ECONOMIC STATISTICS CENTRE OF EXCELLENCE

A COLLABORATION WITH Office for National Statistics

Multilateral index number methods for Consumer Price Statistics

Kevin J. Fox, Peter Levell and Martin O'Connell

ESCoE Discussion Paper 2022-08

April 2022

ISSN 2515-4664

DISCUSSION PAPER

Multilateral index number methods for Consumer Price Statistics
Kevin J. Fox, Peter Levell, Martin O'Connell
ESCoE Discussion Paper No. 2022-08
April 2022

Abstract

The increasing availability of supermarket scanner data covering expenditures and prices on a wide range of products has created new opportunities for national statistical institutes, including the possibility of publishing more reliable indicators of monthly price changes. We discuss and evaluate the properties of different multilateral index numbers for measuring high frequency price changes, drawing on household scanner data. We find that use of the Caves-Christensen-Diewert-Inklaar (CCDI) index, updated using the mean splice, is to preferred for both theoretical and empirical reasons.

Keywords: consumer price index (CPI), multilateral indexes, scanner data

JEL classification: C43, E31

Peter Levell, Institute for Fiscal Studies
Peter_l@ifs.org.uk

Published by:
Economic Statistics Centre of Excellence
National Institute of Economic and Social Research
2 Dean Trench St
London SW1P 3HE
United Kingdom
www.escoe.ac.uk

Multilateral index number methods for Consumer Price Statistics

Kevin J. Fox, Peter Levell, Martin O'Connell*

April 2022

Abstract

The increasing availability of supermarket scanner data covering expenditures and prices on a wide range of products has created new opportunities for national statistical institutes, including the possibility of publishing more reliable indicators of monthly price changes. We discuss and evaluate the properties of different multilateral index numbers for measuring high frequency price changes, drawing on household scanner data. We find that use of the Caves-Christensen-Diewert-Inklaar (CCDI) index, updated using the mean splice, is to preferred for both theoretical and empirical reasons.

Keywords: Consumer price index (CPI), multilateral indexes, scanner data
JEL classification: C43, E31

*This research has been funded by the Office for National Statistics (ONS) as part of the research programme of the Economic Statistics Centre of Excellence (ESCoE). Levell and O'Connell also acknowledge support from the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS (ES/T014334/1). We thank the ONS, especially Sanjiv Mahajan and Helen Sands, and the ONS Advisory Panel on Consumer Price Statistics – Technical for their for their constructive feedback and discussions. The R package IndexNumR by Graham White has been used extensively in producing the results in this report and supplementary results. His assistance is gratefully acknowledged. The report has also benefited from discussions and collaborations with Erwin Diewert, Jan de Haan and Shipei Zeng over the years. Data were supplied by Kantar FMCG Purchase Panel. The use of Kantar FMCG Purchase Panel data in this work does not imply the endorsement of Kantar FMCG Purchase Panel in relation to the interpretation or analysis of the data. The usual disclaimer applies. Supplementary results and corresponding R code, using Dominick's supermarket scanner data and an example of transitory goods from Destatis, will be made available via GitHub.

1 Introduction

The increasing availability of supermarket scanner data covering expenditures and prices on a wide range of products has created new opportunities for national statistical institutes (NSIs). Price changes can now be observed across many products at high frequency, potentially allowing NSIs to, for example, publish more reliable indicators of monthly price changes.

However, as yet there is no universally agreed on method for calculating price indices with high frequency transactions data, such as electronic point-of-sales scanner data. Conventional index numbers, such as those currently used in the UK Consumer Price Index (CPI), are ill-suited to this task. Scanner data potentially allows many more products to be incorporated in the calculation price indices, but must then contend with more product entry and exit and volatile movements in prices and quantities. Rapid product churn means that fixed-based indices can quickly become unrepresentative of consumer spending patterns, while chained indices often report excessive and unrealistic price changes, a phenomenon known as 'chain drift'.

Multilateral index numbers (first suggested in a time series context by Balk (1980)) offer a solution to chain drift: they can control the sometimes massive chain drift bias that can eventuate from using standard chained indexes with transactions data. But there are many possible multilateral index numbers for NSIs to choose from. These indices differ in their theoretical and empirical properties, and there is no consensus on which index is considered 'the best'. What is more, even if consensus could be achieved on the most suitable multilateral index, there are different ways to extend these indices when data from new time periods are added. Extension methods are needed when using multilateral indexes on non-revisable price indices such as the CPI. Unfortunately, the necessary use of extension methods will lead to chain drift in the resulting index, but significantly less than the kind of explosive chain drift that can occur when using traditional methods; see Box 1 in Section 2.

Drawing on stakeholder feedback, the UK Office for National Statistics (2020) proposed a scoring system for different methods for calculating higher frequency indices, assigning weights for different properties on the basis of their perceived importance (Office for National Statistics, 2020). These include characteristics such as ease of explanation and the resources involved in calculation. This framework narrowly favoured the use of the Geary-Khamis index with the 'fixed base monthly expanding window' (FBEW) extension method.

The weights used in such exercises are inherently subjective (e.g. 10%, 5% or 3.33%) and are not typically founded in axiomatic or economic theory. Indeed, altering these weights slightly can affect the ranking.[1] We note that with only three quite reasonable changes to the ONS scoring system, Geary-Khamis loses its top ranking to Caves-Christensen-Diewert-Inklaar ("CCDI" or GEKS-Törnqvist) and GEKS-Fisher, both using the mean splice. Geary-Khamis with the FBEW extension method was assigned top ranking because of the fact Geary-Khamis satisfies the additivity axiom (the quantity index corresponding to the price index can be written as a weighted sum of goods), and because of its perceived ease of explanation to a non-technical audience. However, while additivity is a desirable feature, allowing NSIs to publish sub-indices that sum to give the overall index, other indices also have desirable decomposition properties, including the CCDI index. It is also questionable that Geary-Khamis (or the FBEW extension) method is easier to explain than other indices to non-technical audiences. The FBEW extension method was preferred because it satisfies the property "Is the price comparison accurate with binary time periods being compared?" (a property also satisfied by fixed base and chained monthly indices). It is not the case however that the FBEW extension method will be consistent with purely bilateral comparisons across time periods (and there is no particular reason to think that the FBEW is any more 'accurate' than other extension methods). Were it not for having these points in its favour, the Geary-Khamis index with FBEW extension method would be given the same score as the CCDI and GEKS-Fisher indices with the mean splice.

In this report we discuss and evaluate the properties of different multilateral index numbers. In our empirical work we use UK household level scanner data collected by a market research firm. Specifically, we use Kantar's FMCG Purchase Panel over the period 2013-2019 to empirically assess different approaches, including different index number methods, splicing techniques and window lengths. Our data enables us to test the the performance of different indices across a large number of different of products. We show that the Geary-Khamis index appears to be especially sensitive to the extension method used, suggesting that if it is used, the extension method would have to be chosen with care and with consideration of the particular context, and that the choice may be consequential in the future.[2] Empirically, we find that the CCDI index is by contrast relatively insensitive to the

[1] We acknowledge that the ONS (2020) publication is not a living document, and that the approach to scoring may have subsequently changed given international research developments.

[2] Our findings are consistent with other evidence in the literature. For example, Lamboray (2017b) found that "GEKS is in general less sensitive to the choice of window length and splicing method than the Geary-Khamis or WTPD. The greater sensitivity of the Geary-Khamis and WTPD has also been noticed in other studies (for example Van Loon and Roels (2018))."

choice of extension method, and that the CCDI index with mean splice in particular seems to work well relative to the benchmarks considered.

Additional factors also favour CCDI over alternatives such as the GEKS-Walsh index, weighted time-product dummy index (WTPD) and Geary-Khamis index. In general, multilateral price indices can give misleading impressions of month-to-month price changes when they are driven by transitory products. The intuition for this is most obvious to see for GEKS-type indices, but in these situations the Geary-Khamis index can also produce counter-intuitive results.[3] A common solution to the problem of transitory products is to impute the values of prices for goods in periods when they are not available, but the Geary-Khamis and WTPD indices are invariant to price imputations for products with zero quantity weights. Imputation for missing prices, including hedonic imputation, is a standard procedure for NSIs, so this inconsistency is problematic.[4] The fact that the CCDI index is compatible with imputation again favours the use of this index over others.

In addition to the splicing methods considered by the ONS, we also examine a newly proposed framework for splicing: splicing on the published series. In a new result in the literature, we found that almost exactly the same results are found from using the mean splice in the regular manner as using the mean splice on the published series, regardless of index number formula used; hence we do not report the results from the mean splice on the published series, as they are generally indistinguishable from those for the regular mean splice results which are reported. The same does not hold for other splicing methods considered.

Overall, the CCDI index with the updating using the mean splice may be preferred, for both theoretical and empirical reasons.

The remainder of this report is structured as follows. Section 2 sets out different price indices and their properties when using transactions data. Section 3 describes the Kantar data we use to empirically assess the performance of different indices. Section 4 describes empirical differences between indices calculated on the full data set, and plotting the distribution of differences across product categories. Section 5 quantifies the extent of chain drift for different indices, window lengths and splicing methods across a large set of product categories. Section 6 sets out some of the drivers of chain drift. Section 7 shows the differences in index values over time and for different approaches for a product category that exhibits significant seasonality (chocolate). Section 8 concludes and discusses potential avenues for future research.

[3]In other work, we demonstrate this for an example of transitory goods provided by Destatis. Results and code are available on request, and will soon be made available via Github.

[4]See e.g. Wells and Restieaux (2014) for a UK-focussed review of imputation methods, and the IMF (2020) CPI Manual, Chapter 6 for more on imputation methods.

2 Multilateral price indices and their properties[5]

Motivation. Let $\mathbf{p}^t = (p_1^t, \ldots, p_N^t)'$ denote the vector of prices for N goods in period t, and let $\mathbf{q} = (q_1, \ldots, q_N)'$ be a vector of quantities for the goods purchased by some representative consumer. Suppose we wish to compare how the cost of purchasing the basket of goods changes between t and $t+1$. One way of measuring this change is with $P_{Lo}^{t,t+1} = \frac{\mathbf{p}^{t+1\prime}\mathbf{q}}{\mathbf{p}^{t\prime}\mathbf{q}}$. This is called a Lowe price index and is commonly used in CPI construction. If base period quantities are used (i.e. $\mathbf{q} = \mathbf{q}^t$) the index is known as a Laspeyres index ($P_L^{t,t+1}$) and if end period quantities are used ($\mathbf{q} = \mathbf{q}^{t+1}$) it is known as Paasche index ($P_P^{t,t+1}$). The indices can be re-written in terms of price relatives for each good between t and $t+1$, $\frac{p_n^{t+1}}{p_n^t}$, weighted by the share of expenditure allocated to them, $s_n^t = \frac{p_n^t q_n^t}{\mathbf{p}^{t\prime}\mathbf{q}^t}$:

$$
\begin{aligned}
P_L^{t,t+1} &= \sum_n s_n^t \frac{p_n^{t+1}}{p_n^t} \\
P_P^{t,t+1} &= \left(\sum_n s_n^{t+1} \left(\frac{p_n^{t+1}}{p_n^t} \right)^{-1} \right)^{-1}.
\end{aligned}
$$

A problem with these indices is, that because they use weights that correspond either to the base or end period, they are subject to substitution bias. Superlative indices seek to deal with this problem and entail using a combination of base and final period weights. Three commonly used superlative indices are the Fisher index (a geometric mean of the Laspeyres and Paasche indicies), the Tornqvist index (a geometric mean of price changes weighted by average budget shares in the base and end periods), and the Walsh index (an arithmetic average of price changes weighted by the geometric mean of quantities in the base and end periods). These indices take the form:

$$
\begin{aligned}
P_F^{t,t+1} &= (P_L^{t,t+1} P_P^{t,t+1})^{1/2} \\
P_{Tq}^{t,t+1} &= \prod_n \left(\frac{p_n^{t+1}}{p_n^t} \right)^{0.5(s_n^t + s_n^{t+1})} \\
P_W^{t,t+1} &= \frac{\sum_n \sqrt{q_n^t q_n^{t+1}} p_n^{t+1}}{\sum_n \sqrt{q_n^t q_n^{t+1}} p_n^t}
\end{aligned}
$$

[5] Much of this discussion is based on Diewert and Fox (2022). The interested reader can find more details and references there.

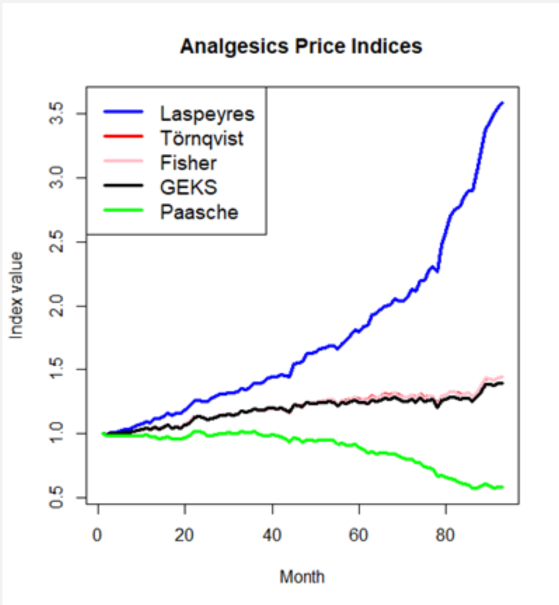
Suppose we have three time periods, $t = 1, 2, 3$ and wish to compare price changes over this time given a preferred choice of index. In the comparison of periods 1 and 3 we can use the direct (fixed-base) comparison, $P^{1,3}$, or the chained comparison $P^{1,2}P^{2,3}$. In general these will give different answers. A significant drawback of the fixed base comparison is that it uses a fixed basket and hence will not capture product churn. This problem is especially serious if the fixed-base comparison is made over many periods, as the index will become increasingly unrepresentative of actual spending patterns. Hence it is generally preferable to used a chained index, as described in the ILO 2004 CPI Manual (p. 407): "rapid sample attrition means that fixed base indices rapidly become unrepresentative and hence it seems preferable to use chained indices which can more closely follow market-place developments."

However, the use of bilateral indicies (where comparisons are made between two periods), like the Laspeyres, Paasche, Fisher, Tornqvist or Walsh index, can give rise to *chain drift* bias, which we define as the difference between the direct comparison of prices between two periods and a chained comparison. More formally, an index that is free of chain drift satisfies the multiperiod identity test. See Box 1 and Box 2 for more on chain drift and the use of multilateral index numbers to address this problem.

Box 1. Chain drift

Rapid sample attrition means that fixed base indices rapidly become unrepresentative. Hence there is an international consensus that it is preferable to use chained indices which can more closely follow market-place developments. Suppose in the three period example, after the third period prices and quantities revert back to their levels in the first period. If a chained index yields a result of 1 for the price change from period 1 to period 3, then it is said to satisfy the multiperiod identity test: $P^{1,2}P^{2,3}P^{3,1} = 1$. Chained bilateral indices are not guaranteed to satisfy this property; they are therefore said to suffer from 'chain drift bias'.

Chain drift bias can become huge when using transactions level data. Consider the following example using Dominick's supermarket data:[a]



The Laspeyres, Paasche, Törnqvist and Fisher indexes are all chained in this example. The GEKS index in the figure is a multilateral method (GEKS-Fisher) calculated over the full sample, which by definition does not suffer from chain drift bias. Deviations from this index (the black line) then indicate change drift through deviating from a chain-drift-free index. Note how the chain drift bias is especially dramatic for indexes commonly used by NSIs, the Laspeyres and Paasche indices. While not evident here given the scale, Ivancic et al. (2011) showed that there could also be significant chain drift bias from superlative indices, such as the Törnqvist and Fisher indices.

