A comparison of index extension methods for multilateral methods

Antonio G. Chessa[1]

Abstract

Multilateral methods have gained a lot of attention in the past five years, as national statistical institutes (NSIs) are getting wider access to transaction (scanner) data. Multilateral methods can be used to calculate transitive index series and have other advantages, such as processing complete large data sets and including new products in the first sales period.

The choice of index formula is just one aspect from a broad spectrum of decisions that have to be made when compiling price indices. Multilateral indices are transitive on a fixed time interval, but an essential question is how the drift-free properties can be preserved when data from the next period become available. The time window has to be adjusted in order to include new data. The indices calculated on the adjusted window may differ from the previously calculated indices, which, however, cannot be revised in the CPI.

This paper presents the results of a comparative study of extension methods for index series that are known to this date: splicing methods and methods that use a fixed base period. Price indices are calculated for transaction data of a supermarket chain, pharmacy products and a department store chain. The results show significant drift for window splice and movement splice even at aggregate chain level. The fixed base methods are drift-free by definition and therefore perform much better than the splicing methods.

Alternative splicing methods are proposed, which link year on year indices of rolling windows to published indices instead of following the classical approach of linking on recalculated indices. The alternative methods have no drift over the length of the time window. The results for window splice and half splice, with a 25-month window in the latter case, are indeed much better than for classical splicing. The half splice method in particular emerges as a very promising, accurate and stable method.

Keywords: CPI, multilateral methods, index extension, splicing, transitivity, chain drift.

1 Introduction

Electronic transaction data sets contain expenditures and quantities sold of items purchased by consumers at physical or online sales points of a retail chain.[2] Sales data are often aggregated by retailers to weekly sales and are specified by the product IDs of items, such as the Global Trade Item Number (GTIN, barcode).[3] Transaction data sets also contain characteristics, such as brand and package volume, of the items sold. While traditional price collection methods typically record prices of several tens of products in shops, electronic transaction data sets may contain several tens of thousands of items at the GTIN level for a single retail chain.

---

[1] CPI department, Statistics Netherlands. Email: ag.chessa@cbs.nl. The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

[2] Statistics Netherlands uses the broader term 'transaction data' instead of 'scanner data', which will also be done in this paper. Scanner data are in fact transaction data specified by barcode.

[3] The term 'item' is used interchangeably with product ID (e.g. GTIN) in this paper.

The large numbers of items purchased by consumers and the availability of sales values and quantities have contributed significantly to the increasing popularity of transaction data at national statistical institutes (NSIs). The use of transaction data in the Consumer Price Index (CPI) has increased rapidly over the past few years. Until 2014, four countries in Europe used transaction data in their CPI, which increased to ten by January 2018. Australia and New Zealand are also using transaction data in their CPI since several years.

Transaction data sets offer possibilities to NSIs to enhance the quality of index numbers. More refined methods can be applied that deal with the dynamics of consumption patterns in a more appropriate way than traditional fixed-basket methods. Multilateral methods can be used to specify monthly weights based on actual sales at product level and new products can be directly included in index calculations. Large transaction data sets can be processed to their full extent. A big advantage of multilateral methods is that transitive index series can be obtained while incorporating the aforementioned dynamics. It is therefore not unsurprising that NSIs and researchers from academia have increasingly focused on multilateral methods in recent years (Ivancic et al., 2009, 2011; de Haan and van der Grient, 2011; Krsinich, 2014; Chessa, 2016; Chessa et al., 2017; ABS, 2017; Diewert and Fox, 2017; Van Loon and Roels, 2018).

The choice of index formula is one aspect of a broad spectrum of choices that have to be made in order to compile index numbers. Each choice aspect may have an impact on the index, which can be quite large (Chessa, 2016, 2017, 2018; Chessa et al., 2017). Products have to be defined before prices and price indices can be calculated and weighting schemes have to be decided upon, beside index formula. Two additional choices have to be considered when using multilateral methods: the length of the time window on which index series are calculated and how to continue an index series when data of the next month become available.

This paper deals with the latter choice aspect. The extension of a series is in fact essential to all time series modelling. In the specific case of the CPI, the problem is that indices published in previous periods cannot be revised. Multilateral methods can be used to calculate transitive index series on an arbitrarily fixed time interval. The interval must be adapted in some way in order to accommodate newly available data. The index series calculated on the adapted interval may differ from the published indices. The question is how the index series of the new window can be linked to the current series, such that the continued series of published indices behaves 'well', in the sense that it does not drift from some transitive 'benchmark index'.

Different methods have been proposed in the present decade for continuing index series, which can be broadly subdivided between 'splicing methods' (de Haan and van der Grient, 2011; Krsinich, 2014; Diewert and Fox, 2017) and 'fixed base methods' (Chessa, 2016). Section 2 gives an overview of the current methods. These methods are included in a broad comparative study, which is presented in Section 3. Price indices are calculated for different extension methods by making use of transaction data of three Dutch retail chains: a supermarket chain, a pharmacy store chain and a chain of department stores. The results are discussed in Section 4. Alternative splicing methods are suggested following the analysis and subsequently applied to the data sets. Section 5 concludes.

2 Methods for extending index series

Multilateral methods were originally introduced to compare price levels across countries or regions. These methods yield transitive price comparisons, which is a desirable property since the results are independent of the choice of base country. Well-known methods are the GEKS method (Gini, 1931; Eltetö and Köves, 1964; Szulc, 1964), the Geary-Khamis method (Geary, 1958; Khamis, 1972), and the Country Product Dummy method proposed by Summers (1973). For details on the various methods, see Balk (1996, 2001), chapter 7 in Balk (2008), Diewert (1999) and Deaton and Heston (2010).

The set of countries is fixed during the comparison. When applying multilateral methods in a temporal setting (first considered by Balk (1981)), the months in a time window will change as soon as information of the next month becomes available, as it has to be included in the time window. The transitive index series of successive windows have to be linked in some way in order to produce a series of published, non-revisable price indices. Different methods have been proposed, which are described in this section. First, a general characterisation of index extension methods is given. From the available literature on this topic it is not always easy to identify the different choices that have to be made in order to specify and implement extension methods. This paper tries to make these choices explicit.

In the rest of this paper a month is taken as time period. Once the length of the time window is chosen, three choices are made 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 for each of these three aspects, which are described below. The first two choice aspects are explicitly stated in different contributions. However, there is a third element (C), which is hardly emphasised, but appears to play a crucial role when analysing the differences and the behaviour of extension methods, as we will see in sections 3 and 4.

Window adjustment

Two methods have been proposed in the literature:

A1. A 'rolling' or 'moving' window. This window keeps its full length when it is shifted each month to include data from the next month;
A2. A monthly expanding window. The window is extended by one month each time monthly data become available. A fixed base month is used, for instance, December of the previous year. So, the window consists 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 change in the current month with respect to 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:

B1. A moving linking month;
B2. A fixed month.

Index in the linking month

Linking index series of subsequent windows generates a series of published indices. But indices are also recalculated after adjusting a window. Some extension methods link onto a 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:

C1. A recalculated index;
C2. A published index.

Index extension methods

Some widely used index extension methods are described below. Index extension methods can be subdivided into two main classes:

