comparative study on index extension methods already showed that the choice of extension method has a big impact on price indices [10]. The present study on seasonal items confirms this finding, and shows that the choice of index extension method has a bigger impact than the choice of index formula, at least, when index formulas are confined to multilateral methods that are suited for processing transaction data.

As research and applications of multilateral methods in the time domain were initiated only in the last decade, also the time-driven problem of extending index series from one period to the next is a topic with a short history. This paper offers an attempt towards characterising and formalising the index extension problem and methods. Index formulas of extension methods, including their year on year indices, are derived and an assessment of these methods is given in terms of their behaviour with regard to drift and other index-theoretical properties. An alternative definition of drift-free index series has been proposed, since the requirement of transitivity is too strong for extended index series.

The analytical properties and the results of the comparative empirical study presented in this paper and in previous studies ([10], [11]) lead to the following main conclusions:

- Index extension methods may produce very different index series, in particular for seasonal items.
- The method HASP, which uses a 25-month window and continues index series over time by linking year on year indices on published indices of one year ago, has given the best results. The superiority of HASP is particularly evident for seasonal items.
- Extension methods that make use of 13-month windows are very sensitive to drift for seasonal items, which is of a predominantly downward nature. Year on year indices may easily be lower by several percentage points compared with HASP and full period transitive indices.
- The average year level correction method AYCO is clearly superior to base month linking, which finds both analytical and empirical support. The advantage of AYCO over year on year linking is that AYCO reproduces the new index series from the second month onward.

The only method that comes close to HASP is the fixed base FBRW method when used with a 25-month rolling window. The year on year indices of this FBRW version are slightly less accurate than the HASP figures. A big advantage of HASP is that inflation figures calculated from each 25-month window will also be the published figures in every month. This holds when applying multilateral methods at COICOP-5 level, which should absolutely be encouraged also in order to make higher aggregate inflation figures less dependent upon dated yearly fixed weights below COICOP. The 25-month FBRW and HASP give the same indices in every base month, so that considering FBRW as an alternative to HASP will hardly generate any added value.

The method HASP makes use of a 25-month window, which means that at least 25 months of data have to be available in order to use the method in the CPI. Situations may occur where NSIs have historical data covering a shorter period, which they may want to introduce in their CPI and HICP without waiting until 25 months of backdata are available. Recent work shows that it is also possible to use HASP in situations with less backdata. This can be achieved by imputing prices in all months for which no data are available in the initial 25-month window. Nonzero quantities need to be assigned to the imputed prices when using the QU-GK or GEKS-Walsh method. It is recommended to set these 'imputed' quantities at values close to zero, so that the adjustment factors

ν_i in the GK method and the indices in the data period will hardly be influenced by the imputed prices. More details can be found in [13].

A field that is currently going through phases of increasing complexity like index theory should welcome any effort that contributes to shared practices across countries. The findings from different comparative studies on index extension methods and related topics could provide opportunities worth considering in this respect. Recommendations include the use of a 25-month rolling window in multilateral methods, the method HASP for extending index series over time and the method AYCO for level corrections when introducing new data sources and/or index methods in the CPI and HICP.

These recommendations naturally link up with the application of the product stratification method MARS to calculate values for product match or continuity 12 months after the comparison period by using rolling windows, which would thus proceed in parallel with HASP. This is an area of ongoing research.

Acknowledgements

This research was funded from Eurostat grant IMCOPS 847540–2018-NL-PRICE, for which the author expresses his gratitude. The author also wants to thank various colleagues at Statistics Netherlands and of other statistical offices and international organisations for the numerous stimulating discussions on this topic.

References

[1] ABS (2017). Making Greater Use of Transactions Data to Compile the Consumer Price Index. Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, 10-12 May 2017, Eltville am Rhein, Germany. Available at https://www.bundesbank.de/en/homepage/making-greater-use-of-transactions-data-to-compile-the-consumer-price-index-australia-635722

[2] Auer, L. von (2014). The Generalized Unit Value Index Family. *Review of Income and Wealth*, 60, 843-861. https://doi.org/10.1111/roiw.12042

[3] Balk, B.M. (1981). A Simple Method for Constructing Price Indices for Seasonal Commodities. *Statistische Hefte*, 22, 72-78. https://doi.org/10.1007/bf02932820

[4] Barb, F.D., Chessa, A.G., Heuvel, E.G. van den, Hoebergen, A., Ketelaar, K., Ooms, L., Verburg, J., and Willenborg, L. (2020). Impact Analysis of Transactions Data Stratification and Classification on HICP Quality. Report written for Eurostat grant project IMPANQ 801915–2017-NL-PRICE, Statistics Netherlands.

[5] Białek, J., and Bobel, A. (2019). Comparison of Price Index Methods for the CPI Measurement Using Scanner Data. Paper presented at the *16th Meeting of the Ottawa Group on Price Indices*, 8-10 May 2019, Rio de Janeiro, Brazil. Available at https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/bialek_bobel_paper_2.pdf

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

[7] Chessa, A.G., Verburg, J., and Willenborg, L. (2017). A Comparison of Price Index Methods for Scanner Data. Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, 10-12 May 2017, Eltville am Rhein, Germany. Available at https://www.bundesbank.de/en/homepage/a-comparison-of-price-index-methods-for-scanner-data-635922

[8] Chessa, A.G. (2018). *Product Definition and Index Calculation with MARS-QU: Applications to Consumer Electronics*. Report written for a Eurostat Grant agreement, Statistics Netherlands.