[a]Available from the Kilts Center, University of Chicago Booth School of Business: https://www.chicagobooth.edu/research/kilts/datasets/dominicks.

Box 2. Multilateral methods as a solution to chain drift

1. Ivancic et al. (2011) proposed using multilateral indexes for controlling for chain drift bias.[a] Multilateral index numbers will satisfy the multiperiod identity test, regardless of product churn in the intervening periods.

2. CPIs are typically non-revisable. If a new period of data is added to the sample period, values of an index calculated on the new (longer) data set will be different than on the old data set. This is would require re-writing CPI history, which is not possible if there is a policy of not revising.

3. A method of extending the multilateral index as new data becomes available was suggested by Ivancic et al. (2011): a rolling window approach, where the new window of data is spliced onto the old window, allowing the calculation of the growth in prices from the end of the old window to the new observation of data.

4. This has already changed NSI practice in several countries. Yet there remains active research on the use of (i) alternative multilateral index formulae, (ii) alternative splicing methods and (iii) the choice of window length.

5. A multilateral index extended in this way will not in general satisfy the multiperiod identity test. Hence, the relative extent to which chain drift bias is introduced through the choice of methods is a key focus of current research. An aim of our report is to contribute evidence regarding this.

[a] The discussion paper version was published in 2009; see Ivancic et al. (2009). de Haan and van der Grient (2011) provided further evidence of the problem of chain drift and the effectiveness of multilateral methods in addressing the problem.

Consistent with the literature, we use the term 'chain drift bias' to refer to the failure to satisfy the multiperiod identity test. As the introduction of chain drift bias in a multilateral method arises from using a rolling window approach with splicing, it could be referred to by alternative terms, such as 'splicing bias'[6].

[6] This term suggested by Jan de Haan in personal communications

Multilateral index numbers. We focus our attention on five multilateral index numbers. The first three are based on the Fisher, Tornqvist and Walsh indices, and are called the GEKS-Fisher index, CCDI (or the GEKS-Tornqvist index) and the GEKS-Walsh index respectively. In each case, the price level in period τ is given by a geometric mean of the corresponding bilateral index that compares period τ with all other periods $t = 1, \ldots, T$. Hence, the measured price level in period τ under the indices is given by:

$$\begin{aligned}
\mathbb{P}^{\tau}_{GEKS-F} &= \prod_t \left[P_F^{\tau,t}\right]^{1/T} \\
\mathbb{P}^{\tau}_{CCDI} &= \prod_t \left[P_{Tq}^{\tau,t}\right]^{1/T} \\
\mathbb{P}^{\tau}_{GEKS-W} &= \prod_t \left[P_W^{\tau,t}\right]^{1/T}.
\end{aligned}$$

In addition we consider an un-weighted multilateral index for cases when quantity information is not available, the GEKS-Jevons index:

$$\mathbb{P}^{\tau}_{GEKS-J} = \prod_t \left[P_J^{\tau,t}\right]^{1/T},$$

where $P_J^{\tau,t}$ is the geometric mean of price relatives, $\prod_t \frac{p_n^{t+1}}{p_n^t}$.

The fifth multilateral index number we consider is the Geary-Khamis index. The Geary-Khamis index is an implicit price index, defined as total expenditure divided by a volume or quantity index, with 'quality adjustment factors' determining how many units of good i are equivalent to a unit of good j. The index is implicitly defined by the solution to a set of equations that jointly determine price levels, $\mathbb{P}^t_{GK}$, for $t = 1, \ldots, T$ and quality adjustment factors, b^n for $n = 1, \ldots, N$. It is helpful to denote total quantity on good n across all time periods by $q_n \equiv \sum_t q_n^t$. The $N + T$ equations that determine the price levels and quality adjustment factors are:

$$\begin{aligned}
b_n &= \sum_t \left(\frac{q_n^t}{q_n}\right)\left(\frac{p_n^t}{\mathbb{P}^t_{GK}}\right) \quad \text{for } n = 1, \ldots, N \\
\mathbb{P}^t_{GK} &= \frac{\mathbf{p}^{t'}\mathbf{q}^t}{\mathbf{b}'\mathbf{q}^t} \quad \text{for } t = 1, \ldots, T.
\end{aligned}$$

Each adjustment factor b_n is a share-weighted average of inflation-adjusted prices for each commodity n over all t periods. The usual method for obtaining a solution to these two equations is to iterate between them. However,Diewert and Fox (2022) derive an alternative method which is more efficient (p. 360, footnote 24). This more efficient method is the default in IndexNumR. The final index is expenditure divided by the sum of quality adjusted quantities purchased in each period.

Another multilateral index examined by Ivancic et al. (2009) was (what is now known as) the weighted time product dummy method, that is derived from the country product dummy method; see Summers (1973) and Rao (2005). We do not examine this index in the empirical comparisons that follow as NSIs appear reluctant to use regression-based methods for index construction.

For each of the multilateral index numbers, it is common to rebase the price levels relative to the first period of date: $P^{1,t} = \mathbb{P}^t/\mathbb{P}^1$. The comparison of prices in period t and $t+1$ is given by $P^{t,t+1} = \mathbb{P}^{t+1}/\mathbb{P}^t = P^{1,t+1}/P^{1,t}$.

Multilateral index numbers and substitution bias. The GEKS-Fisher index, CCDI and the GEKS-Walsh index are based on superlative bilateral indexes (i.e., Fisher, Tornqvist and Walsh). This means they are consistent, to a second order approximation, with arbitrary homothetic consumer preferences, and therefore limit substitution bias. In contrast, the Geary-Khamis index (as an index which satisfies the additivity property) is based on a linear preference structure. This means that consumers view goods as perfect substitutes. Diewert and Fox (2022) showed that the Geary-Khamis index was also consistent with consumers viewing goods as being not at all substitutes. For most product categories for which there are scanner data, such extreme consumer preferences are very unlikely.

The linking problem. Suppose we use a multilateral index to compute price levels over a given time period, $1, \ldots, T$. If data for period $T+1$ becomes available, re-computing the index over $1, \ldots, T+1$ will lead to a revision of price levels over the initial T periods. NSIs typically regard such revisions to past CPI levels as undesirable.

One approach to avoid this is to use a *rolling-window*, which entails computing a new set of prices levels over $2, \ldots, T+1$ and basing the measure of the period $T+1$ price level on the price levels computed over $1, \ldots, T$ linked with those computed over $2, \ldots, T+1$. More concretely, let $\mathbb{P}^1, \ldots, \mathbb{P}^T$ and $\widetilde{\mathbb{P}}^2, \ldots, \widetilde{\mathbb{P}}^{T+1}$ represent the two sequences of multilateral index numbers. The linked measure of the price level in period $T+1$ is given by:

$$\rho^{T+1}(s) = \frac{\mathbb{P}_s}{\mathbb{P}_1} \frac{\widetilde{\mathbb{P}}^{T+1}}{\widetilde{\mathbb{P}}^s},$$

where s denotes the period used to link the two series. Different choices of s correspond to different linking methods. We consider the following rolling-window methods:

- The movement splice: $s = T$
- The window splice: $s = 2$
- The half splice: $s = \frac{T}{2}$ (or, when T is an odd number $s = \frac{T+1}{2}$)
- The mean splice, which entails a geometric mean over each choice $s = 2, \ldots, T$ so $\rho^{T+1}(\text{mean}) = \prod_{s=2}^{T} \left(\rho^{T+1}(s)\right)^{1/(T-1)}$

Without structure on the underlying price and quantity data, it would seem that each choice of a linking period s running from $s = 2$ to $s = T$ is an equally valid choice of a period to link the two sets of price levels. The mean splice is a compromise of splices in different periods and thus reduces the risk of splicing on the "wrong" period.[7]

In addition we consider the fixed base expanding window (Chessa, 2016) and fixed base moving window splices (Lamboray, 2017a). Both of these use a fixed base month (say December).

The fixed base moving window (FBMW) approach uses the price change between the base period and each new period t calculated, using $\widetilde{\mathbb{P}}^t/\widetilde{\mathbb{P}}^{base}$, to extend the index to each period in the window. The new value of the index in $T+1$ is

$$\rho^{T+1} = \rho^{base} \frac{\widetilde{\mathbb{P}}^{T+1}}{\widetilde{\mathbb{P}}^{base}}.$$

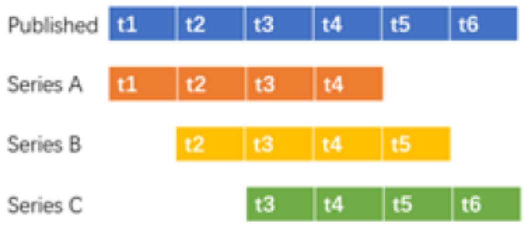
If the base period is December and the new period being calculated is January, then this method is identical to the movement splice.

The fixed base expanding window (FBEW), is similar to the fixed based moving window, but *expands* the window each period to include the latest period of data. For example with monthly data, and a December base month, then the window used to compute the new data point in January includes only December and January. In February it will include December, January and February, and so on until it includes all months in a given window.

[7] Another approach to linking is to use a measure of similarity, the idea being that it makes sense to link windows at the observation which has the most similar price and quantity data; see Diewert (2009), Diewert and Fox (2022) and Diewert (2021a). We do not pursue this approach in this report as we agree with the conclusion of Diewert (2021a) that, while promising, in comparison the use of the mean splice seems "safest" given the current state of knowledge. A potential drawback of similarity linking for NSIs is that the linking period is not known ex ante.

Another approach besides the rolling-window approach is to *splice on the published series* (Chessa, 2021). In this case, rather than splicing on a previously calculated window, we splice on the previously calculated series. We illustrate this with a simplified example in Figure 2.1.

Figure 2.1: *Splicing on the published series*



Consider updating published $blue_{t5}$ to get $blue_{t6}$ using (for simplicity of exposition) a window splice.

Representing splicing as the published end-point times a splicing factor, for the ordinary window splice the splicing factor is $(green_{t6}/green_{t3})/(yellow_{t5}/yellow_{t3})$. Then,

$$blue_{t6} = (green_{t6}/green_{t3})/(yellow_{t5}/yellow_{t3})blue_{t5} \tag{2.1}$$

For the "linking on the published series" approach with the window splice (or "WISP"):

$$\begin{aligned} blue_{t6} &= (green_{t6}/green_{t3})/(blue_{t5}/blue_{t3})blue_{t5} \\ &= (green_{t6}/green_{t3})(blue_{t3}/blue_{t5})blue_{t5} \\ &= (green_{t6}/green_{t3})blue_{t3} \end{aligned} \tag{2.2}$$

Analogous series can be calculated using movement, half and mean splices linking on the published series.

A perceived attraction of this approach is that the year-on-year index of the published series is the same as the year-on-year index of the new window, i.e. $(blue_{t6}/blue_{t3}) = (green_{t6}/green_{t3})$.

However, any standard index (e.g. fixed base Laspeyres) can be used to update the published series and keep the year-on-year growth rate in the published series the same as the new index. That is, we do not need multilateral indexes if that is all that we require. NSI's using scanner data are more likely to be interested in publishing accurate month-to-month indices (e.g. for monetary policy).

Splicing on the published series may be less appropriate for this purpose. The approaches of (2.1) and (2.2) use growth rates between different periods with (typically) different products available. There is likely a loss characteristicity (that is, a greater difference between bilateral and multilateral indices between two periods) from using (2.2) relative to using (2.1). To update $blue_{t5}$, comparing the adjacent green and yellow windows to get the splicing factor in (2.1) uses (adjusted) price growth between periods $green_{t6}$ and $yellow_{t5}$, where the product coverage is likely to be more similar. That is, more similar than using only the green window growth between $t6$ and $t3$ to update $blue_{t3}$, as is done in (2.2), as the green and blue window are more chronologically separated.

All of the linking procedures listed above avoid the need to revise past price levels. But using these linking methods comes at the cost of introducing chain drift bias into the price index. The degree of chain drift bias introduced by this procedure depends on the window length chosen, the link method used and the nature of the underlying prices and quantities. The extent of this can only really be examined empirically. This is what we examine in Section 5, where we compare the difference between a multilateral index computed over all periods in our data with the same index number computed with a linking procedure. The former satisfies the multiperiod identity test and therefore does not suffer from chain drift bias. The difference between the series provides an empirical quantification of the extent to which linking introduces chain drift bias into the price index.

Transitory/seasonal products. Multilateral index numbers have the advantage that they can readily incorporate seasonal products (such as Easter eggs or Christmas cakes). However, they may misstate the importance of such goods in driving monthly price changes (as for example measured by fixed-basket bilateral indices calculated from one month to the next). This is most easily seen for GEKS formulae, which are the geometric means of bilateral indices calculated between each month in a given window. By weighting each bilateral month-to-month index equally, GEKS indices will implicitly downweight the importance of items which are only included in a subset of these comparisons. In other work (available on request), we show that the Geary-Khamis index also has problems in dealing with transitory products. Essentially, the use of these methods should not be seen as a solution to the problem of transitory products.

One solution to this problem is to *impute* the prices of missing items in the months they are no longer available (and thus in which the quantities are zero).[8]

Missing prices can occur due to product churn as well as (seasonal) goods being temporarily unavailable. Restrictions under Covid-19 lockdowns also resulted in missing prices due to goods (e.g. international travel) being unavailable. Imputed prices are estimated prices of goods in periods when they are not sold any more or when they are not yet sold, i.e. in periods when the quantities sold are zero, and hence expenditures are also zero. The Geary-Khamis index is an adjusted unit value index, and so imputed prices cannot affect the index. The same is true for WTPD index with expenditure-share weights; when the expenditure shares are zero, imputed prices will not affect the index. Hence, if there is a purpose in NSIs imputing prices, then Geary-Khamis and WTPD indices seem inconsistent with standard NSI practice.

Comparing index number methods Price indices can be compared across a number of dimensions. One approach is to assess whether a given method satisfies particular common-sense properties ('axioms'). All of the multilateral index number methods we consider here satisfy many of the basic axioms expected of multilateral indices, but none satisfy all; see Diewert (1999), Balk (2001) and Lamboray (2021). However, the relevance of the importance of the various properties represented by the axioms is more contested. Hence the axiomatic approach to choosing a multilateral index number method is inconclusive.[9]

In addition assessing performance relative to these axioms, it might be beneficial if indices met other criteria, such as being easy to explain to the public, or being easy to compute without much human input.

Office for National Statistics (2020) proposes an extensive set of criteria for selecting the appropriate index number method. This takes a broad set of desirable criteria for index numbers and weights them according to their (subjective) importance following consultations with experts and CPI users. Although all the weighted multilateral indices that we consider achieve similar scores in this exercise, the Geary-Khamis method was deemed most desirable, partly as a result of the fact this index satisfies the property of 'additivity'.

[8]A multitude of other methods for dealing with seasonal products are available; see Diewert (2021b) for a comprehensive review.

[9]An alternative approach is the 'economic approach' which assesses how well an index approximates changes in the costs consumers of achieving a given level of welfare over time. Simulations in Diewert and Fox (2022) show that the CCDI index with the mean splice performs better than alternatives at matching changes in the cost of living for realistic consumer preferences, and so would be favoured under this approach.