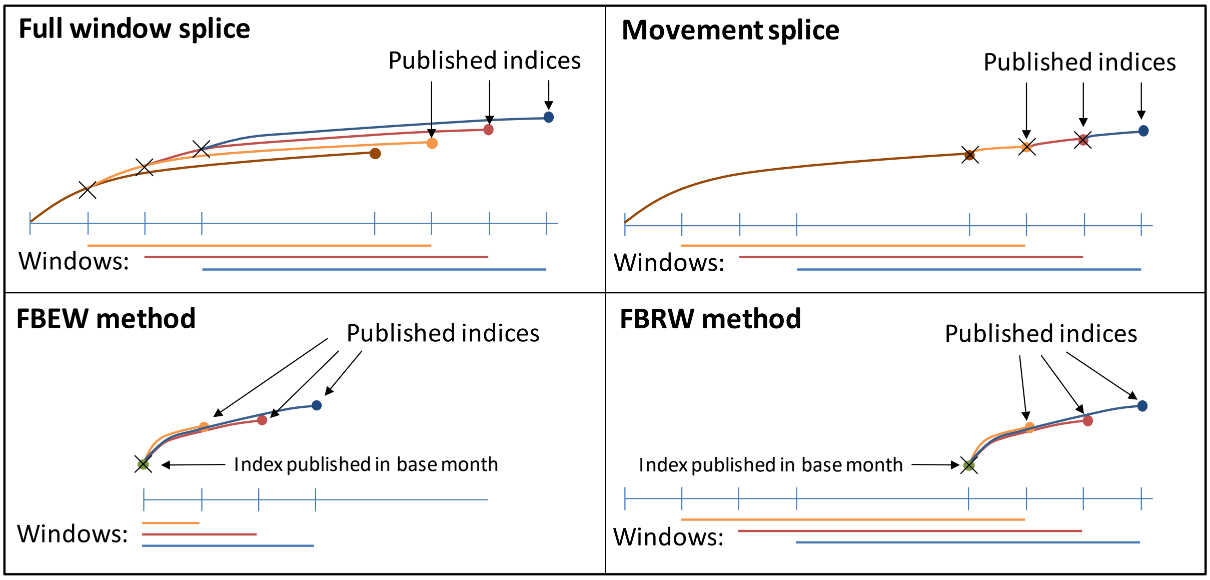
1. Splicing methods. These methods all use a rolling window and a moving linking month. Two well-known methods are:
    a. Window splice. This is a type A1-B1-C1 method and was proposed by Krsinich (2014). The first month of the adjusted window is used as linking month. By this property, the method is sometimes also called 'full window splice'. The most recent recalculated index is used as linking index, that is, of the most recently linked series.
    b. Movement splice. This is a type A1-B1-C2 method and was proposed by de Haan and van der Grient (2011) as part of the Rolling Year GEKS (RGEKS or RYGEKS). The penultimate month of the adjusted window is taken as linking month, and the month on month index of the adjusted window is chained to the published index of the previous month.
2. Fixed base methods. These methods use a fixed base month as linking month, which is usually December of the previous year. The published index in the base month is the linking index. We will consider two versions of this type of method:
    a. Fixed base expanding window (FBEW). This method is of type A2-B2-C2 and was proposed by Chessa (2016). A direct index is calculated with respect to the base month. The index in December of the present year will be equal to the index of the transitive series when linking this series to the base month, for a 13-month window. This implies that the series of published indices will be free of chain drift by construction, at least, in a piecewise sense.
    b. Fixed base rolling window (FBRW). This method is of type A1-B2-C2. The only difference with the FBEW method is that in this case a rolling window is used. Also this method is free of chain drift, for the same reasons as the FBEW method.

These methods, which are illustrated in Figure 1, will be applied to transaction data in Section 3.

Of course, other extension methods can be constructed by making different choices with regard to window adjustment, linking month and/or linking index. For instance, choosing the central month in adjusted time windows (with an odd number of months) as linking month in a splicing method results in the 'half splice' method (de Haan, 2015). The 'mean splice' method calculates an index for the current month as an unweighted geometric mean of the indices that are obtained by taking each month of the adjusted window as linking month (the current month is obviously excluded). This method has been suggested in Diewert and Fox (2017).

European regulations concerning index compilation prescribe the use of a fixed base month. This contributed to suggesting and investigating the suitability of monthly expanding windows, as proposed by Chessa (2016). The more elegant fixed-length rolling windows can also be combined with a fixed base month. But the novelty of multilateral methods in the CPI, together with the choice of using the base month both as the first month in the time window and for index calculations, was thought to contribute to transparency and simplicity at the time when a multilateral method was introduced in the Dutch CPI (January 2016).

Figure 1. Illustration of index extension methods, with × denoting linking month and index.



3 Comparative study

3.1 Transaction data sets

The extension methods described in the previous section are applied in the present section to transaction data sets of three Dutch retail chains: a supermarket chain, a department store chain and a chain of pharmacy stores. The department and pharmacy store data were already used in previous studies (Chessa, 2016, 2017); the data used in this study cover periods of 47 and 43 months, respectively. The supermarket data cover a period of 48 months. The supermarket and pharmacy store data have January as initial month, while the department store data sets have February as starting month.

In this study, a total of 65 product categories are defined, which are distributed over 6 COICOP divisions. For the three chains we have the following numbers of categories and COICOP divisions:

- For the supermarket chain, 11 product categories are defined that belong to COICOP 01, 05 (detergents) and 12 (personal care products);
- The data used for the pharmacy store chain are distributed over 20 product categories and 2 COICOP divisions (06 and 12);
- The department store chain has a broad assortment. In this case, 34 product categories are defined, which are distributed over 5 COICOP divisions (01, 03, 05, 11 and 12).

Product categories can be COICOPs or lower aggregates. For example, the COICOPs in division 01 are more refined than in division 03, so that COICOPs like bread, fruit, vegetables, meat, coffee and tea are taken as product categories in this study. The department stores sell men's and ladies' clothing. Because of the large variation in items, these COICOPs are subdivided into smaller product categories (e.g. underwear, socks, T-shirts and sweaters for menswear). Refined product categories are also defined for toiletries and beauty items sold by the pharmacy stores.

For example, toiletries are subdivided into seven categories: hair care, oral care, shaving care, razors, soap products, bath and shower products, and body care products for infants.

3.2 Index comparisons and results

Price indices are calculated with the Geary-Khamis (GK) method. Rates of churn differ quite a lot across product categories, so that GTINs are not always suitable as products. Before calculating indices, products were therefore defined for each of the 65 categories. Applications with the method MARS in a previous study already pointed out that GTINs are suitable as products for COICOP 01, while GTINs are grouped into products by common characteristics for clothing and pharmacy items (Chessa, 2019). Similar choices are made in the present study:

- GTINs are chosen as products for COICOP 01 and also for non-clothing items sold by the department stores;
- Products are defined in terms of sets of product characteristics for clothing and pharmacy items. For example, hair care products are characterised by brand, package volume and hair type (normal, dry, damaged, etc.).

In the latter case, unit values of products are calculated by summing expenditures and numbers of items sold over the GTINs that are assigned to the same product and by dividing these two aggregate figures.

The extension methods applied in this section will also be applied in combination with a second multilateral method (the TPD method) in order to assess the possible impact of the choice of index method on the differences between extension methods. The following choices and comparisons are made for the extension methods:

- Time windows of 13 months are used in each extension method;
- The fixed base extension methods FBRW and FBEW are applied with December as base month. The FBEW method thus reaches its maximum length of 13 months in December;
- Index extension with the FBEW method is also carried out in the first calendar year. The effects of index extension for the other three methods take shape after 13 months;
- Price indices obtained with the extension methods are compared with two index series: (i) indices that are transitive on 13-month windows, with December as base month, and (ii) indices that are transitive on the full length of the time period covered by the data sets.

The four index extension methods are compared in a series of graphs shown on the following pages. Comparisons are made at retail chain level. Two aggregate price indices are shown for the department store chain: one for clothing and one for non-clothing items. The price indices of the product categories calculated with the GK method are combined with yearly fixed weights to indices for the higher aggregates by applying traditional Laspeyres type methods. The impact of the index extension methods at product category level is also shown in a set of graphs.

The indices for the splicing methods show substantial differences with the transitive 'benchmark' indices for 13-month windows, even at chain level (Figure 2). The results in Table 1 show that the differences between the year on year indices for the window splice method are close to one percentage point for the supermarket chain in the third and fourth year, averaged over all months. This is a big deviation at such a high level of aggregation. The movement splice method lies above the piecewise transitive index for supermarkets. It shows smaller absolute differences than the window splice method, but the differences are still large. The movement splice method also deviates substantially at aggregate level for clothing. Only the differences for the third and fourth year are shown in Table 1, since the full effect of index extension with 13-month windows is obtained from the third year.

Figure 2. Price indices for window splice and movement splice at highest aggregate level, compared with the transitive indices on 13-month windows.

