

Article

# New index number methods in consumer price statistics

Research into the use of new index number methods to calculate price indices using web-scraped and scanner data.

Contact:  
Helen Sands  
cpi@ons.gov.uk  
+44 (0)1633 456900

Release date:  
1 September 2020

Next release:  
To be announced

## Table of contents

1. [Main points](#)
2. [Introduction](#)
3. [Overview of index number methods](#)
4. [Framework for shortlisting index number methods](#)
5. [Stress-testing shortlisted index number methods](#)
6. [Case study using real-world data](#)
7. [Conclusions and future work](#)
8. [Related links](#)
9. [Annex A: Technical descriptions of index number methods](#)
10. [Annex B: Stress-testing of additional features](#)
11. [Annex C: Case studies of indices produced using real-world data](#)

# 1 . Main points

- We have produced a quality framework and carried out subsequent testing of different methods that can be used to produce consumer price indices at the lowest level of aggregation (index number methods), including emerging multilateral methods that are in use by other national statistical institutes (NSIs) when using scanner and web-scraped data; this research will help us to understand whether different index number methods are needed for different data sources (for example, scanner and web-scraped data) and different spending categories (for example, groceries, clothing or package holidays).
- Our framework and subsequent testing have shown that at the lowest level of aggregation, consumer price indices produced using multilateral indices for web-scraped and scanner data will be more comprehensive and accurate than those made using fixed-base or chained bilateral methods; Quality-adjusted Geary Khamis (QU-GK) was the highest-scoring method against our index number method framework's criteria, and it performed well under our testing; therefore, the QU-GK method is our current proposed method to use with scanner and web-scraped data when expenditure information, or approximations thereof, are available.
- In the absence of expenditure information, we propose the GEKS-Jevons to be a suitable alternative, although we stress the importance of research into [approximating product-level expenditures](#) as a more effective approach for use with web-scraped data, provided suitable expenditure approximations can be made.
- The move to use scanner and web-scraped data for consumer price statistics is a current point of interest among many other NSIs, and we will continue to produce indices for our favoured methods in parallel while research, both in the UK and internationally, progresses; this means that should international consensus converge on a preferred method, we will likely follow international guidance and best practice.

## 2 . Introduction

We are investigating automated ways of collecting data for consumer price statistics with increased product coverage and frequency of collection relative to our traditional data sources. The data sources we plan to use, by 2023, are [scanner and web-scraped data](#). These new data sources have several features that mean that new price index number methods may be required to maximise their use in our statistics. However, there is an ever-growing number of index number methods that can be used, with little international consensus, currently, as to what the optimal method is.

The choice of which index number methods to use at the lowest level of aggregation will depend heavily on the data source and information available as well as the desired properties of the method itself.

## 3 . Overview of index number methods

### Traditional price collection

Currently, the price quotes used in UK consumer price statistics are collected manually from physical stores, websites and by phoning the retailer or business. This way of collecting prices means that traditional data sources do not typically contain information on the number of each product sold and, as such, most indices use what we refer to as unweighted index number methods at the lowest level of aggregation.

To calculate an index using an unweighted index number method, a sample of products and retailers is chosen that is considered representative of consumer spending in each region and country of the UK. Prices for this same sample of products are collected each month and an average<sup>1</sup> of price movements is calculated to produce an initial price index, known as an elementary aggregate (EA) index. For example, a price collector will observe the price of a commonly bought loaf of white bread in London every month. An EA index is created for bread, based on the average price movement of all loaves of bread sampled in London. Above this EA level, expenditure weights are used to aggregate price movements of London bread together with price movements of bread in other regions and countries of the UK. Price movements for all bread are then weighted together with price movements of all other goods and services in the UK in calculating the headline rates of consumer price inflation.

For more information of how prices are collected and price indices are constructed in our current measures of consumer price inflation, please refer to our [Consumer Prices Indices Technical Manual](#).