The weights used in this exercise are inherently subjective and can be contested. For example, while the additivity property of the Geary-Khamis price index means that the corresponding quantity index can be easily decomposed, decompositions for other (price and quantity) indices - including the CCDI index - also exist (Webster and Tarnow-Mordi (2019)).[10] Moreover, while Office for National Statistics (2020) test how highly ranked indices behave in different (simulated) environments, their weighting exercise does not account for other important characteristics such as the degree of chain drift exhibited by different spliced indices in realistic settings or the sensitivity of different indices to the choice of extension method. These properties cannot be discerned *a priori* but must be assessed empirically. We turn to this in the following sections.

3 Household scanner data

We use household level scanner data that is collected by the market research firm Kantar's FMCG Purchase Panel. The data cover purchases of fast-moving consumer goods (FMCG) brought into the home by a sample of households living in Great Britain (i.e. the UK excluding Northern Ireland). This sample includes all food and drinks (including alcohol), as well as toiletries, cleaning products, and pet foods. We use data covering years 2013–2019. In each year the dataset contains purchase records of around 30,000 households. Participating households are typically in the data for many months. Each household records all barcodes that they purchase using a handheld scanner, and they send their receipts (either electronically or by post) to Kantar.[11] For each transaction we observe quantity, expenditure, price paid and barcode characteristics (including product category).[12]

For each of 13 broad product groups we select the top four product categories in each according to spending, and compute price indices. Figure 3.1 shows the full list of products and the shares of spending devoted to each (calculated over the whole period). The price indices we compute are all monthly, and treat the underlying barcodes as the elementary products. To obtain underlying elementary

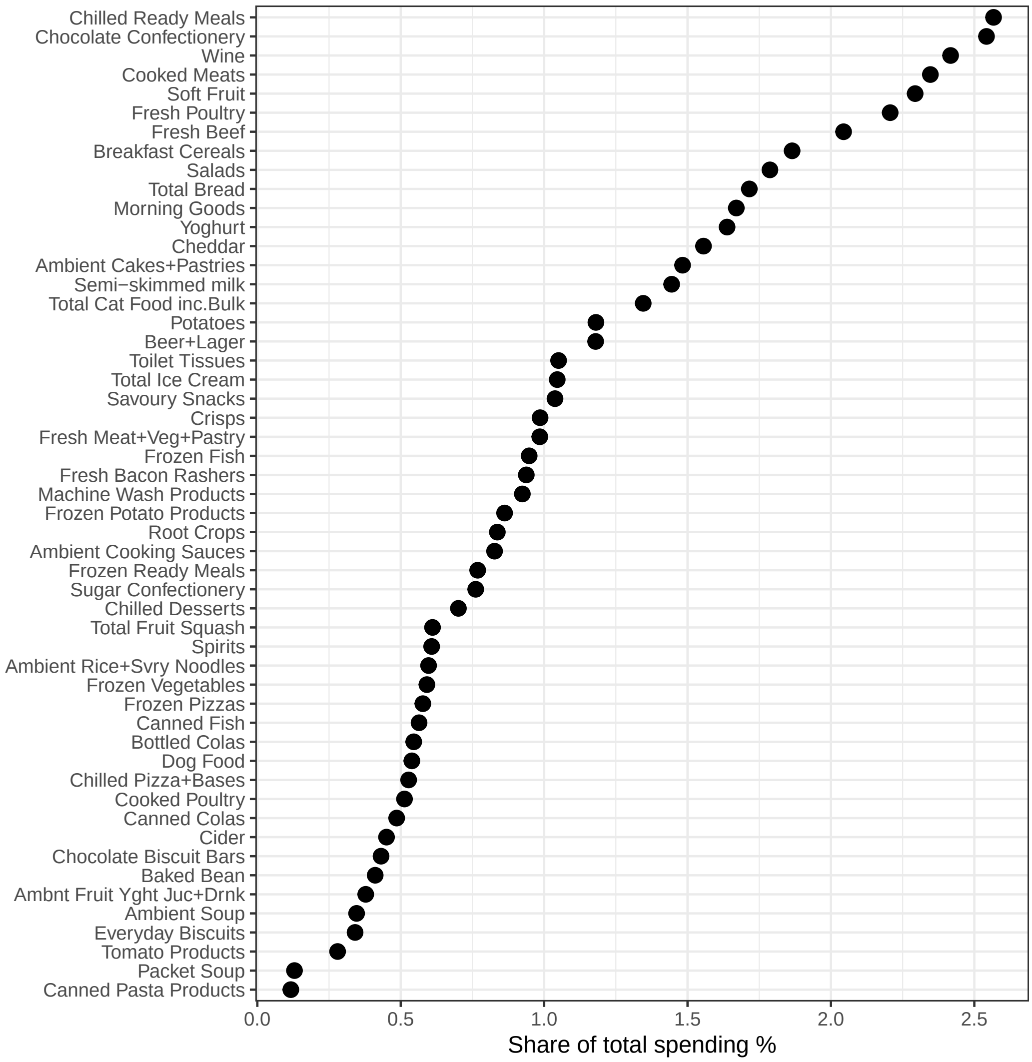
[10] Also, a decomposition of the Geary-Khamis in its entirety is not readily available as standard decompositions do not account for the contribution of transient products.

[11] Non-barcoded items (e.g., loose fruit and vegetables) are recorded by scanning a code in a book provided by Kantar.

[12] NSIs such as the ONS are likely to have access to store level rather than household scanner data in practice. Store level data has several advantages when it comes to calculating index numbers relative to household scanner data, including larger sample sizes, and less of an issue of missing products (which may simply be 'missing' in a particular as no households in the sample are observed buying them, rather than that they are no longer available). These advantages may result in lower chain drift than what we find here.

prices in these indices we take the sum of expenditure and quantity on each barcode in each time period (year-month). Elementary prices are equal to the ratio of total monthly expenditure on the barcode over total monthly quantity.

Figure 3.1: *Share of total spending on product groups*



Note: Shares of spending are calculated over the whole period 2013-2019.

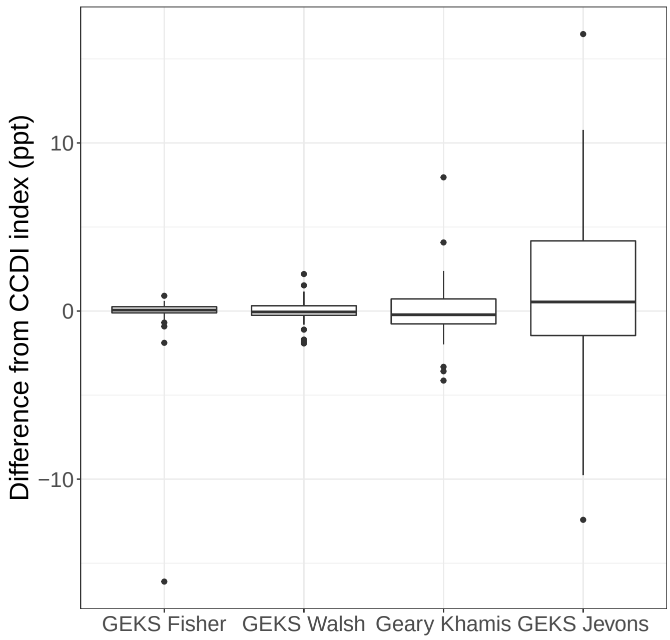
4 Quantifying the difference between indices

In this section we quantify difference in measured inflation across the five multilateral price indices we consider - CCDI, GEKS-Fisher, GEKS-Walsh, Geary-Khamis and GEKS-Jevons. In each case we compute price indices using all 84 year-months of the data, meaning they do not exhibit chain drift bias. For each product category we compute the difference in each of the GEKS-Fisher, GEKS-Walsh, Geary-Khamis and GEKS-Jevons index numbers with CCDI in the last period of our data (December 2019). Figure 4.1 shows box plots for the distribution of differences across product categories (one for each of GEKS-Fisher, Geary-Khamis, GEKS-Walsh and GEKS-Jevons).

The difference in GEKS-Fisher and CCDI price indices in the final period of data are small for almost all product categories. There is one significant outlier, that has a 16.1 percentage point difference between the two indices. This is the product category "spirits". The large difference is a consequence of one particular barcode that has a price that falls by several orders of magnitude half way through our time period. This very large price change is a consequence of measurement error in the data. It has very little impact on the CCDI price index, or indeed the other multilateral indices, but leads to a very large fall in GEKS-Fisher. The differences between GEKS-Walsh and CCDI are small.

Differences between Geary-Khamis and CCDI are typically larger than either the GEKS-Fisher or GEKS-Walsh indices; the 25th and 75th percentiles of differences are -0.8 ppt and 0.7 ppt (compared with -0.1ppt and 0.3ppt between GEKS-Fisher and CCDI, and -0.3 and 0.3ppt between the GEKS-Walsh and CCDI). Unsurprisingly, the unweighted GEKS-Jevons exhibits the largest variances in differences with CCDI.

Figure 4.1: *Index values relative to CCDI index*



5 Quantifying chain drift bias

In this section we quantify the chain drift bias that results from using multilateral price indexes, with either a spliced rolling window or splicing on the published series. We do this by comparing spliced indices with multilateral indices computed using all 84 year-months of data. As the latter do not exhibit chain drift bias, this gives a precise quantification of the extent of chain drift bias that results from using rolling window and splicing methods. We make these comparisons for each product category for each of the CCDI, GEKS-Fisher, GEKS-Walsh, Geary-Khamis, and GEKS-Jevons index numbers. In contrast to the last section, all comparisons made in this section are between different variants (i.e. non-spliced and spliced) of the same index number.

5.1 Splicing methods

We firstly hold the window length at 13 months and compare different splicing methods. We consider the window, half, movement and mean splice, as well as fixed base expanding window (FBEW) and fixed base moving window (FBMW). In addition, we consider the HASP (half splice implemented on the published series). In Figure 5.1 we show boxplots summarising the distribution of differences between spliced and transitive "benchmark" indexes — calculated as one window over the whole sample period — in December 2019 across product categories. Each of the five panels corresponds to a different index number, and each boxplot within a panel corresponds to a different splicing method.

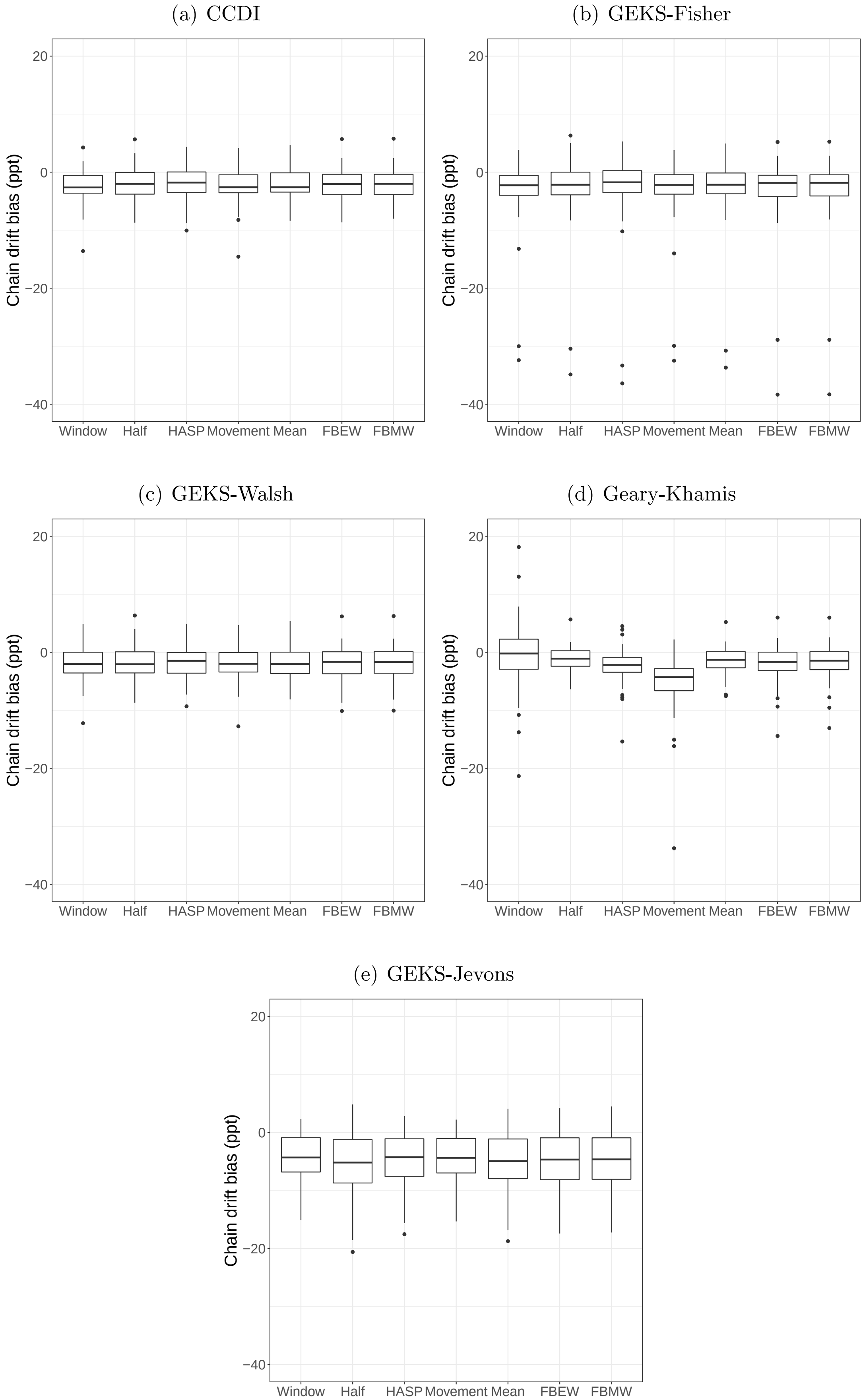
Panel (a) shows that the distribution of chain drift bias is relatively stable across different splicing methods for the CCDI. For each method (except the HASP) at least 75% of product categories exhibit negative chain drift bias, though a minority exhibit positive bias. The median bias ranges from -2.6 ppt for the movement splice to -1.8 ppt for the HASP, and interquartile range ranges from 3.1 ppt for the movement splice to 3.8 ppt for the half splice.

Panel (b) shows that the amount of chain drift bias under GEKS-Fisher index, like CCDI, is relatively insensitive to splicing method, however unlike under CCDI, GEK-Fisher does have some outliers that exhibit very large change drift bias. Panel (c) shows that results for the GEKS-Walsh index are similar to those for CCDI.

Panel (d) shows that Geary-Khamis is more sensitive to the splicing method than CCDI (or GEKS-Fisher and GEKS-Walsh indices). While the distribution of chain drift bias is relatively small (and comparable to the CCDI case) with the half, mean, FBEW and FBMW splices, it is considerably larger for the window and movement splices.

GEKS-Jevons (panel (e)) exhibits considerably more chain drift bias (regardless of splicing method) than any of the weighted price indexes. In contrast to Geary-Khamis, the distribution of chain drift bias under GEKS-Jevons is smallest with the window and movement splice.