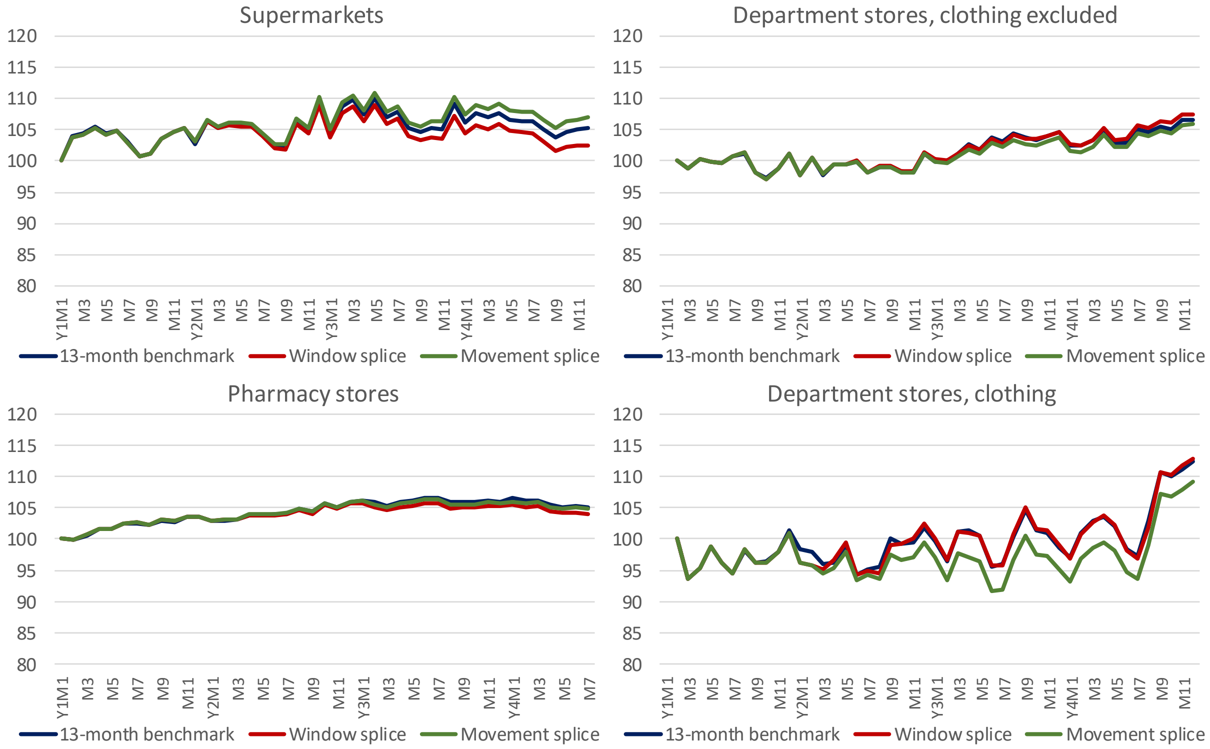


Figure 3. Price indices for the fixed base extension methods at highest aggregate level, compared with the transitive indices on 13-month windows.

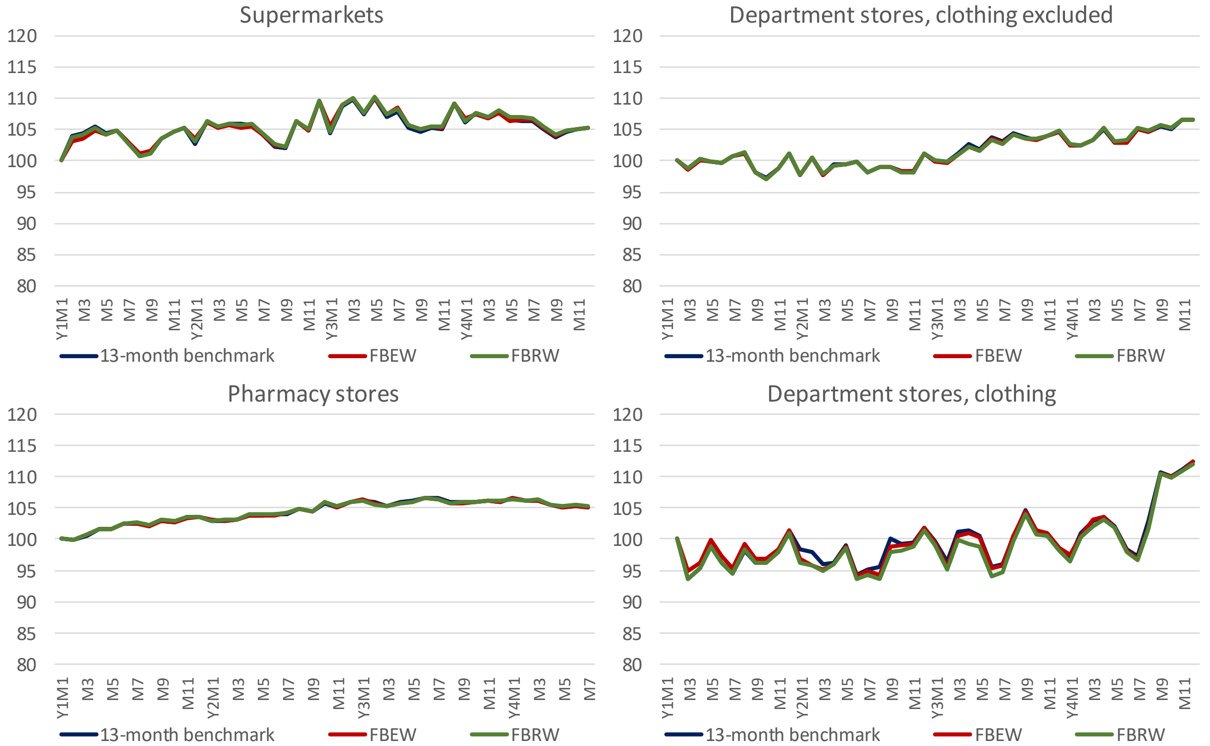


Figure 4. Price indices for window splice and movement splice for four product categories, compared with the transitive indices on 13-month windows.

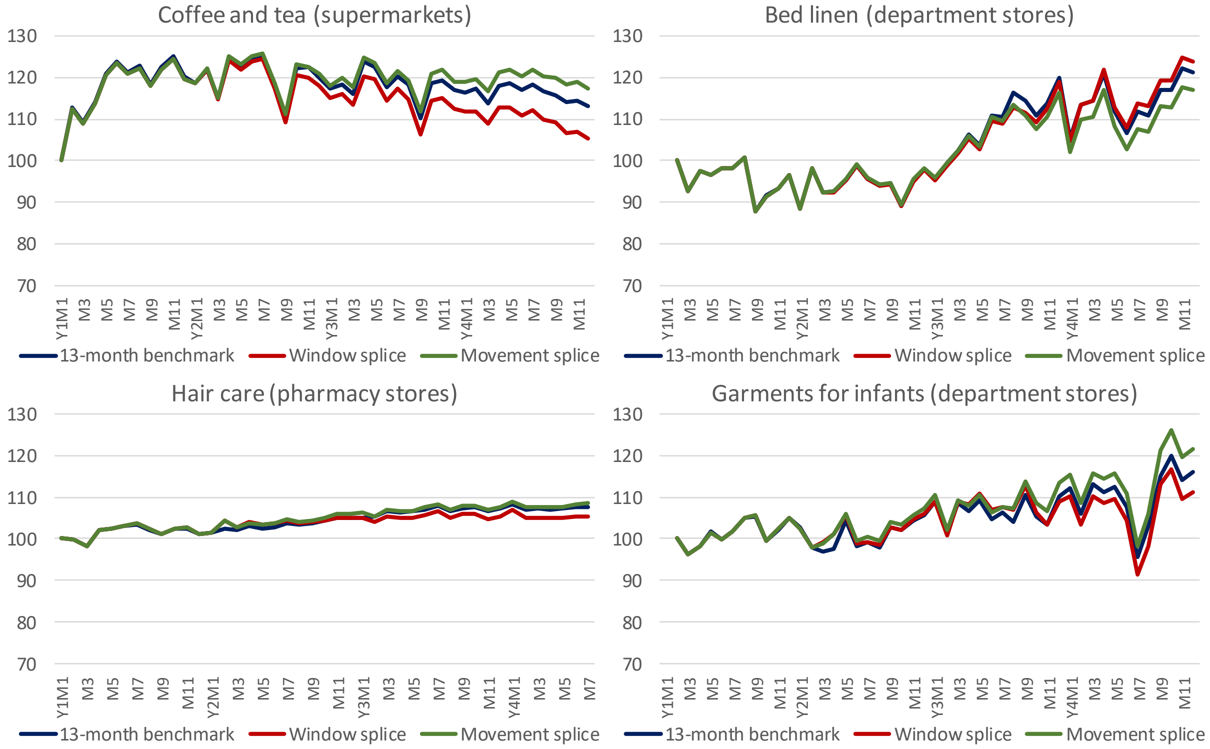


Figure 5. Price indices for the fixed base extension methods for four product categories, compared with the transitive indices on 13-month windows.

Table 1. Differences between the year on year indices (extension method minus 13-month transitive index) at chain level. The average differences over the months in the third and fourth year are given and also the smallest and largest differences within the two years.