## New data sources

Web-scraped data are collected from retailers' websites. Compared to traditional data sources, they can be collected more frequently, cover a broader range of products, and can contain a wealth of information about the product and its attributes. A monthly price can be calculated for web-scraped data by taking an average of the observed prices for each product across a month. Research is ongoing into how this average should be calculated. Web-scraped data lack information on the number of products sold. This means that for web-scraped data, where we have not been able to [approximate the likely number of each product sold](#) we will need to use unweighted index number methods.

In contrast, scanner data are collected by retailers at the point of sale, providing information on the number and type of products sold. These data allow us to use weighted index number methods, meaning that products that have a higher value of sales will have greater influence over the inflation rate. A monthly price is calculated for each product by taking its total expenditure across each month and dividing it by the number of sales. For example, if a loaf of the retailers' own-brand white bread had a total sales value of £100 in March, and 100 loaves of this type of bread had been sold, we would calculate that the average price of that loaf of bread in March would have been £1. Once we have a price for each product in each month, we can aggregate the price changes for individual products using the total sales value as an indication of how much weight to give each product in the resulting index. This means that, for scanner data, we can use weighted index number methods to produce item-level inflation estimates that account for the number of each product that have been sold.

## Weighting period used in constructing price indices

For weighted indices, we need to decide what period to use the expenditure weights from. For example, to calculate inflation between January and February, the price movements can be aggregated based on sales values for either January, February, or an average of both. Using the first period (in this case, January) would result in a "base weighted" index number method, such as a Laspeyres. Using the second period (in this case, February) would result in a "current-period weighted" index number method, such as a Paasche. Using a combination of the weights in each period would result in a "superlative" index number method, such as a Fisher or a Törnqvist (technical descriptions of the methods included in this article can be found in [Annex A](#)).

In normal economic conditions, consumers tend to substitute towards cheaper products when prices increase or towards discounted products. In a base weighted index, the weights are calculated before this substitution occurs, meaning that a base weighted index could overstate the true cost of living. The reverse is true for a current-period weighted index as the weights are calculated after substitution away from more expensive products has taken place. Superlative indices are better approximations of the cost of living as they better account for substitution behaviour between the base and current periods. Current-period weighted and superlative indices have not been used historically in UK consumer price indices because of a lack of reliable timely expenditure information in the current period.

## Time periods used in constructing price indices

Another consideration is the time periods that should be used in calculating the index. Bilateral methods consider price changes for a consistent sample of products between two time periods, although these time periods are not necessarily consecutive. For example, in the current calculation of consumer price indices, prices in each month of a year are expressed relative to the price of the same product in January of the same year. This current method is referred to as a fixed-base bilateral index number method.

Fixed-base methods only measure price movements for products that were available in the base month, or products that have been used as replacements if the original products are out of stock. Frequent chaining can be used to incorporate new products more regularly, to ensure that new and disappearing products can be accounted for so that the sample remains representative of the market over time. Monthly chaining is where consistent product sets for pairs of months are taken throughout the year and their price movements are chained together to form a continuous series. For example, the price change for a set of products between January and February, a set of products between February and March, and a set of products between March and April would be calculated, and these movements would be chained together to show the overall price change between January and April. While frequent chaining has the advantage over fixed-based methods in that it can account for new and disappearing products, it also typically suffers from a phenomenon referred to as “chain-drift”. This is explored more in [Section 5: Stress- testing shortlisted index number methods](#).

In comparison, multilateral methods simultaneously make use of all data over a given time period. The use of multilateral methods for calculating temporal price indices is relatively new internationally, but these methods have been shown to have some desirable properties relative to their bilateral method counterparts, in that they account for new and disappearing products (to remain representative of the market) while also reducing the scale of chain-drift. Multilateral methods can use a specified number of time periods to calculate the resulting price index; the number of time-periods used by multilateral methods is commonly defined as a “window length”.

## Varieties of bilateral index number methods (comparing two time periods)

All weighted and unweighted bilateral methods that simply compare prices between two chosen time periods can use both fixed-base and chained varieties. Table 1 provides a list of the bilateral methods considered in this article, grouped by the period from which their weights are derived. Technical descriptions of all methods considered in this article are provided in [Annex A](#).