Figure 5.1: *Chain drift bias with different splicing methods (13 month window)*



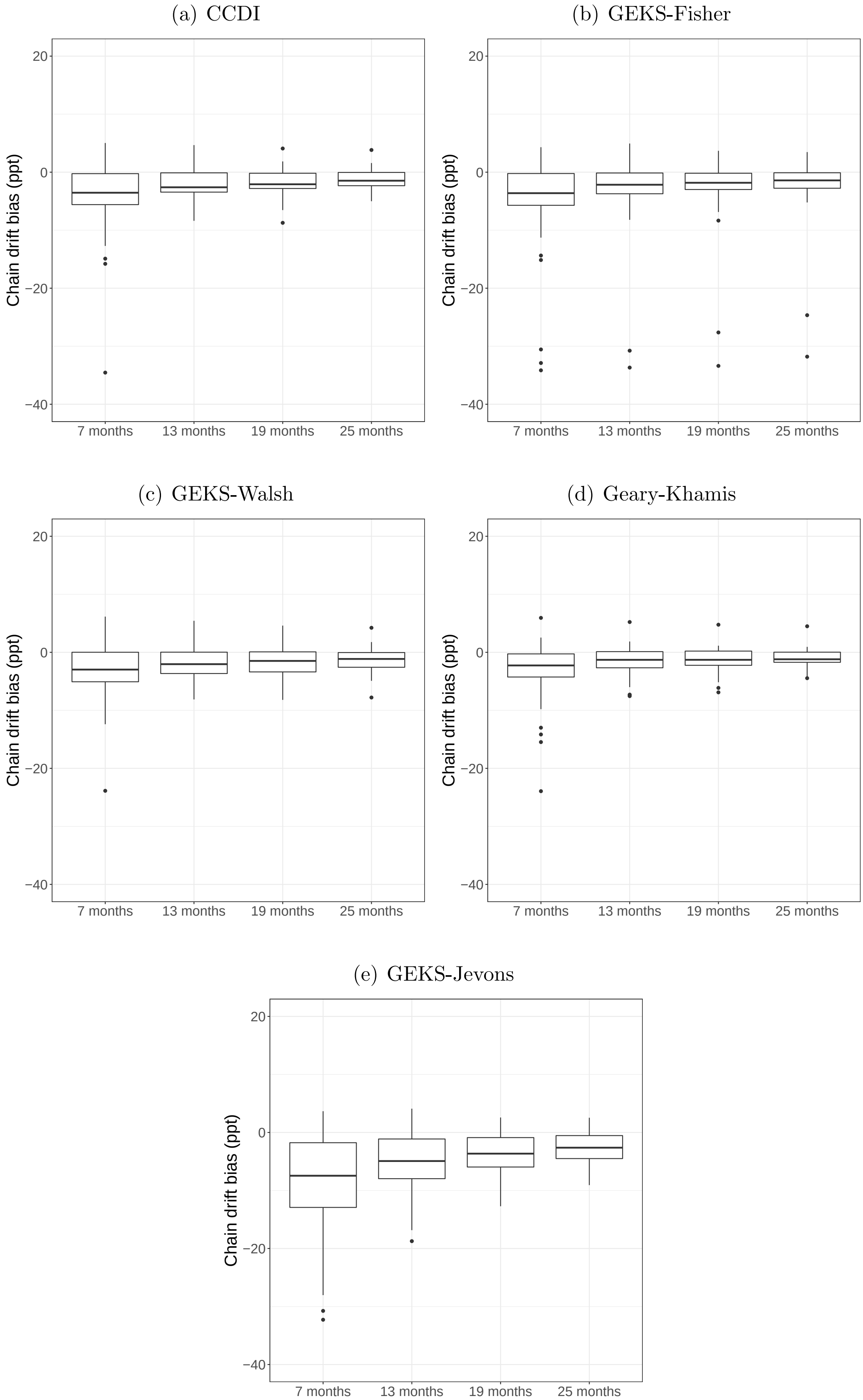
5.2 Different window lengths

Here we compare the impact of different window lengths on chain drift bias, holding fixed the splicing method (using the mean splice throughout). Figure 5.2 summarises the results. It is structured similarly to Figure 5.1 – each panel represents a different index number, and within panel the boxplots correspond to different window lengths. For all index numbers, longer windows lengths lead to considerably less chain drift bias. For a 25 month window (the longest we consider) the distribution of chain drift bias, under CCDI, GEKS-Fisher, GEKS-Walsh and Geary-Khamis are similar.

The rolling window CCDI with the mean splice has a median chain drift bias at 25 months of -1.5 ppt and an interquartile range 2.3 ppt. By contrast, the median chain drift bas for a 13 month window is -2.6 ppt (interquartile range 3.3 ppt). For GEKS-Fisher, while the distribution of chain drift bias shrinks with window length, even at a 25 month window length there remain two big outliers. The median chain drift bias for GEKS-Fisher at 25 months is -1.1 ppt (interquartile range 2.7 ppt). For GEKS-Walsh the median chain drift at 25 months is -1.1 ppt (interquartile range 2.5 ppt), while for Geary-Khamis it is -1.2 ppt (interquartile range 1.75 ppt).

Long window lengths appear to be particularly important for the GEKS-Jevons. Median chain drift bias falls from -4.9 ppt to 2.6 ppt, and the interquartile range declines from 6.8 ppt to 4.0 ppt, as the window length increases from 13 to 25 months.

Figure 5.2: *Chain drift bias using different window lengths (using mean splice)*



Box 3. Summary of evidence on chain drift bias from Section 5

This section has presented estimates of chain drift bias that results from using multilateral price indexes with alternative extension methods. The following key conclusions can be drawn:

1. The CCDI (or 'GEKS-Törnqvist') index seems preferable. In Figure 5.1 it is very stable under all extension methods considered. This lack of sensitivity provides assurance against making a choice of extension method that later turns out to be consequential. GEKS-Walsh provides similar stability, but has the drawback that the use of the Walsh index rules out the possibility of (quality adjusted) price imputations for missing products.

2. The mean splice seems preferable. In Figure 5.1, for the weighted multilateral methods (panels (a) to (d)), it appears to be associated with the least outliers (the dots in the figures), and makes the Geary-Khamis index perform like the CCDI index does under all methods (albeit with greater risk that it could be distorted by anomalous results associated with a particular link period). The half splice also does this, but the mean splice provides greater assurance of not splicing only on a period which is peculiar in some respect (e.g. a pandemic-related lockdown period). If the results are approximately the same between the half splice and the mean splice, as here, then that suggests using the mean splice.

3. A 25-month window length seems preferable. Using the mean splice as the splicing method, Figure 5.2 shows that chain drift bias falls significantly as the window size increases, until it is very close to zero across for most products across all multilateral indexes considered.

4. The above three points suggest using CCDI, with the mean splice and a 25-month window.

6 Drivers of chain drift bias

The results from the previous section highlight the importance of long window lengths to reduce the degree of chain drift bias. Chain drift biases decline noticeably from 13 months to 25 months across index number methods.

Here we assess when chain drift bias is likely to be more of an issue.

In particular we consider how chain drift bias at different window lengths relates to three possible drivers:

- **Seasonality in pricing ('weak seasonality'):** These are estimated from fixed effects regressions of log prices on monthly dummies. The degree of seasonality is the difference between the largest and smallest seasonal effects.

- **Monthly churn**: share of spending on products which are new this month. New products are new product codes in the data (this may sometimes reflect minor changes in the nature of products that nonetheless result in the products being assigned a new barcode).

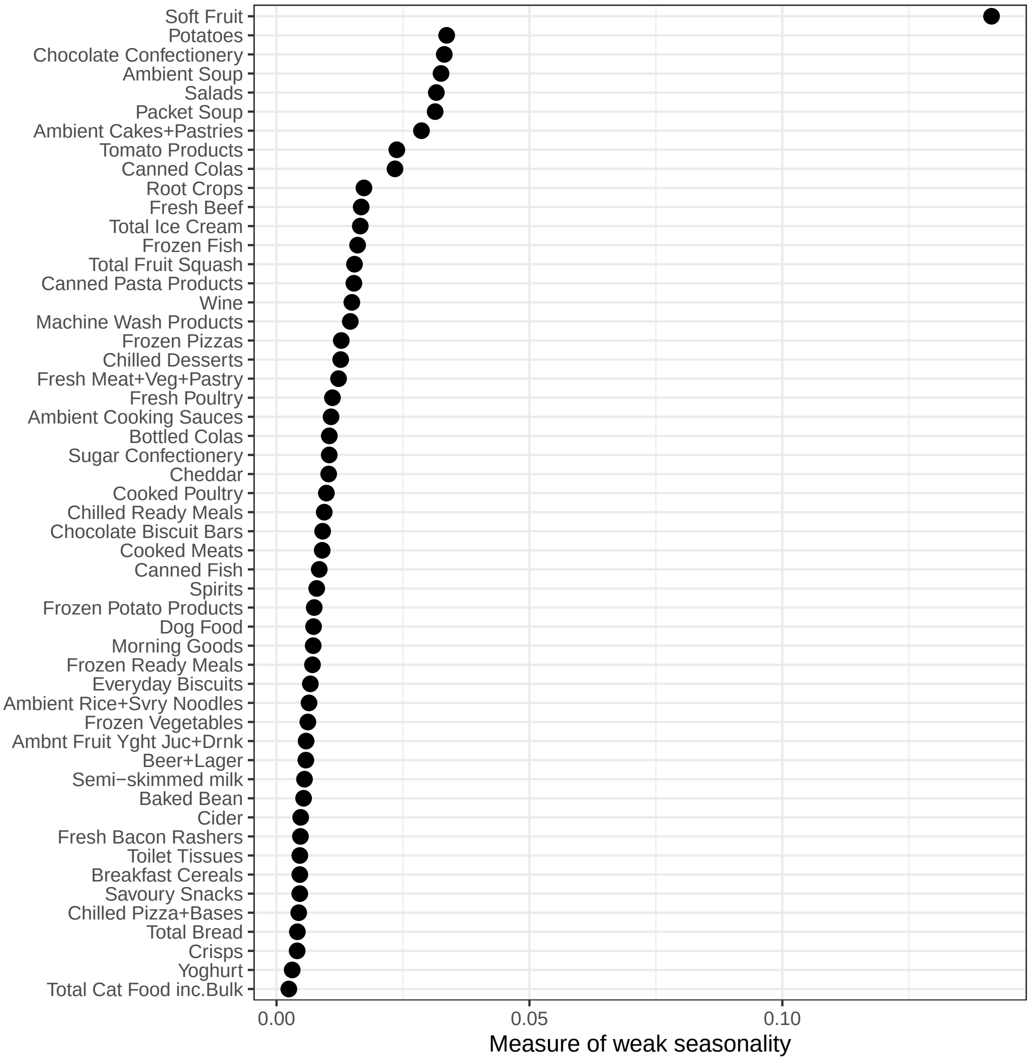
- **Annual churn**: share of spending on products which are new this year. New products are new product codes in the data (although these may sometimes reflect quite minor changes). Run-out sales at the end of product life-cycles have been identified as a potentially important cause of chain drift (Melser and Webster (2021)).

Figures 6.1 - 6.3 show the degree of weak seasonality, monthly churn and annual churn across different product groups.

There is particularly large seasonal price variation for soft fruit, which also shows a high amount of monthly churn. Other goods that show a high degree of monthly churn are wine, spirits and ambient cakes and pastries. It is low for simpler products like beans and milk.

Annual churn is greatest for chocolate confectionary, ice cream and frozen ready meals.

Figure 6.1: *Measure of weak seasonality by product*



Note: Weak seasonality is the difference between the smallest and largest coefficients on monthly dummies from regressions of log prices on monthly dummies and product fixed effects.

Figure 6.2: *Measure of monthly churn by product*

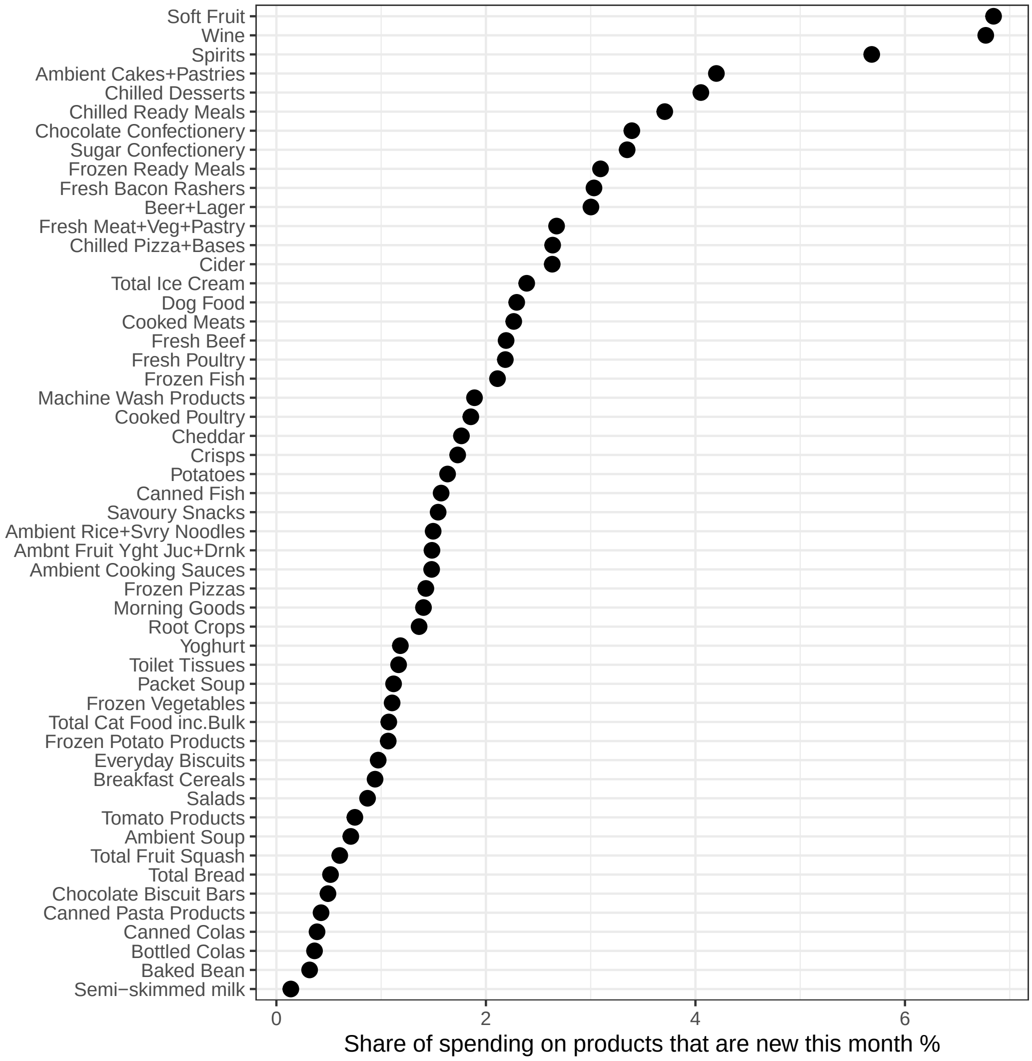
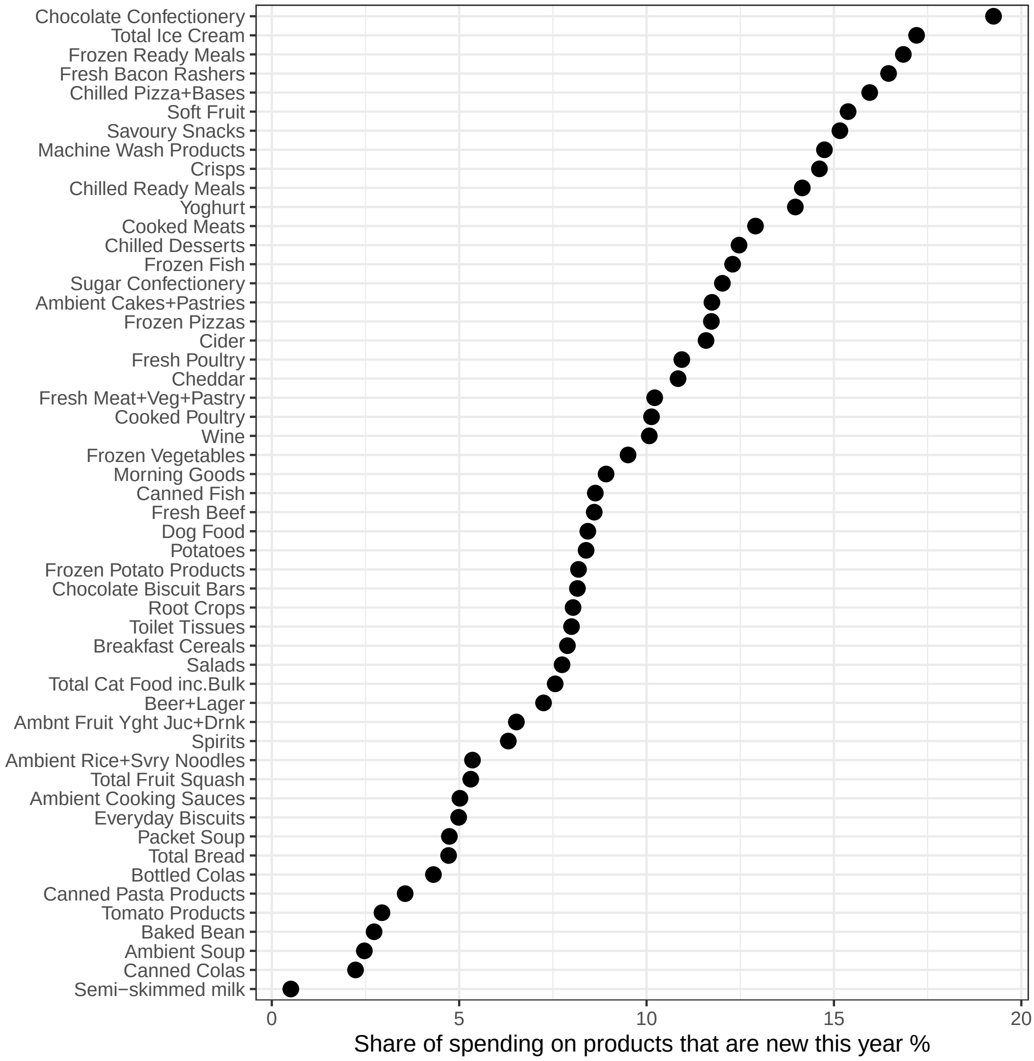