Extension method	Year 3			Year 4		
Supermarket chain	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Window splice	-0.93	-1.21	-0.76	-0.81	-1.03	-0.66
Movement splice	0.47	0.23	0.65	0.61	0.52	0.69
FBEW	0.32	0.00	0.60	-0.26	-0.42	0.01
FBRW	0.14	-0.21	0.32	-0.04	-0.19	0.06
Pharmacy stores	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Window splice	-0.66	-1.12	-0.44	-0.19	-0.66	0.07
Movement splice	-0.34	-0.76	-0.12	-0.01	-0.46	0.22
FBEW	-0.10	-0.33	0.05	0.10	-0.15	0.27
FBRW	-0.17	-0.69	0.00	0.25	-0.17	0.53
Department stores, clothing excluded	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Window splice	-0.16	-0.49	0.34	0.61	-0.23	1.04
Movement splice	-0.68	-1.09	-0.07	-0.04	-0.80	0.36
FBEW	-0.17	-0.36	0.04	0.19	-0.07	0.46
FBRW	-0.17	-0.56	0.18	0.33	-0.06	0.77
Department stores, clothing	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
Window splice	0.65	-0.97	2.78	-0.22	-1.51	0.66
Movement splice	-1.93	-3.51	-0.41	0.14	-1.18	0.94
FBEW	0.51	-0.39	1.43	0.07	-1.15	0.72
FBRW	0.13	-2.03	1.59	0.58	-0.56	1.77

The differences for the two splicing methods may become very large at product category level, as Figure 4 shows, in particular for coffee and tea. The results for the fixed base extension methods are much more stable. These methods are free of drift by construction, which is clearly illustrated in Figure 3 and Figure 5. This means that also the results at product category level are accurate.[4]

The smallest and largest differences in the year on year indices compared to the transitive benchmark indices give an idea of the volatility of the indices produced with the extension methods. The variability in the year on year differences is larger for the splicing methods. When looking at the fixed base methods, it is interesting to note that the FBEW method produces more stable results than the FBRW method, in spite of the expanding window character of the former method. The FBEW method gives better results than the FBRW method in three of the four aggregate levels, both in terms of average differences and with regard to the variability in the differences of the year on year indices. The FBRW method shows better performance on the supermarket data.

[4] The 13-month benchmark indices for hair care and infant garments in Figure 4 and Figure 5 differ slightly from the indices calculated after defining products with the method MARS in Chessa (2019). In the present paper the same characteristics were used for the product definitions in each year, while in Chessa (2019) the selection of attributes was allowed to vary from year to year.

Figure 6. Price indices that are transitive on 13-month windows compared with transitive indices for the full period as time window, at highest aggregate level.

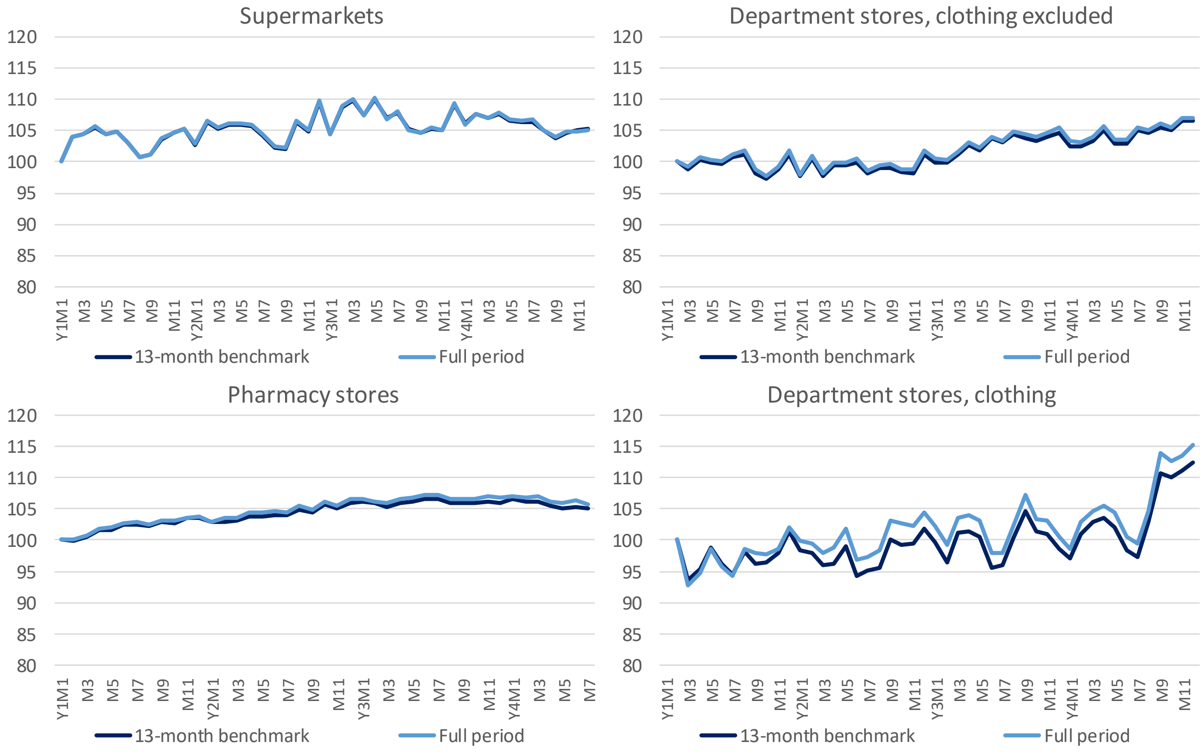
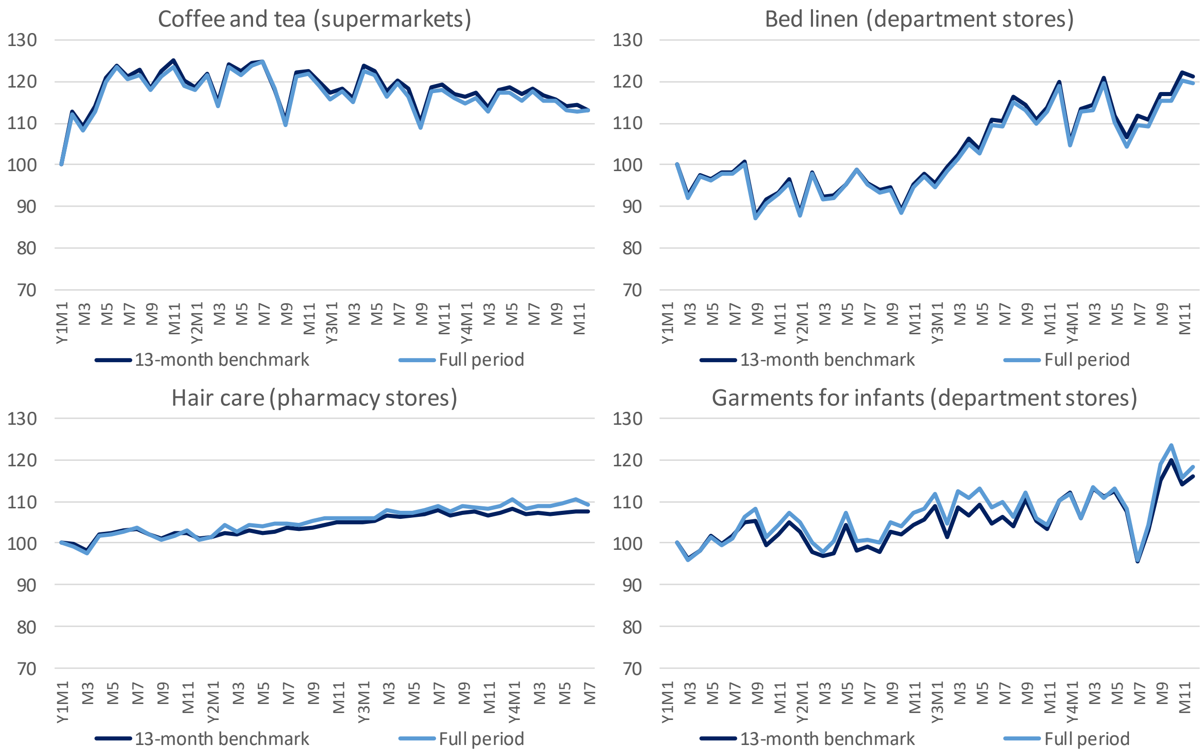


Figure 7. Same comparisons as in Figure 6, but now for the four product categories shown in previous graphs.



In Figure 6 and Figure 7, the piecewise transitive indices on 13-month windows are compared with price indices that are calculated simultaneously on the full time windows that vary between 43 and 48 months in length for the different data sets. The latter indices are therefore transitive over the whole time period. The graphs show small to very small differences, also at product category level. Detailed comparisons between the indices obtained with the extension methods and the full-window transitive indices are therefore omitted. The average values shown in Table 1 change between 0 and 0.3 percentage point for all methods when the full-period transitive indices are used as benchmark.