[9] Chessa, A.G. (2019). MARS: A Method for Defining Products and Linking Barcodes of Item Relaunches. Paper presented at the *16th Meeting of the Ottawa Group on Price Indices*, 8-10 May 2019, Rio de Janeiro, Brazil. Available at https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/product_definition_with_mars_chessa_og19.pdf

[10] Chessa, A.G. (2019). A Comparison of Index Extension Methods for Multilateral Methods. Paper presented at the *16th Meeting of the Ottawa Group on Price Indices*, 8-10 May 2019, Rio de Janeiro, Brazil. Available at https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/index_extension_methods_compared_chessa_og19_1.pdf

[11] Chessa, A.G., Nan, L., Barb, F., Link, K., and Ketelaar, K. (2020). *Towards a Comprehensive Methodology for Index Calculation in the CPI with Transaction Data*. Report written for Eurostat grant project IMCOPS 847540–2018-NL-PRICE, Statistics Netherlands.

[12] Chessa, A.G. (2021). MARS: A Method for Defining Products and Linking Barcodes in Transaction Data. To appear in the *Journal of Official Statistics*.

[13] Chessa, A.G. (2021). Index calculation with GK-HASP with limited backdata. Presentation prepared for the *Eurostat Task Force on Multilateral Methods*, 25 February 2021. Presentation is uploaded to CIRCABC: EUROPA > European Commission > CIRCABC > Eurostat > Price Statistics.

[14] Diewert, W.E., and Fox, K.J. (2017). Substitution Bias in Multilateral Methods for CPI Construction Using Scanner Data. Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, 10-12 May 2017, Eltville am Rhein, Germany. Available at https://www.bundesbank.de/en/homepage/substitution-bias-in-multilateral-methods-for-cpi-construction-using-scanner-data-635978

[15] Eltetö, Ö., and Köves, P. (1964). On an Index Computation Problem in International Comparisons. *Statiztikai Szemle*, 42, 507-518. [in Hungarian]

[16] Geary, R.C. (1958). A Note on the Comparison of Exchange Rates and Purchasing Power between Countries. *Journal of the Royal Statistical Society A*, 121, 97-99. https://doi.org/10.2307/2342991

[17] Gini, C. (1931). On the Circular Test of Index Numbers. *Metron*, 2, 3-24.

[18] de Haan, J. (2015). A Framework for Large Scale Use of Scanner Data in the Dutch CPI. Paper presented at the *14th Meeting of the Ottawa Group on Price Indices*, 20-22 May 2015, Urayasu City, Japan. Available at https://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf

[19] de Haan, J., and van der Grient, H.A. (2011). Eliminating Chain Drift in Price Indices Based on Scanner Data. *Journal of Econometrics*, 161, 36-46. https://doi.org/10.1016/j.jeconom.2010.09.004

[20] de Haan, J., Willenborg, L., and Chessa, A.G. (2016). An Overview of Price Index Methods for Scanner Data. Paper presented as "room document" at the *Meeting of the Group of Experts on Consumer Price Indices*, 2-4 May 2016, Geneva, Switzerland. Available at 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

[21] ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004). *Consumer Price Index Manual: Theory and Practice*. Geneva: ILO Publications. https://doi.org/10.5089/9787509510148.069

[22] Ivancic, L., Diewert, W.E., and Fox, K.J. (2011). Scanner Data, Time Aggregation and the Construction of Price Indexes. *Journal of Econometrics*, 161, 24-35.

[23] Khamis, S.H. (1972). A New System of Index Numbers for National and International Purposes. *Journal of the Royal Statistical Society A*, 135, 96-121. https://doi.org/10.2307/2345041

[24] Krsinich, F. (2014). The FEWS Index: Fixed Effects with a Window Splice – Non-revisable Quality-adjusted Price Indices with no Characteristic Information. Paper presented at the *Meeting of the Group of Experts on Consumer Price Indices*, 26-28 May 2014, Geneva, Switzerland. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/New_Zealand_-_FEWS.pdf

[25] Lamboray, C. (2017). The Geary Khamis Index and the Lehr Index: How Much Do They Differ? Paper presented at the *15th Meeting of the Ottawa Group on Price Indices*, 10-12 May 2017, Eltville am Rhein, Germany. Available at https://www.bundesbank.de/en/homepage/the-geary-khamis-index-and-the-lehr-index-how-much-do-they-differ--636014

[26] Summers, R. (1973). International Price Comparisons Based upon Incomplete Data. *Review of Income and Wealth*, 19, 1-16. https://doi.org/10.1111/j.1475-4991.1973.tb00870.x

[27] Szulc, B. (1964). Index Numbers of Multilateral Regional Comparisons. *Przeglad Statysticzny*, 3, 239-254. [in Polish]

[28] Van Loon, K., and Roels, D. (2018). Integrating Big Data in the Belgian CPI. Paper presented at the *Meeting of the Group of Experts on Consumer Price Indices*, 7-9 May 2018, Geneva, Switzerland. Available at https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Belgium.pdf

[29] Willenborg, L. (2017). *Transitivizing Elementary Price Indexes for Internet Data using the Cycle Method*. Discussion Paper, Statistics Netherlands.

[30] Willenborg, L. (2017). *From GEKS to Cycle Method*. Discussion Paper, Statistics Netherlands.

[31] Zhang, L.C., Johansen, I., and Nygaard, R. (2019). Evaluating Unit-Value Price Indices in a Dynamic Item Universe. Paper presented at the *16th Meeting of the Ottawa Group on Price Indices*, 8-10 May 2019, Rio de Janeiro, Brazil. Available at https://eventos.fgv.br/sites/eventos.fgv.br/files/arquivos/u161/statistics_norway_paper.pdf

List of Figures

List of Tables