Figure 6.3: *Measure of annual churn by product*



6.1 Seasonality in pricing

Figures A1–A5 in Appendix section A.1 show the relationship between chain drift and seasonality in pricing for the different multilateral indices we consider. In each case there is evidence of an upward sloping relationship for short window lengths (7 months). This relationship is substantially reduced once the window length is extended to 13 months and further still when it is extended to 25 months. We exclude soft fruits as an outlier in terms of its high degree seasonal pricing (but report the chain drift bias in the notes to each graph). This product is not associated with particularly high chain drift with any index.

6.2 Monthly churn

Figures A6–A8 in Appendix section A.2 show how chain drift bias varies according to the amount of product churn relative to the previous month. The pattern is similar to that seen for seasonality in pricing; for short window lengths higher monthly churn is associated with more chain drift bias, but at longer window lengths there is little evidence of a relationship between monthly churn and chain drift.

6.3 Annual churn

Figures A11–A13 in Appendix section A.3 show how chain drift bias varies according to the amount of product churn relative to the previous year. As a comparison of Figures 6.2 and 6.3 shows, the patterns of annual and monthly churn are quite different across products. Chain drift bias appears to be more of a problem for products with greater annual churn, but as with monthly churn this problem and its relationship with measures of churn are greatly reduced by using a longer window length when splicing.

The extent to which the product churn considered here is representative of what an NSI might face is unclear. For example, the ONS uses 'stock-keeping units' (SKUs) for product identification, which offer a broader product definition than barcodes. Further, the ONS carrys out a product re-launch linking exercise to further increase matching between products. Hence, our results may exaggerate problems relating to product churn relative to the context faced by the ONS.

7 Results for a single product category: Chocolate and confectionery

In this section we consider time series for a single product category for which transitory products appear to particularly significant: chocolate and confectionery.

We start by considering how the index is affected by the choice of different window lengths (given the use of the mean splice). We show this for different indices in Figure A16 of the Appendix section A.4. These figures illustrate the extent of chain drift bias, especially when short window lengths are used. The use of a 7-month window length is associated with chain drift biases of around 20-30 percentage points over a six year period (depending on the index). With a 25 month window, the amount of chain drift bias falls considerably across all indices – to around 5 percentage points for all indices.

Figure A17 in Appendix section A.4 shows indices for different splicing methods using a 13 month window. Across indices the movement and window splices do worst, while the half splice tends to do best. The mean splice and HASP perform similarly and only slightly worse than the half splice. FBEW and FBMW tend to perform worse that the mean splice. As before, we find that Geary-Khamis tends to be more sensitive to the choice of extension methods than other indices.

This highlights an important point regarding the choice of extension method. It may be the case that different splicing dates perform better than others, depending on context. This may depend on the nature of annual churn (for example when new products tend to enter and exit). In the absence of a good reason to prefer one splicing period over another, however the mean splice provides a compromise over different possible splicing dates.

8 Conclusions and avenues for future research

The use of multilateral index numbers with scanner data was suggested as a method for controlling the dramatic chain drift that can arise if standard bilateral index numbers are used, as illustrated in Box 1 of Section 2. Hence, the focus on this report is on the ability of different multilateral index numbers and extension methods to control for chain drift. That is, this report examines methods that will enable the use of massive amounts of data to enhance the quality of the UK CPI.

Using UK data, our findings are consistent with what is found in much of the emerging empirical literature on this topic internationally. We also provide some new perspectives. Our key results can be summarised as follows.

1. The ONS (2020) framework for assessing multilateral methods relied on weighting selected index number properties. All the indices examined here received similar scores, with the Geary-Khamis index favoured slightly over the CCDI index because it's supposed ease of explanation to a "non-technical audience" and because the Geary-Khamis index can be additively decomposed. However, it is questionable that the Geary-Khamis index is any easier to explain to the public than other indices, and decompositions are available for non-additive indices including CCDI. Moreover, the CCDI (with the mean splice) has other attractive properties not accounted for within this framework.

2. We find that the Geary-Khamis method is very sensitive to the extension method. This finding is consistent with the emerging empirical literature.

3. Empirical evidence suggests that either the regular mean splice or the mean splice on the published series are to be preferred. For each index number formula, they give essentially the same results in our applications. This near equivalence is not found for other splicing methods.

4. After extensive empirical examination using the UK data reported here, the GEKS-Törnqvist (or "CCDI") index with the mean splice is found to be the preferred method.

We add two more observations, as follows.

1. The GEKS-Jevons index performed poorly. In cases where the quantities corresponding to the prices are unobservable, such as in the case of web-scraped data, the use of multilateral indexes based on unweighted bilateral comparisons is undesirable. There is a need to either find alternative weights, or to sample goods relative to their economic importance. For example, Theil (1967) showed that if price relatives are sampled according to the average consumer expenditure between the two periods being compared, the resulting index is a bilateral Törnqvist index. We believe that attempting to sample the prices of goods relative to their economic importance is currently quite standard NSI practice, and to the extent it is possible, it would seem more appropriate than using an unweighted multilateral index on all collected data. However, more research on this topic is needed.

2. Strong evidence is found in favour of using a 25-month window. This is consistent with international evidence. If a new consumption segment is defined for inclusion into the CPI, such a long window may delay its introduction due to a lack of historical data. A potential strategy in such a case is to sample from scanner data as it becomes available, much as NSIs have done when first starting to utilise scanner data; that is, simply replacing sampling from field collections by using sampling from electronically collected data. Standard bilateral index numbers can then be used as soon as there are two periods of data. Of course, this means discarding a large amount of data for immediate purposes. Once data for 25 months is available, then the price series can continue through using a multilateral method, using all the collected information.

Evidence on the impact of varying the window length and the drivers of chain drift is also provided. We believe that this evidence on the drivers of chain drift is novel to the literature.

Hence, we believe that this report has advanced knowledge of the use of multilateral methods for use in CPI construction. However, as always, there remain some areas for future research. These include the following:

- When does product 'missingness' become a problem? That is, at what point does product entry and exit become problematic to the extent that solutions such as imputation are required?

- Could the choice of splicing method depend on the timing of product entry and exit? Are there some contexts when methods other than the mean splice should be preferred?

- More empirical examination of "similarity" linking methods using different data sets would aid in the understanding of their performance and their potential for implementation by NSIs.

- In the case of web-scraped data, if prices can be sampled according to their economic importance, is it better to apply a bilateral index to the resulting data or a multilateral index?

References

Balk, B. M. (1980). A method for constructing price indices for seasonal commodities. *Journal of the Royal Statistical Society. Series A 143*(1), 68–75.

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

Chessa, A. (2016). A new methodology for processing scanner data in the Dutch CPI. Technical Report No.1/2016 EURONA.

Chessa, A. (2021). Extension of multilateral index series over time:analysis and comparison of methods. Technical report, Department of Consumer PricesStatistics Netherlands.

de Haan, J. and H. van der Grient (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics 161*, 36–46.

Diewert, W. (1999). Axiomatic and economic approaches to international comparisons. In A. Heston and R. Lipsey (Eds.), *International and Interarea Comparisons of Income, Output, and Prices*, Chapter 1.

Diewert, W. (2009). Similarity indexes and criteria for spatial linking. In D. Rao (Ed.), *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, pp. 183–216. Cheltenham, UK: Edward Elgar.

Diewert, W. (2021a). *The Chain Drift Problem and Multilateral Indexes*, Chapter 7.

Diewert, W. (2021b). *Seasonal Products*, Chapter 9. International Monetary Fund.

Diewert, W. E. and K. J. Fox (2022). Substitution Bias in Multilateral Methods for CPI Construction. *Journal of Business & Economic Statistics 40*, 355–369.

Ivancic, L., W. Diewert, and K. J. Fox (2009). Scanner data, time aggregation and the construction of price indexes. Technical report, Department of Economics, University of British Columbia, Vancouver, Canada.

Ivancic, L., W. Diewert, and K. J. Fox (2011). Scanner data, time aggregation and the construction of price indexes. *Journal of Econometrics 161*(1), 24–35.

Lamboray, C. (2017a). The Geary Khamis index and the Lehr index: how much do they differ? Technical report.

Lamboray, C. (2017b). Index compilation techniques for scanner data. Technical report, EuroStat.

Lamboray, C. (2021). Guide on the use of multilateral methods in the HICP. Technical report, Eurostat. Draft paper presented to the Eurostat Workshop on Scanner Data and Webscrapping, 12 October.

Melser, D. and M. Webster (2021). Multilateral methods, substitution bias, and chain drift: Some empirical comparisons. *Review of Income and Wealth 67*(3), 759–785.

Office for National Statistics (2020). New index number methods in consumer price statistics. pp. 49.

Rao, D. (2005). On the equivalence of the weighted country product dummy (CPD) method and the rao system for multilateral. *Review of Income and Wealth 51*, 571–580.

Summers, R. (1973). International comparisons with incomplete data. *Review of Income and Wealth 19*(1), 1–16.

Van Loon, K. and D. Roels (2018). Integrating big data in the belgium cpi. Technical report, STATBEL, Geneva.

Webster, M. and R. C. Tarnow-Mordi (2019). Decomposing multilateral price indexes into the contributions of individual commodities. *Journal of Official Statistics 34*(2), 461–486.

Wells, J. and A. Restieaux (2014). Review of hedonic quality adjustment in UK consumer price statistics and internationally. Technical report.

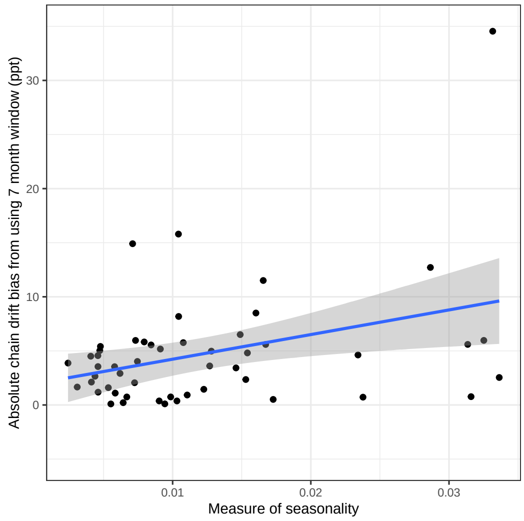
A Appendix

A.1 Figures for Seasonality in Pricing

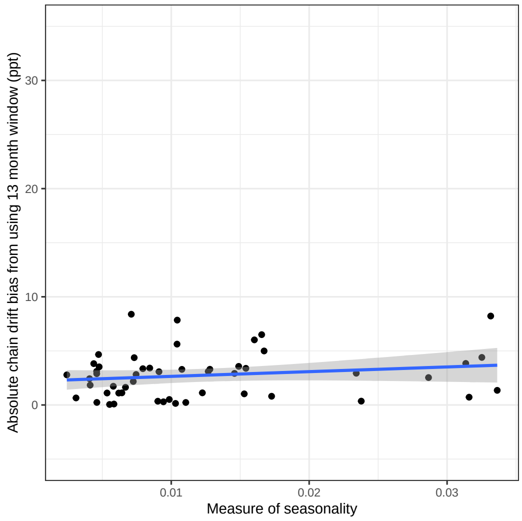
The following figures are referred to in Section 6.1.

Figure A1: *Chain drift in CCDI index vs seasonality in pricing*

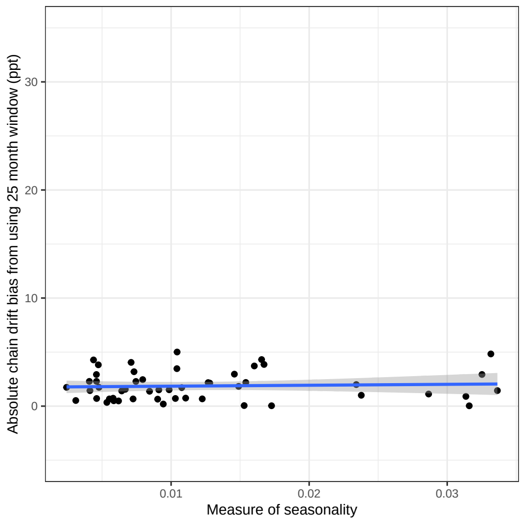
(a) 7 month window



(b) 13 month window



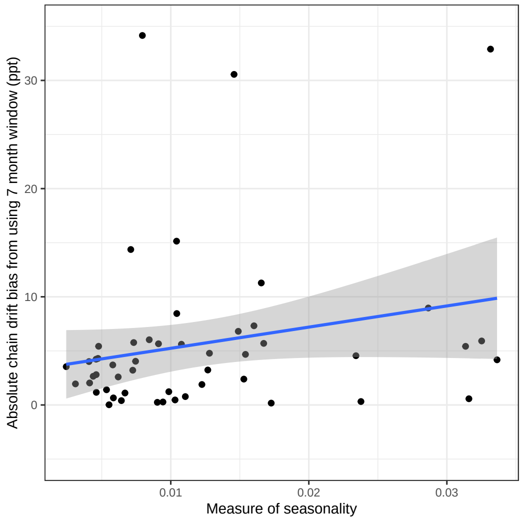
(c) 25 month window



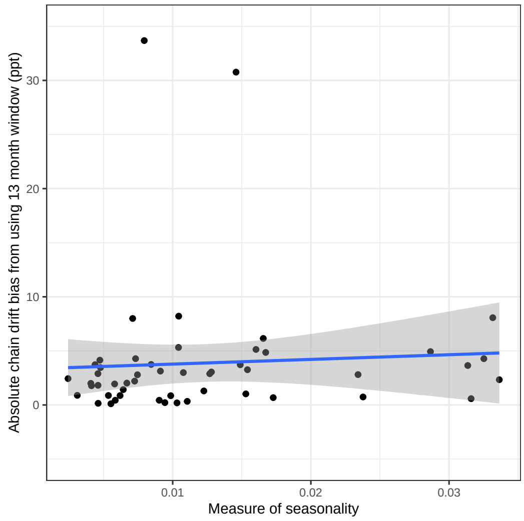
Note: Soft fruits not included (seasonality: 0.14, absolute chain drift bias: 3.97 (7 months), 2.5 (13 months), 1.6 (25 months)).