The largest differences are found for clothing in Figure 6. The differences can be ascribed to a large extent to the introduction of new collections of T-shirts at the beginning of the second year of the time period, which had high introduction prices and immediately generated high expenditures. The prices of these collections decreased rapidly within a few months, which results in different values of the 'reference' product prices (i.e. the v_i in Chessa (2016)) for the 13-month and full-period indices, especially in the second year.

4 Analysis and splicing reconsidered

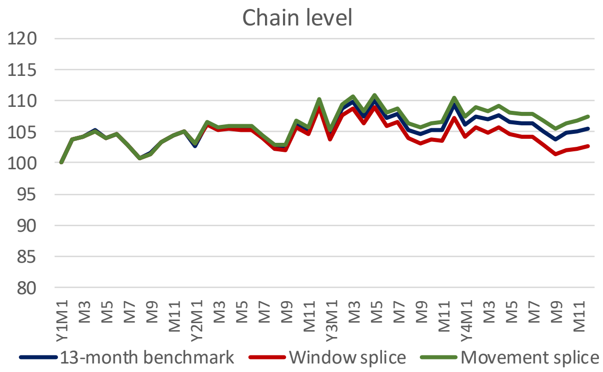
4.1 Analysis of results

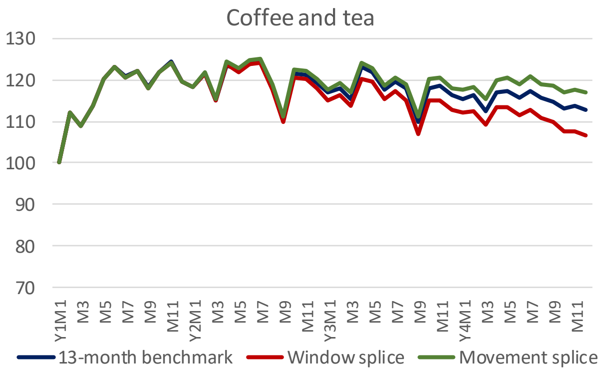
The results in the previous section clearly show that splicing has a high risk of drift and should therefore not be recommended when calculating price indices with multilateral methods. This conclusion holds when splicing is applied in its classical form, according to the characterisation given in Section 2. The fixed base methods give the best results and are free of drift.

The results are obtained with the Geary-Khamis method. The choice of index method may affect the differences between the index extension methods. An interesting question therefore is whether similar results are obtained with another multilateral method. This question is first addressed before analysing the results of Section 3.2 in more detail.

The Time Product Dummy (TPD) method was applied, which is the counterpart of the Country Product Dummy method in the time domain (Summers, 1973; Krsinich, 2014; de Haan, 2015). This was only done for the supermarket data, which show the largest differences for window splicing, while the differences for this data set are also large for movement splice. The results for the two splicing methods are shown in Figure 8. The indices at chain level are shown and also the indices for coffee and tea.

Figure 8. TPD indices for window splice and movement splice for the supermarket data, compared with the transitive indices on 13-month windows.





The results are almost the same as for the GK method. The differences between the year on year indices for window splice and the benchmark index are even about 0.1 percentage point larger on average than with the GK method. The differences for movement splice are almost the same as in Table 1. The fixed base methods give practically the same results as for the GK method and are therefore omitted.

The shortcomings of splicing methods can be formulated as follows:

- Movement splice is a monthly chained method. This property makes movement splice sensitive to drift like monthly chained bilateral methods.
- The main problem with classical window splice is the choice of index in the linking month, which is a recalculated index. The year on year indices calculated on rolling windows may therefore differ from the year on year indices with respect to the published indices of 12 months ago, which introduces a source of potential drift.

The second bullet also holds for splicing variants with a different linking month. The same can be claimed for mean splice (Diewert and Fox, 2017), since it takes an average of spliced indices over the different choices of linking month. Drift can therefore not be excluded for this method either. There is no certainty that drift will be eliminated by taking an average.

In the remainder of this section the focus will be on window splice. One of the reasons for the drifting behaviour of window splice is the increased influence of decreasing prices of exiting products on the spliced indices as the time window is shifted each month. The following notation is introduced in order to explain this in more detail. Let $\tilde{p}_t$ denote the 'price level' (unnormalised index) in month t, $p_{i,t}$ the price of item or product $i \in G_t$ and let $s_{i,t}$ denote its expenditure share in month t, with G_t representing the set of items or products sold in the corresponding month. Like in Chessa et al. (2017), the symbol v_i is used to denote the 'reference price' of i, which is calculated as an average of (deflated) prices over some time window $[0, T]$.

The price levels $\tilde{p}_t$ can be expressed as functions of the ratios $p_{i,t}/v_i$ for three multilateral methods that are being studied by NSIs: the GK, TPD and the GEKS-Törnqvist (GEKS-T) method. For GK we can write:

$$\tilde{p}_t = \left(\sum_{i \in G_t} s_{i,t} \left(\frac{p_{i,t}}{v_i} \right)^{-1} \right)^{-1}, \tag{1}$$

for TPD we have a similar expression:

$$\tilde{p}_t = \prod_{i \in G_t} \left(\frac{p_{i,t}}{v_i} \right)^{s_{i,t}}, \tag{2}$$

while for the GEKS-T the price level takes the form:

$$\tilde{p}_t = \left\{ \prod_{i \in G} \left(\frac{p_{i,t}}{v_i} \right)^{s_{i,t}} \right\}^{1/2} \left\{ \prod_{i \in G} p_{i,t}^{\frac{1}{T+1} \sum_{z=0}^{T} s_{i,z}} \right\}^{1/2}. \tag{3}$$

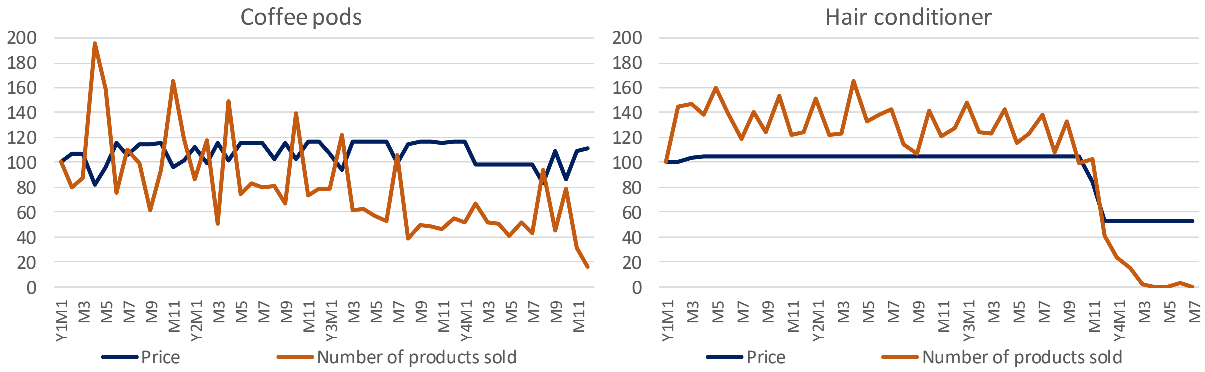
Price indices in month t with respect to a comparison month can be obtained by dividing the price levels in the two months. Note that the v_i are defined in a different way for the three methods (Chessa et al., 2017).

The terms within brackets in (3) take products over a fixed set G. Expression (3) holds for the situation where prices are known for each product in every month, or when prices are estimated in all months without sales, in which case prices are not available in transaction data sets. This version of the GEKS-T is used in the Australian CPI (ABS, 2017).

Figure 9 shows the development of the prices and numbers sold for two products. The price of the hair conditioner starts to drop rapidly after month 10 of the third year, and also the numbers of products sold. When the first month of the time window gets closer to month Y3M10 as it is shifted, the influence of the lower prices in the v_i increases. This influence is bigger for the GEKS-T, since this method applies equal weights to the monthly prices (Chessa et al., 2017). But the increasing influence of the lower prices on the v_i is also notable for the GK and TPD in this example, since the hair conditioner is still sold in fairly large quantities in the first months after prices have decreased.

The v_i for products like the hair conditioner will decrease, so that $p_{i,t}/v_i$ will increase. Linking an index in a month when the regular prices still apply causes the numerator ($\tilde{p}_t$) of a price index to increase by a smaller factor than the denominator because of the decreasing expenditure shares $s_{i,t}$. This results in a downward effect on the index. The downward bias may persist when continuing window splice in subsequent months, because the linking is done on recalculated indices, which tend to go down.