Figure A2: *Chain drift in GEKS-Fisher vs seasonality in pricing*

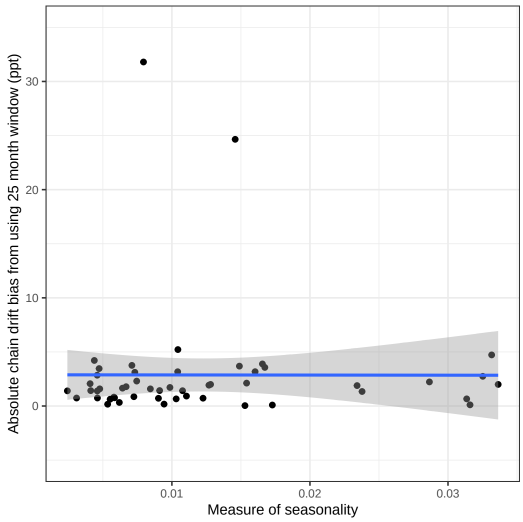
(a) 7 month window



(b) 13 month window



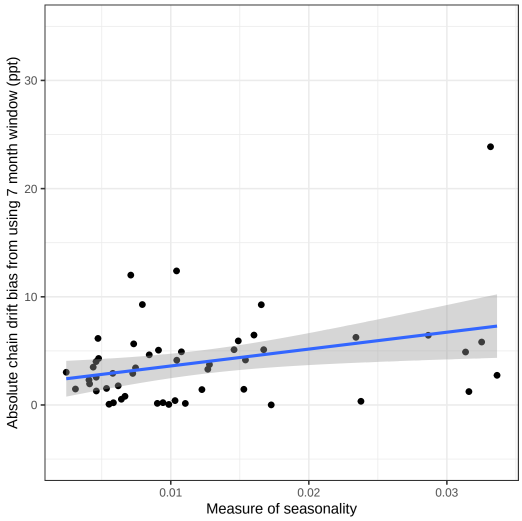
(c) 25 month window



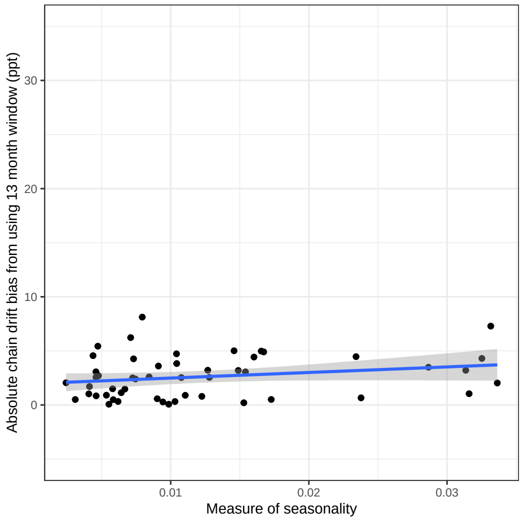
Note: Soft fruits not included (seasonality: 0.14, absolute chain drift bias: 0.6 (7 months), 1.1 (13 months), 0.9 (25 months)).

Figure A3: *Chain drift in GEKS-Walsh vs seasonality in pricing*

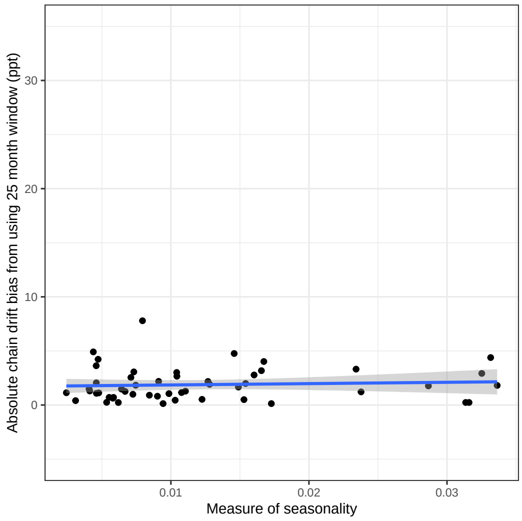
(a) 7 month window



(b) 13 month window



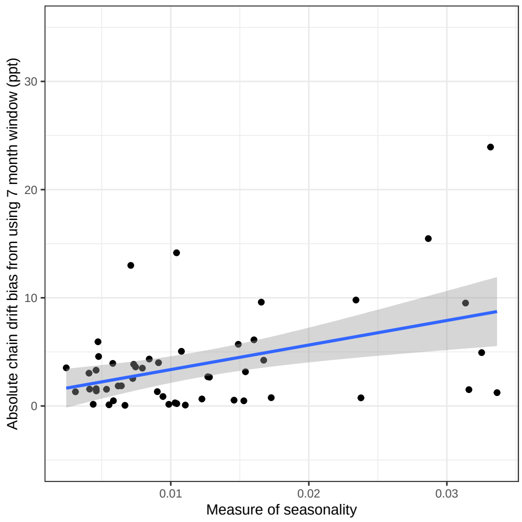
(c) 25 month window



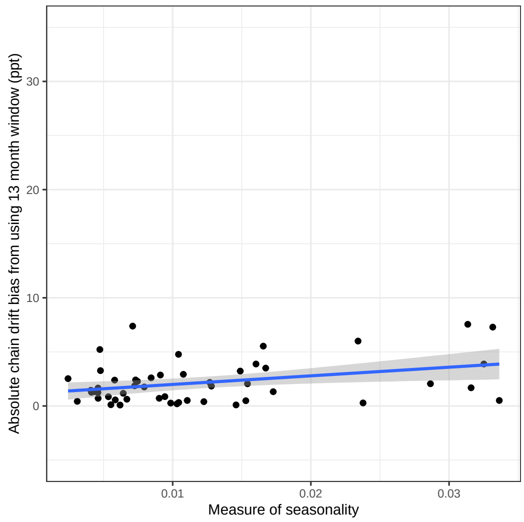
Note: Soft fruits not included (seasonality: 0.14, absolute chain drift bias: 1.3 (7 months), 1.4 (13 months), 1.2 (25 months)).

Figure A4: *Chain drift in Geary-Khamis vs seasonality in pricing*

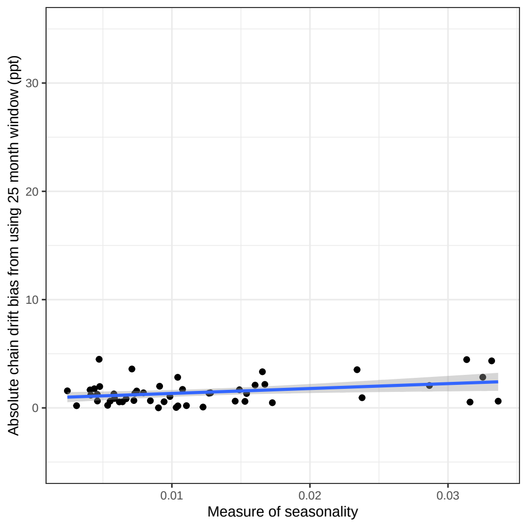
(a) 7 month window



(b) 13 month window



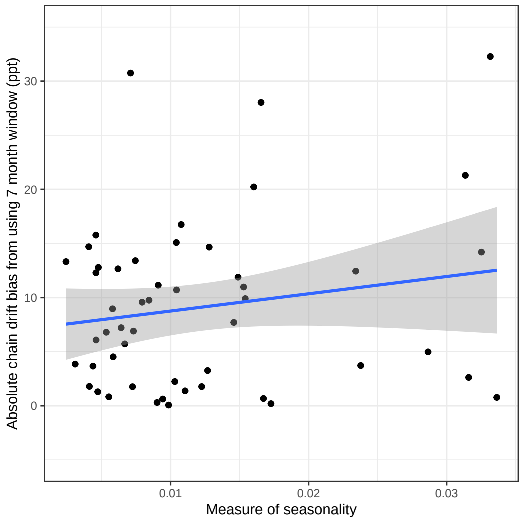
(c) 25 month window



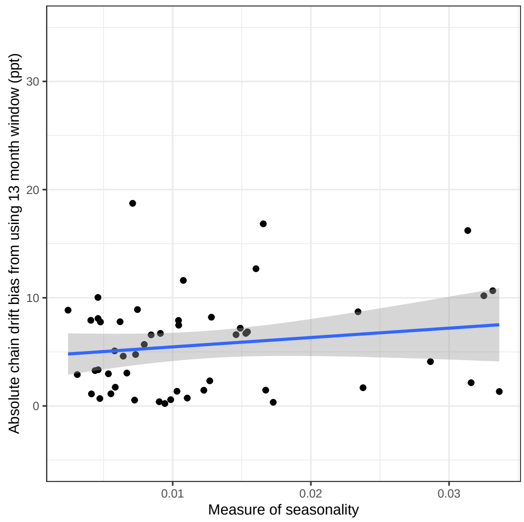
Note: Soft fruits not included (seasonality: 0.14, absolute chain drift bias: 3.90 (7 months), 0.6 (13 months), 0.2 (25 months)).

Figure A5: *Chain drift in GEKS-Jevons vs seasonality in pricing*

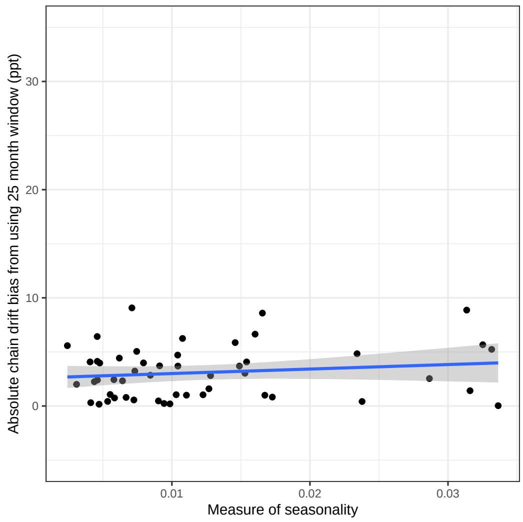
(a) 7 month window



(b) 13 month window



(c) 25 month window



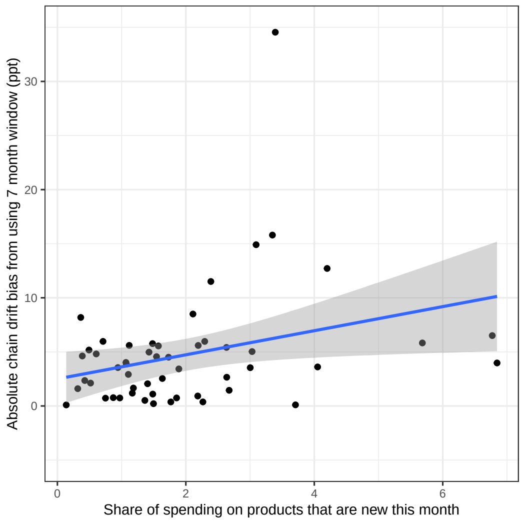
Note: Soft fruits not included (seasonality: 0.14, absolute chain drift bias: 0.05 (7 months), 2.2 (13 months), 1.5 (25 months)).

A.2 Figures for Monthly Churn

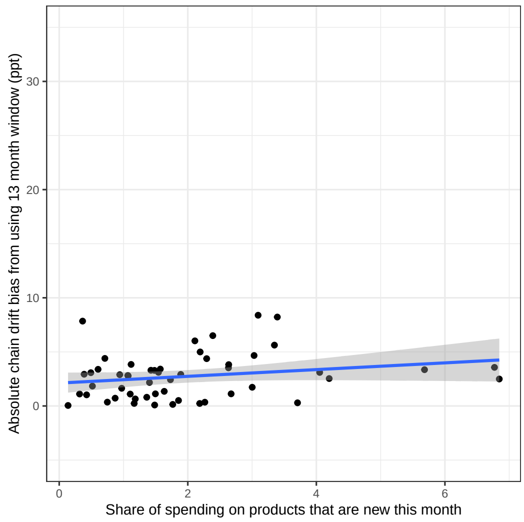
The following figures are referred to in Section 6.2.

Figure A6: *Chain drift in CCDI index vs monthly churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

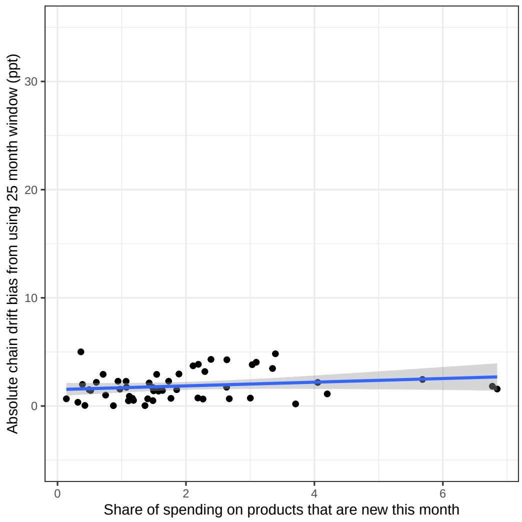
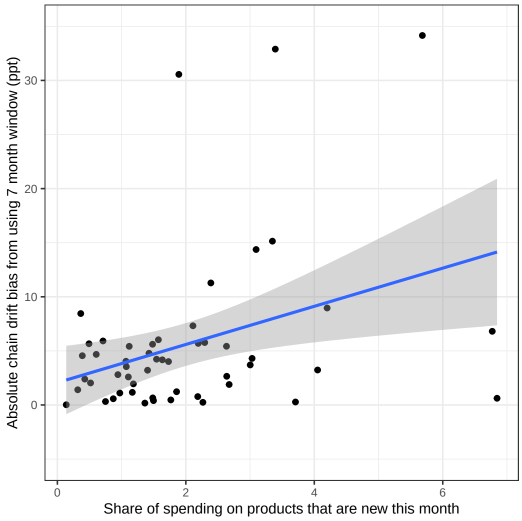
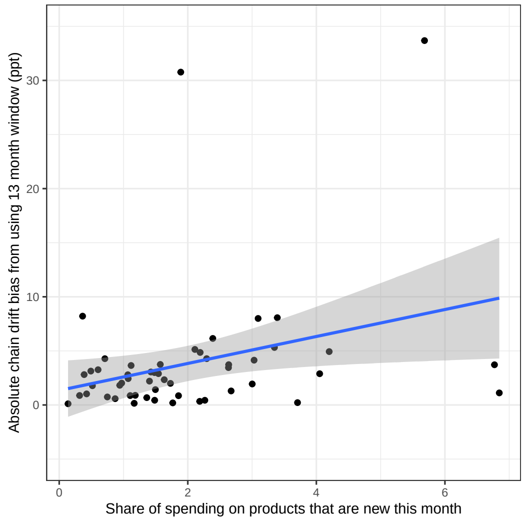


Figure A7: *Chain drift in GEKS-Fisher index vs monthly churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

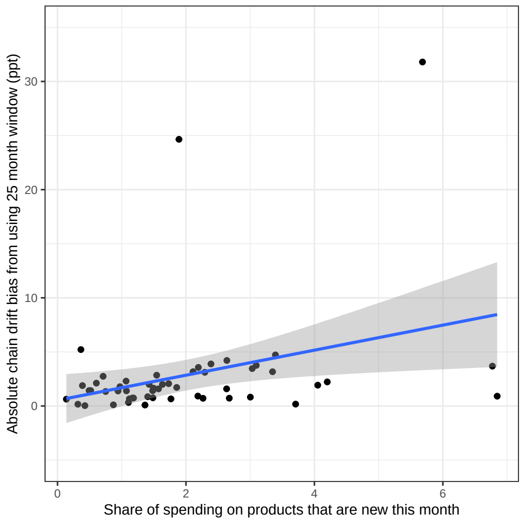
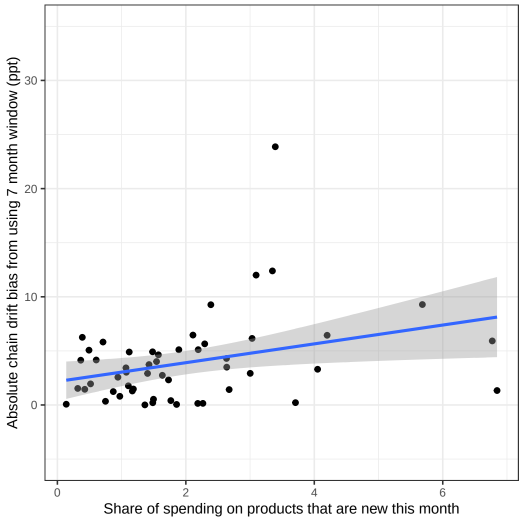
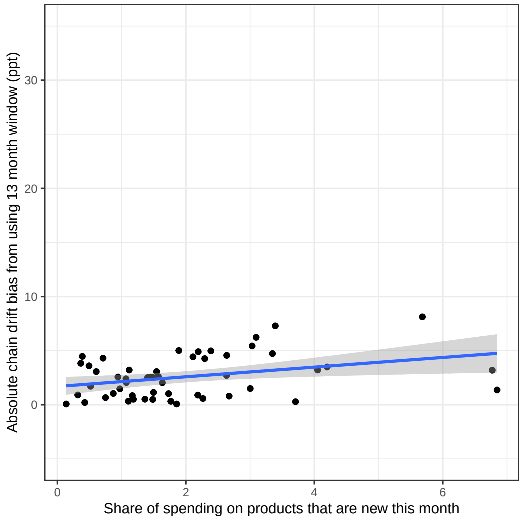


Figure A8: *Chain drift in GEKS-Walsh vs monthly churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

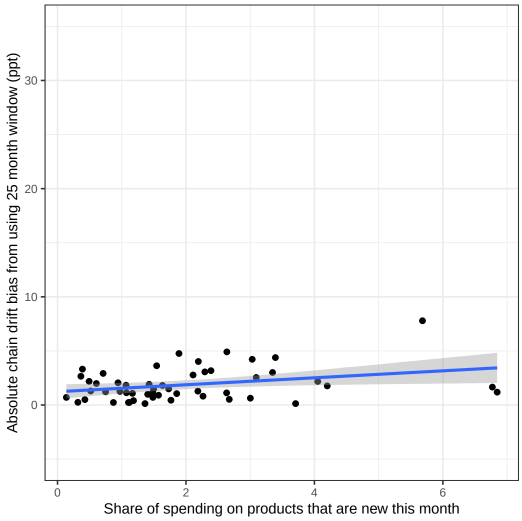


Figure A9: *Chain drift in Geary-Khamis vs monthly churn*

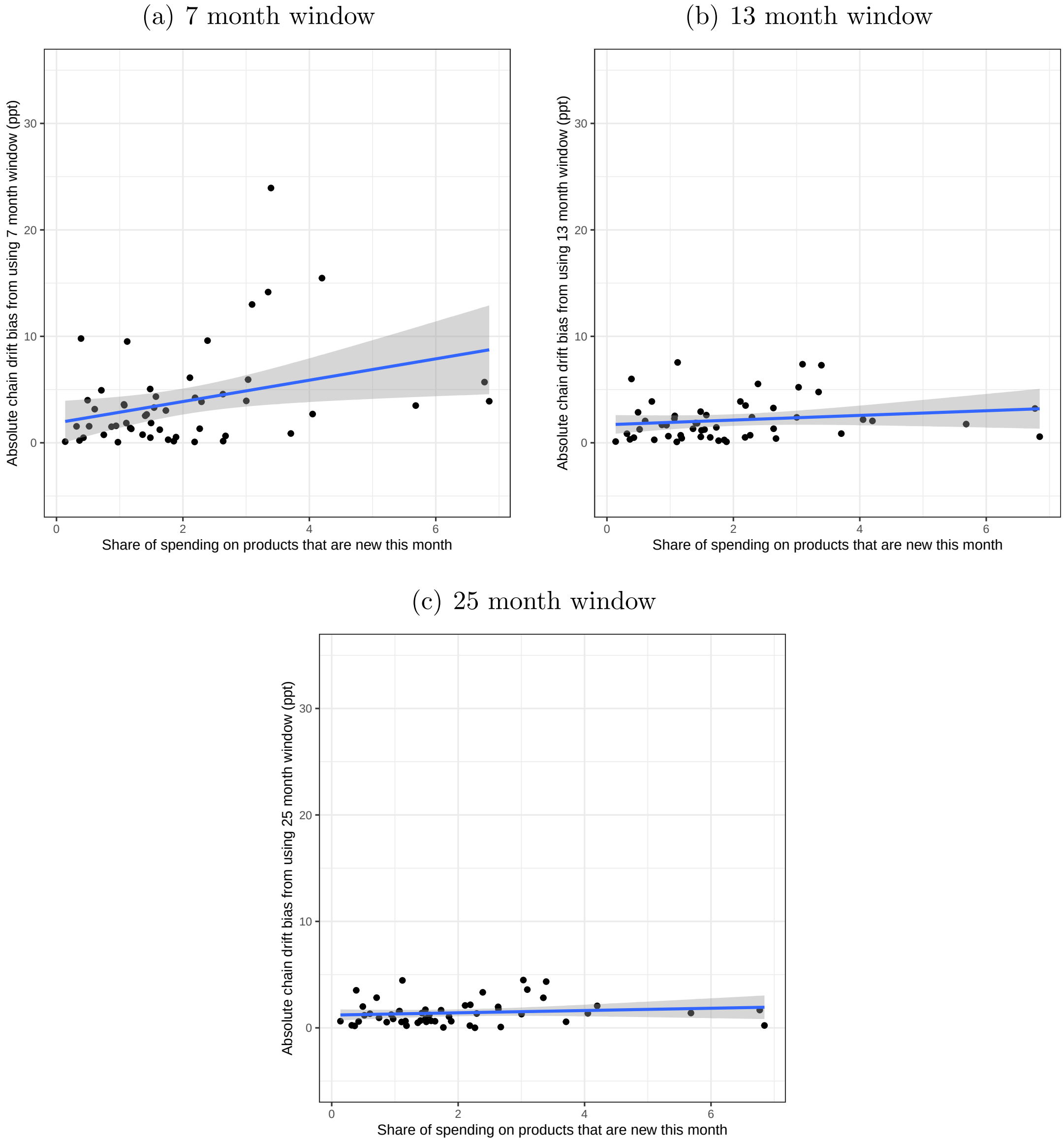
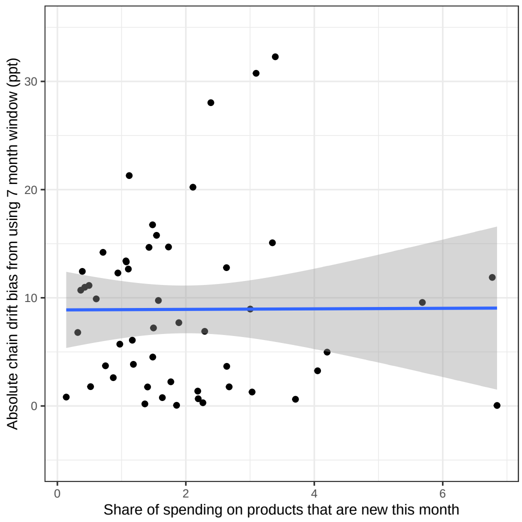
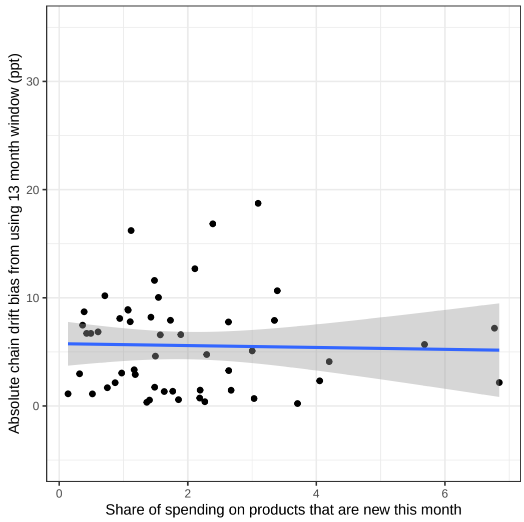


Figure A10: *Chain drift in GEKS-Jevons vs monthly churn*

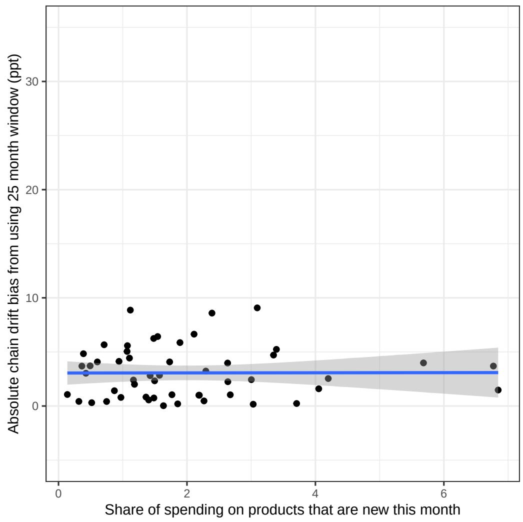
(a) 7 month window



(b) 13 month window



(c) 25 month window

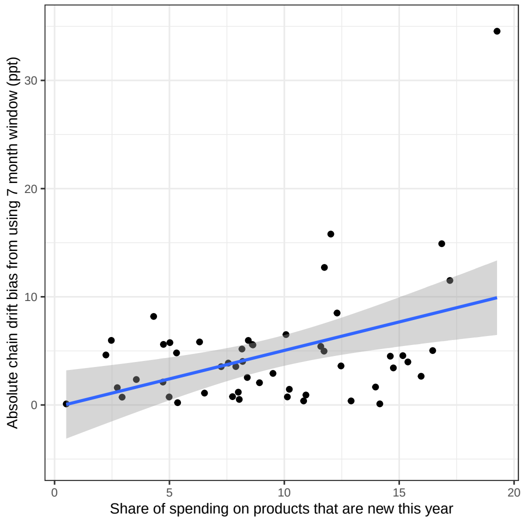


A.3 Figures for Annual Churn

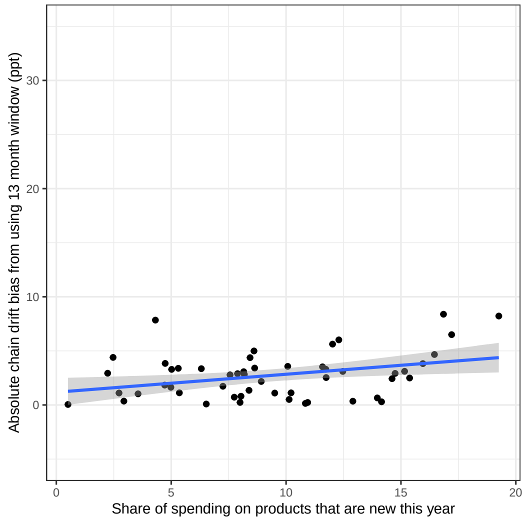
The following figures are referred to in Section 6.3.

Figure A11: *Chain drift in CCDI index vs annual churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

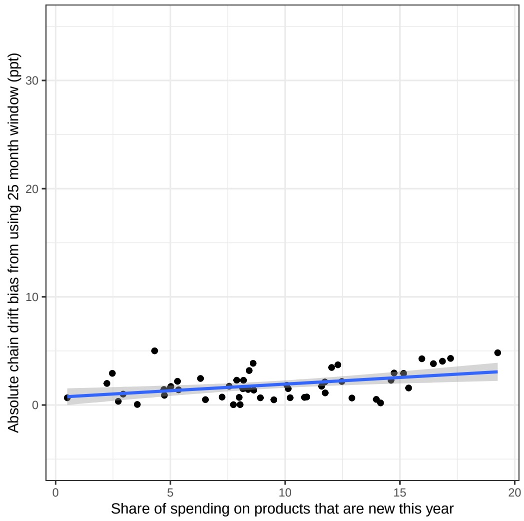
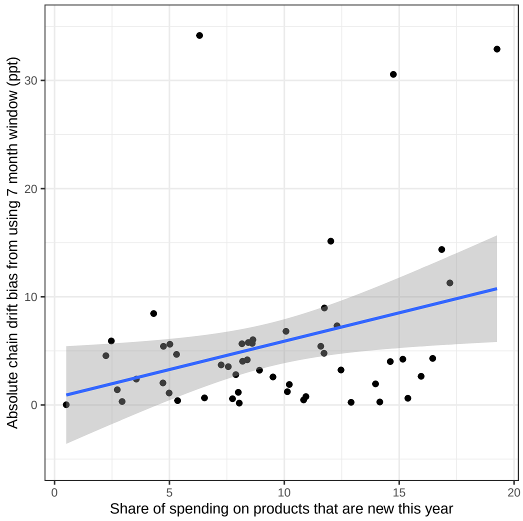
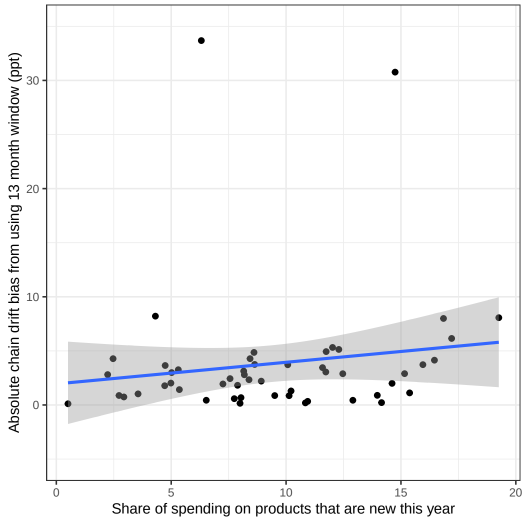


Figure A12: *Chain drift in GEKS-Fisher index vs annual churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