Figure 9. Development of prices and numbers of products sold for a GTIN of coffee pods (supermarket data) and a hair conditioner (pharmacy store data), with Y1M1 = 100.



The drifting behaviour of splicing methods is related to the degree of churn, which characterises pharmacy products. But the occurrence of drift is apparently not restricted to cases like the hair conditioner example. COICOP 01 types of product generally show much lower degrees of churn. So the question is what causes the huge amount of drift in the spliced indices for coffee and tea.

Figure 9 shows the behaviour of the price and number of products sold for coffee pods at GTIN level. The sales quantities show a spiking pattern that is typical for months with discounts. The results for window splice for coffee and tea are more difficult to understand, so let's proceed by steps. First, note that the spikes reach lower values over time. Next, imagine that the first month of a 13-month window has arrived at month M4 of year Y2, which is a month with a discount and a huge increase in quantities sold. The 13-month window contains three of such spikes. When the window is shifted, the new window contains one discount price and one spike of sales quantities less than the preceding window.

Subsequent window shifts cause temporarily lower values of the v_i for products like the coffee pods example as the number of spikes in a rolling window varies. Window splice produces similar effects as in the hair conditioner case. If Y2M4 is the first month of the window, the huge increase in the number of products sold in that month results in an upward effect on the denominator of the price index, which is larger than the numerator because the sales, and consequently also the expenditure shares, show an overall decrease, including those for spikes in subsequent months. As was argued for the hair conditioner, splicing on the resulting index continues to produce downward biased results in subsequent window splices, which may be aggravated because of the periodically higher numbers of spikes in subsequent rolling windows. This may offer an explanation for the serious drift in Figure 4 and Figure 8.

4.2 Adjusted splicing methods

If the index in the linking month is such an influential choice aspect, then one could ask: why not take the published indices as linking indices? Modifying classical window splice on this aspect has a number of important advantages:

- The year on year changes calculated on rolling windows will also be the published year on year indices in each month;
- The year on year indices on each rolling window are index changes of transitive series. Linking these changes to published indices results in index series that are free of drift on the length of the time window.

The results for the window splice method with a 13-month window for the GK method, with the splicing carried out on the published indices in the linking month, are shown in Figure 10 at chain level and in Figure 11 for the four product categories shown in previous graphs. The indices are a considerable improvement over the classical window splice indices shown in Figure 2 and in Figure 4. The drift-free property of this alternative window splice method, as formulated under the second bullet above, can be clearly noted.

The index for clothing at chain level shows some variability (and to a lesser extent also for pharmacy products). This may be caused by the same factors that lead to drift in the classical window splice indices, such as clearance prices, discount prices and also out-of-season prices. A longer window could reduce this variability, as the inclusion of prices from additional past months should suppress the aforementioned low prices. A 25-month window was considered, with the central month of the window chosen as linking month. This gives rise to the half splice method, which was originally proposed by de Haan (2015).

Figure 10. Price indices for window splice (WS) on published indices of 12 months ago at chain level, compared with the transitive indices on 13-month windows.

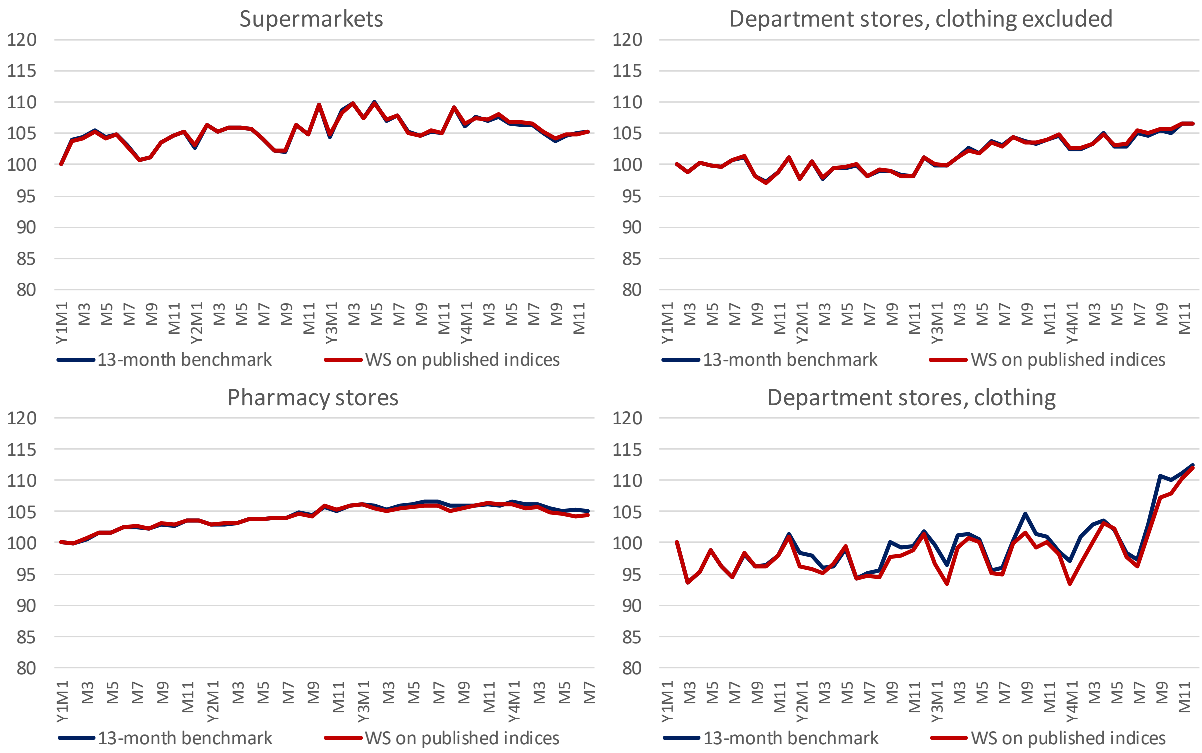
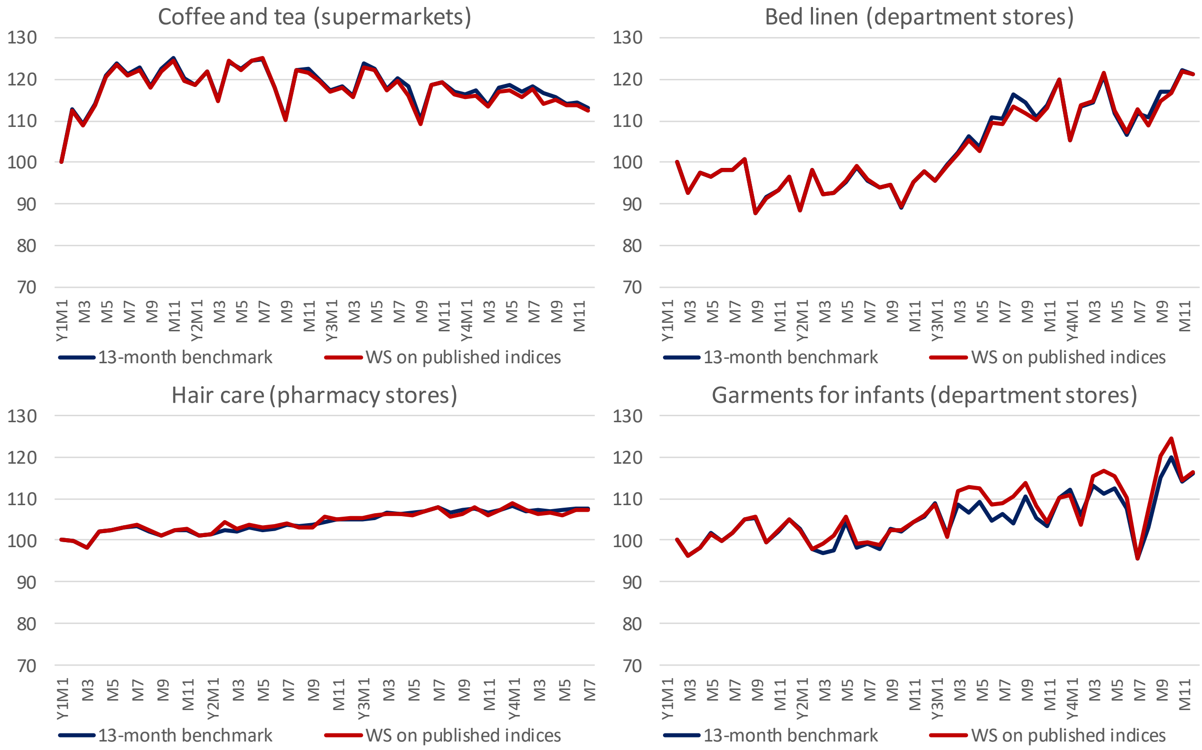


Figure 11. Price indices for window splice (WS) on published indices of 12 months ago for four product categories, compared with the transitive indices on 13-month windows.



But also the half splice method was used in the present study to link year on year indices onto the published indices of 12 months ago, as was done with the window splice method. The results are shown in Figure 12 and Figure 13. The half splice indices do not seem to suffer from the variability shown in the window splice indices for clothing. The indices for the half splice method also seem to tend more towards the transitive indices on the full time period, which are also included in both figures. The effects of using a longer time window are visible in the graphs.

Table 2 quantifies the impact of the window and half splice methods on the year on year indices, by comparing these indices with those for the index series that are transitive on 13-month windows and on the full period. A comparison with the results for the classical splicing methods in Table 1 leads to the conclusion that splicing onto the published indices of 12 months ago yields a considerable improvement in the accuracy of the year on year indices.

Both window splice and half splice compare very well with the year on year indices of both the 13-month and full period transitive benchmark indices. The half splice method shows the best overall performance of the two splicing methods, in particular for the pharmacy store data. Also note the relatively small variability in the differences of the monthly year on year indices, not only when compared with the 13-month benchmark indices, but also with respect to the indices that are transitive on the full period. These observations make the half splice method, when splicing onto the published indices of 12 months ago, a very interesting method to consider by NSIs.

Figure 12. Price indices with a half splice (HS) on published indices of 12 months ago at chain level, compared with the transitive indices on 13-month windows and for the full period.

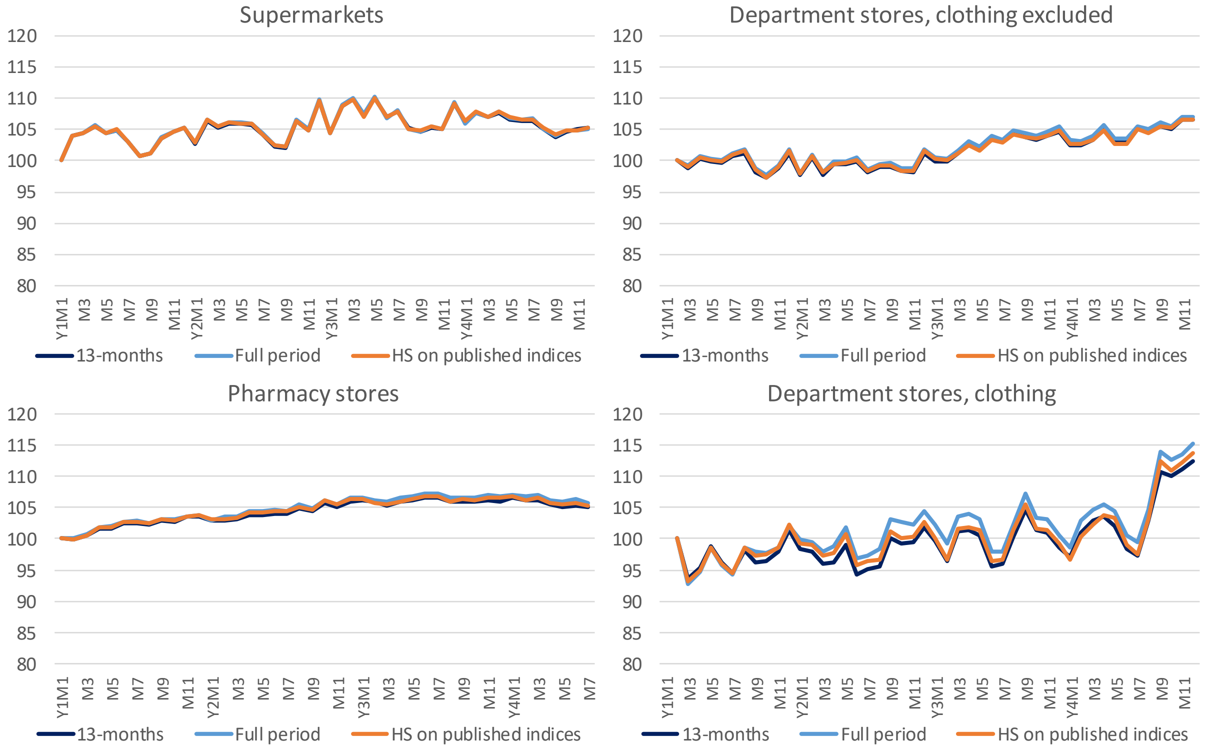


Figure 13. Price indices with a half splice (HS) on published indices of 12 months ago for the four product categories shown previously, compared with the transitive indices on 13-month windows and the full period.

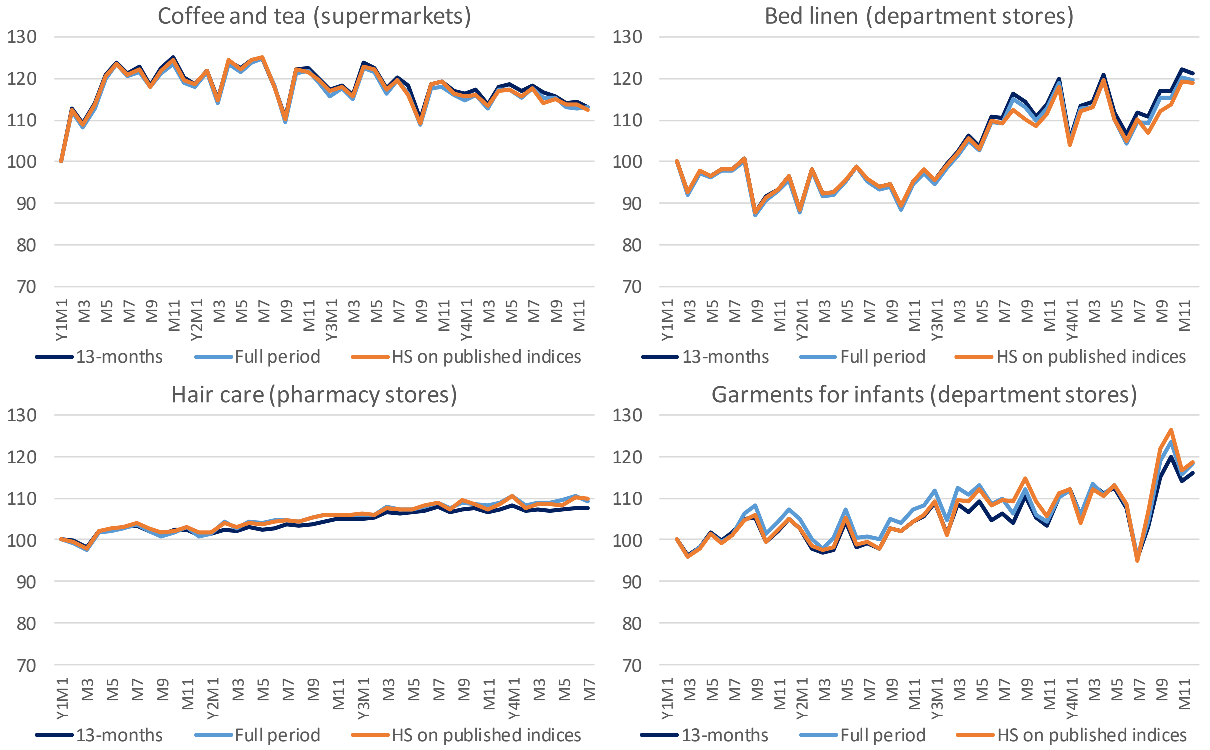


Table 2. Differences between the year on year indices (extension method minus 'benchmark' index) at chain level. The average differences over the months in the third and fourth year are given for both the 13-month and full period benchmarks and also the smallest and largest differences within the two years.