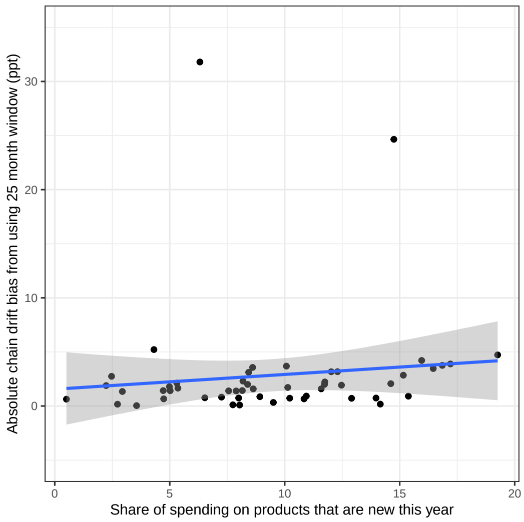
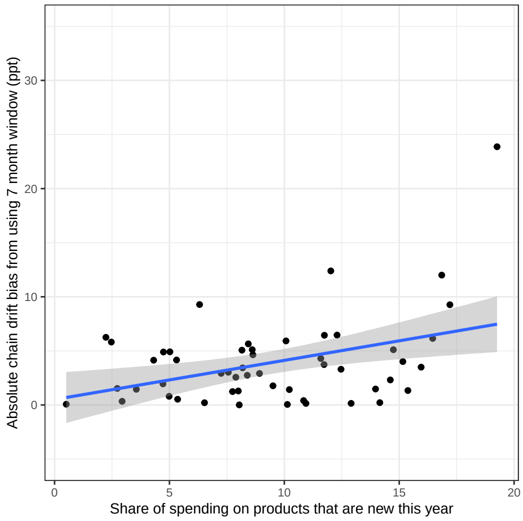
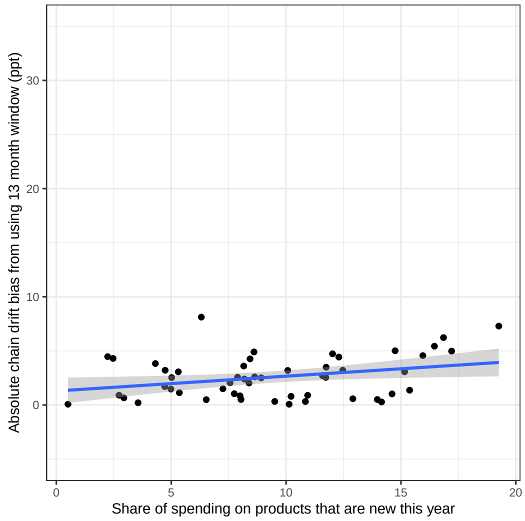


Figure A13: *Chain drift in GEKS-Walsh index vs annual churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

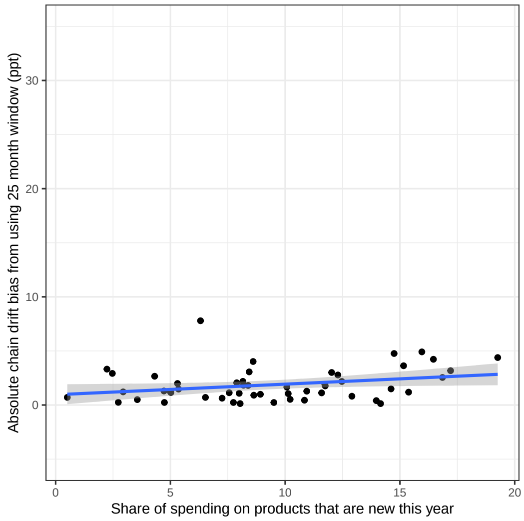
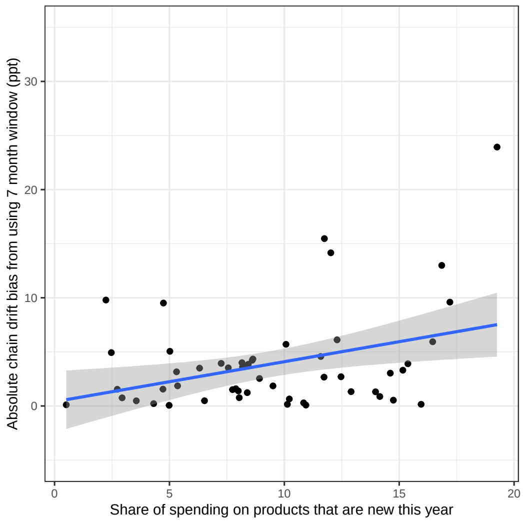
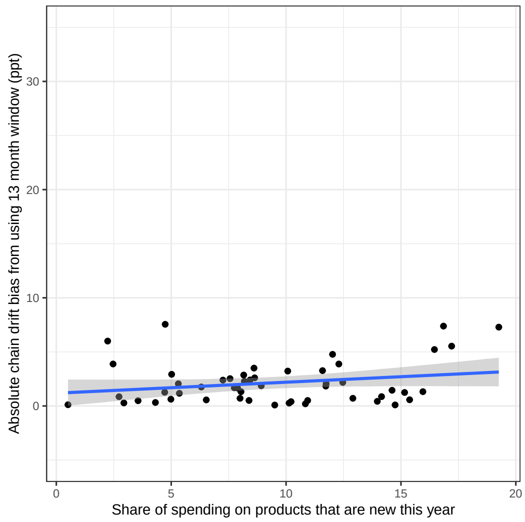


Figure A14: *Chain drift in Geary-Khamis vs annual churn*

(a) 7 month window



(b) 13 month window



(c) 25 month window

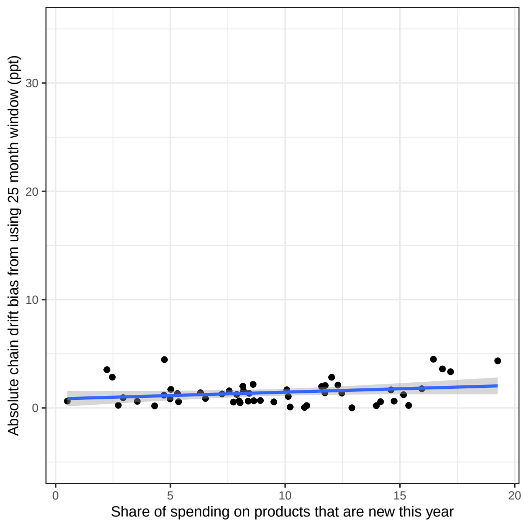
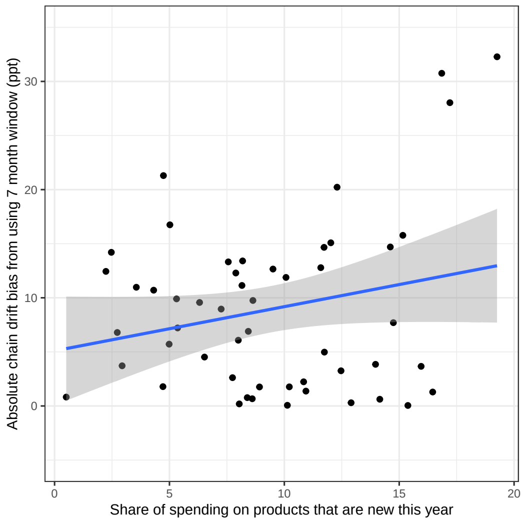
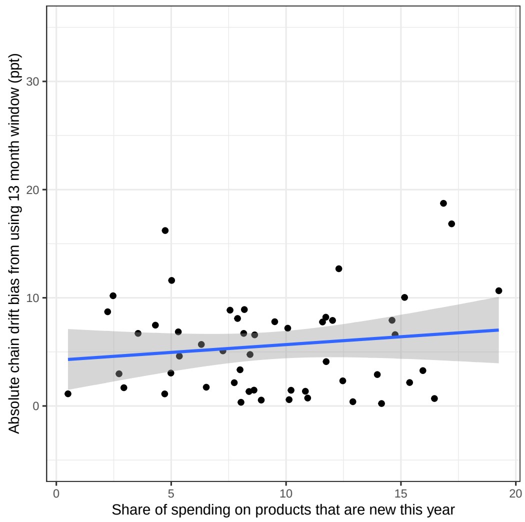


Figure A15: *Chain drift in GEKS-Jevons index vs annual churn*

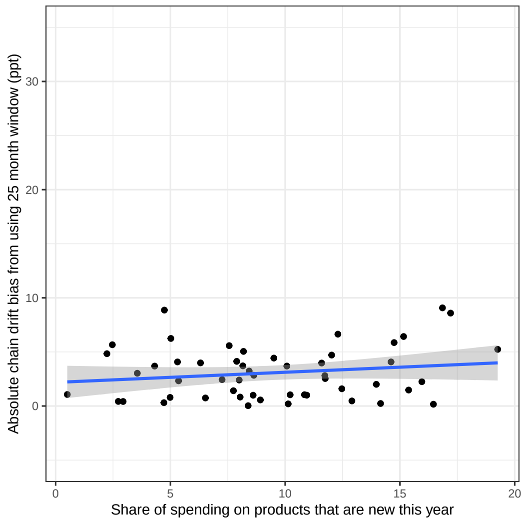
(a) 7 month window



(b) 13 month window



(c) 25 month window

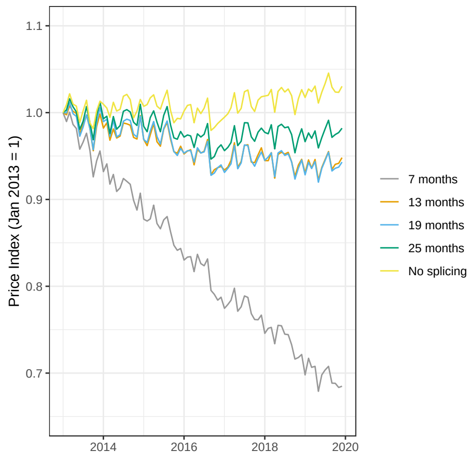


A.4 Figures for a Single Product Category: Chocolate and confectionary

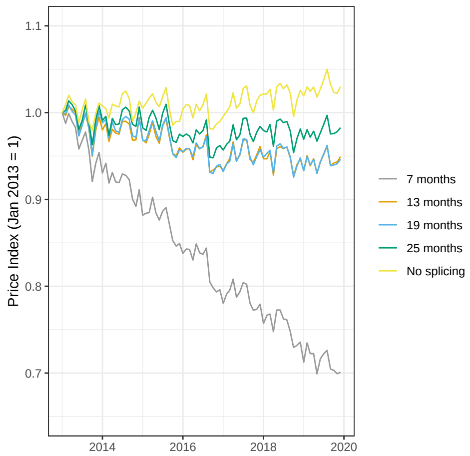
The following figures are referred to in Section 7.

Figure A16: *Chocolate confectionery: different window lengths (using the mean splice)*

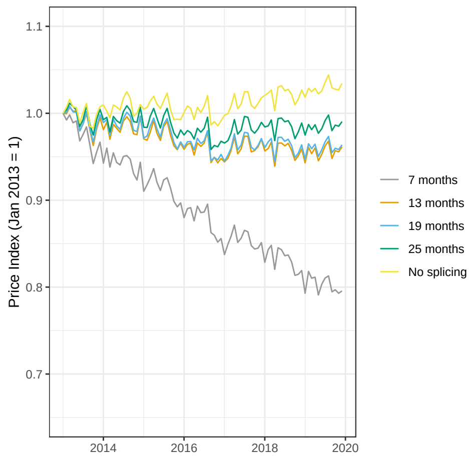
(a) CCDI



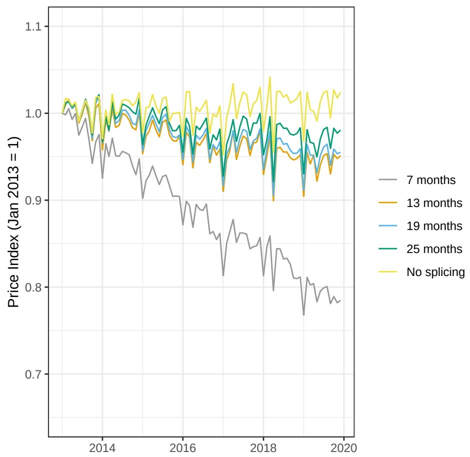
(b) GEKS-Fisher



(c) GEKS-Walsh



(d) Geary-Khamis



(e) GEKS-Jevons

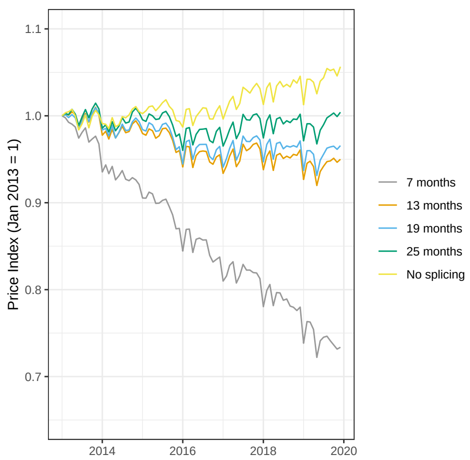
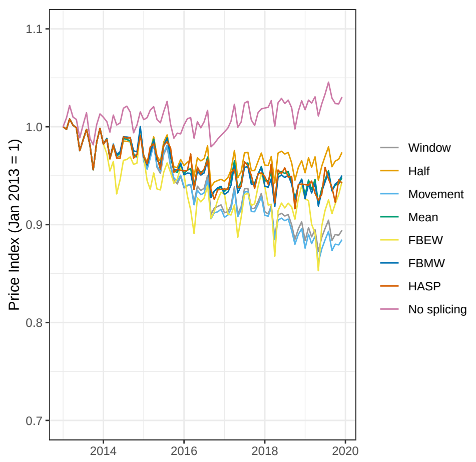
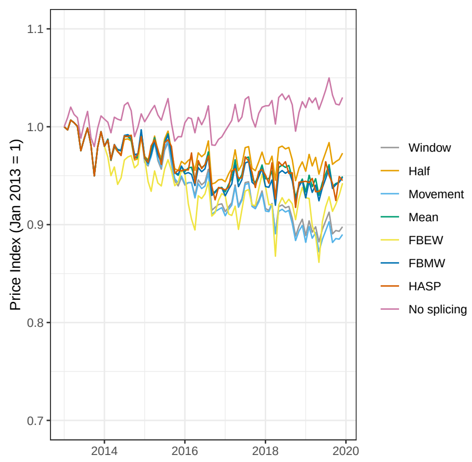


Figure A17: *Chocolate confectionery: different splicing methods (using a 13 month window)*

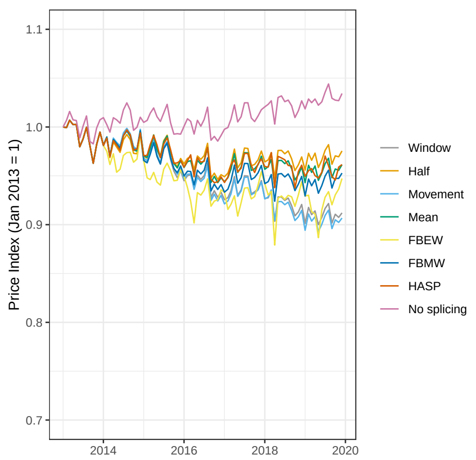
(a) CCDI



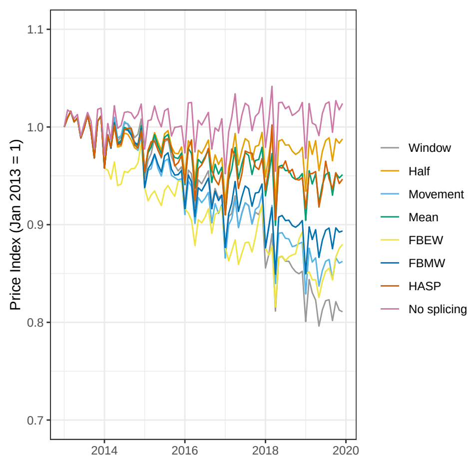
(b) GEKS-Fisher



(c) GEKS-Walsh



(d) Geary-Khamis



(e) GEKS-Jevons