Extension method	Year 3			Year 4		
Supermarket chain	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
WS on published indices (13 months)	-0.05	-0.32	0.19	0.19	-0.19	0.48
HS on published indices (13 months)	-0.15	-0.47	0.18	0.20	-0.12	0.46
WS on published indices (full period)	0.09	-0.26	0.41	0.17	-0.29	0.52
HS on published indices (full period)	-0.01	-0.28	0.37	0.18	-0.17	0.51
Pharmacy stores	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
WS on published indices (13 months)	-0.32	-0.65	0.00	-0.17	-0.26	-0.02
HS on published indices (13 months)	-0.05	-0.43	0.17	0.18	0.00	0.32
WS on published indices (full period)	-0.42	-0.58	-0.20	-0.44	-0.72	-0.26
HS on published indices (full period)	-0.16	-0.26	-0.02	-0.09	-0.24	-0.01
Department stores, clothing excluded	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
WS on published indices (13 months)	-0.13	-0.50	0.19	0.30	-0.06	0.75
HS on published indices (13 months)	-0.26	-0.62	0.14	0.02	-0.25	0.18
WS on published indices (full period)	-0.11	-0.30	0.10	0.20	-0.14	0.61
HS on published indices (full period)	-0.25	-0.47	-0.08	-0.08	-0.33	0.23
Department stores, clothing	**Mean**	**Min**	**Max**	**Mean**	**Min**	**Max**
WS on published indices (13 months)	-0.54	-1.14	0.49	-0.27	-1.26	0.87
HS on published indices (13 months)	-0.59	-1.25	-0.18	-0.12	-1.23	0.92
WS on published indices (full period)	-0.32	-2.15	1.80	-0.11	-0.87	0.91
HS on published indices (full period)	-0.37	-2.13	1.14	0.04	-0.70	0.56

5 Conclusions

The number of NSIs that are already using transaction data in their CPI or are investigating this data source for future use has grown rapidly over the past years. In line with this development, the awareness about the potential of these data, both in terms of expanding product coverage in the CPI and the broader range of index methods that can be considered for CPI compilation, is growing as well.

The broader range of possible choices implies that the problem of compiling price indices increases in complexity. On the one hand, transaction data allow NSIs to compile more accurate index numbers, but on the other hand the broader range of choices can lead to divergent results. How to make methodologically sound choices therefore becomes increasingly important, also with regard to comparability of methods and indices across countries.

In this changing landscape, a growing number of statistical institutes are investigating multilateral methods for their CPI. This class of methods has important advantages: transitive indices can be calculated, these methods can deal with assortment dynamics, with products leaving and entering stores, new products can be directly included in index calculations, and price imputations and filters are no longer needed when using appropriate weighting schemes.

It is important to emphasise that multilateral methods come with a broad palette of choice aspects. Beside the problem of defining products, which is independent of index method, choices have to be made about the length of the time window, the index method and weighting scheme, and how to extend or continue index series from month to month. This paper has focused on the latter choice aspect. Of course, the length of the time window is part of this problem. The results of this paper can also be used to draw conclusions about a suitable choice for the length of the window, which appears to be less influential than the type of extension method.

Different extension methods have been proposed in the literature, which can be broadly subdivided into two main classes: splicing methods and fixed base extension methods. Methods from both classes were included in an extensive comparative study. The main conclusion from this study is that fixed base methods give much better results than window splice and movement splice. Fixed base methods are free of drift by definition. The two splicing methods have shown considerable drift, even at retail chain level. Previous studies gave similar results for different splicing methods at COICOP and overall CPI level (Van Loon and Roels, 2018) and a downward drift of window splice was found for lower aggregates (ABS, 2017). The bigger differences found in this study between extension methods for higher aggregates could therefore be considered to be a novel empirical finding.

The characterisation of index extension methods presented in Section 2 may be helpful in finding an explanation for the drifting behaviour of splicing methods and also for suggesting and investigating potentially better alternative ways of splicing. Index series can be continued from month to month by making choices about three aspects: (1) the type of window adjustment, (2) the linking month, and (3) the index in the linking month, on which an index change resulting from the new time window will be linked.

The third aspect is barely stressed and remains hidden in the literature, but it proves to be crucial in the whole exercise. Splicing methods link onto the latest index, which is a recalculated index (except for movement splice which, by nature, links to the index published in the previous month). Eventually, we are interested in the behaviour of the series of published indices. Index series on rolling windows are all transitive. But splicing onto recalculated indices results in a different index change with respect to the published index in the linking month compared to the index change calculated on a rolling window. This creates a potential source of drift.

Splicing onto published indices sounds like a logical remedy, because the calculated index change on the rolling window will be equal to the published index change. At the same time, we have the certainty that the published series will not drift on the time interval that runs from the linking month until the current month (i.e. the full length of the window in window splice).

The results obtained in Section 4.2 for the remedied window and half splice methods are in line with these expectations. The results are much better than for movement splice and classical window splice. The alternative splicing methods compete very well with the fixed base methods. While the alternative window splice method exhibits some variability in its month to month movements, the half splice method with a 25-month window hardly suffers from this. A possible explanation for its more stable behaviour is that the impact of reduced prices, such as clearance prices, discounts and out-of-season prices, is suppressed by including prices from 12 months before the linking month.

The results of this study recommend fixed base methods and the half splice method that links year on year indices onto the published indices of 12 months ago. This version of the half splice method has given accurate results and is particularly appealing for a number of reasons:

- The generated series of published indices is free of drift on 13-month intervals;

- The year on year indices calculated on the rolling windows will also be the published year on year indices;
- The calculation of product contributions to year on year indices should be easier;
- The calculation of indices for strongly seasonal items will benefit from the use of a longer 25-month window;
- Unpredictable effects in year on year indices that may result from changes to new data sources and/or methods from one year to another in the CPI could be suppressed, since the role of the fixed base month is in fact bypassed.

An additional note that may be worth pointing out is that the longer time window used in the half splice method could also contribute to reducing the differences between the price indices calculated with different multilateral methods. The impact of low prices like clearance prices on diverging results between methods should be reduced by taking a longer time window.

The second point greatly contributes to transparency when using multilateral methods, as the CPI is a year on year statistic in the first place. The last point is becoming more relevant now that more countries are considering new data sources and new methods. Also for this reason, this paper encourages NSIs to include the half splice method in their research plans. Given the promising results obtained with the half splice method, it could also be interesting to consider a variant of the FBEW expanding window method that adds the entire previous year to the window and extends the window to its maximum length of 25 months during the current year.

Acknowledgements

This research is funded by a Eurostat grant. The author expresses his gratitude to Eurostat for assigning this grant.

References

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.

Balk, B.M. (1981). A simple method for constructing price indices for seasonal commodities. *Statistische Hefte*, 22:1, 72-78.

Balk, B.M. (1996). A comparison of ten methods for multilateral international price and volume comparison. *Journal of Official Statistics*, 12, 199-222.

Balk, B.M. (2001). Aggregation methods in international comparisons: What have we learned? ERIM Report, Erasmus Research Institute of Management, Erasmus University Rotterdam.

Balk, B.M. (2008). *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*. Cambridge University Press.

Chessa, A.G. (2016). A new methodology for processing scanner data in the Dutch CPI. *Eurostat Review on National Accounts and Macroeconomic Indicators*, issue 1/2016, 49-69.

Chessa, A.G. (2017). The QU-method: A new methodology for processing scanner data. *Proceedings of Statistics Canada 2016 International Methodology Symposium*.

Chessa, A.G. (2018). Product definition and index calculation with MARS-QU: Applications to consumer electronics. Statistics Netherlands.

Chessa, A.G. (2019). MARS: A method for defining products and linking barcodes of item relaunches. Paper presented at the NTTS Conference, 11-15 March 2019, Brussels, Belgium.[5]

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.

Deaton, A., and Heston, A. (2010). Understanding PPPs and PPP-based national accounts. *American Economics Journal; Macroeconomics*, 2, 1-35.

Diewert, W.E. (1999). Axiomatic and economic approaches to international comparisons. In A. Heston and R.E. Lipsey (eds.), *International and Interarea Comparisons of Income, Output and Prices*, Studies in Income and Wealth, Vol. 61, pp.13-87. University of Chicago Press.

Diewert, W.E., and Fox, K.J. (2017). Substitution bias in multilateral methods for CPI construction using scanner data. Discussion paper 17-02, University of British Columbia, Vancouver, Canada.

Eltetö, Ö, and Köves, P. (1964). On an index computation problem in international comparisons. *Statiztikai Szemle*, 42, 507-518. [in Hungarian]

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.

Gini, C. (1931). On the circular test of index numbers. *International Review of Statistics*, 9, 3-25.

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.

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.

Ivancic, L., Diewert, W.E., and Fox, K.J. (2009). Scanner data, time aggregation and the construction of price indexes. Discussion paper 09-09, University of British Columbia, Vancouver, Canada.

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.

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.

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.

Summers, R. (1973). International price comparisons based upon incomplete data. *Review of Income and Wealth*, 19, 1-16.

Szulc, B. (1964). Index numbers of multilateral regional comparisons. *Przeglad Statystyczny*, 3, 239-254. [in Polish]

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.

[5] Also submitted to the 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro, Brazil, 8-10 May 2019.