Table 1: Bilateral index number methods under consideration

<b>Unweighted</b>	<b>Base weighted</b>	<b>Current-period weighted</b>	<b>Superlative</b>
Jevons	Arithmetic	Laspeyres Paasche	Fisher
Dutot	Geometric	Laspeyres	Törnqvist
Carli			

Source: Office for National Statistics

### Notes

1. All methods listed in this table can use both fixed-base and chained varieties.

While bilateral methods are relatively simple to understand, they can be problematic in certain conditions, particularly when there are a high number of products entering and leaving the market (referred to as churn). The weaknesses of bilateral methods are demonstrated further in [Section 5: Stress- testing shortlisted index number methods](#).

## Varieties of multilateral methods (comparing multiple time periods simultaneously)

Multilateral methods overcome some of the problems experienced in bilateral methods by simultaneously making use of all data available in all time periods. But while multilateral methods have many advantageous properties compared to their bilateral counterparts, in their purest form they are subject to revisions as newer data become available to inform the calculation of previous periods. For example, a multilateral index calculated for March 2020 could be calculated using price changes in all available time periods between January 2020 and January 2021 (using a 13-month window length). Therefore, as each future month between March 2020 and January 2021 becomes available, there is more information to inform the March index value and it would likely be revised.

Using the same example, at the time of publishing a multilateral price index for March 2020 we would lack the information requirements from the remainder of the window until all data were collected by the end of January 2021. The practical way in which we can extend our time series is known as an extension method. Extension methods can be used in combination with multilateral methods to overcome the need for revisions (details of extension method calculations can be found in [Annex A](#)), which are impractical and undesirable for many users of consumer price statistics. A range of multilateral methods paired with a range of extension methods is considered in this article and is presented in Table 2. Multiple combinations of these methods can be used, for example, the GEKS-Jevons can be combined with a movement splice and Quality-adjusted Geary Khamis can be combined with a fixed-base monthly expanding window.

Table 2: Multilateral methods and extension methods under consideration

<b>Multilateral methods</b>	<b>Extension methods</b>
GEKS-Jevons (GEKS-J) <sup>2</sup>	Direct extension (DE)
GEKS-Törnqvist (GEKS-T)	Movement splice (MS)
GEKS-Fisher (GEKS-F)	Window splice (WS)
Quality-adjusted Geary Khamis (QU-GK)	Half window splice (HWS)
Time product dummy (TPD) <sup>2</sup>	Geometric mean splice (GMS)
Time product dummy hedonic (TPH) <sup>2</sup>	Fixed base monthly expanding window (FBME)

Source: Office for National Statistics

### Notes

1. For the purposes of this assessment we have chosen to use a 13-month window length for all our extension methods. Future research will investigate the impact that extending the window length has on our results
2. GEKS-Jevons is the only unweighted multilateral method considered in this research, hedonic methods can be calculated both with and without weighting information

### Notes for: Overview of index number methods

1. Different methods of averaging can be used, such as geometric (Jevons) or arithmetic (Carli) averaging. Another respected method, known as Dutot, calculates the ratio of average prices, rather than taking the average of price movements.

## 4 . Framework for shortlisting index number methods

In total, the combination of different multilateral and extension methods, along with the fixed-base and chained bilateral methods, gives rise to over 50 potential methods that we could use in our consumer price statistics at the lowest level of aggregation. To decide on an appropriate index number method for each data source (for example, scanner and web-scraped data) and each category of spending (for example, clothing, groceries or package holidays), we intend to complete the following steps.

### Step 1: Shortlist methods

- Exclude methods that do not meet minimum resource requirements.
- Exclude methods that do not meet minimum interpretability requirements.
- Apply theoretical framework to remaining methods: theoretical properties (55%), resource (20%), interpretability (15%) and flexibility (10%).
- Methods that score within the top 10 are chosen for the shortlist.

### Step 2: Assess shortlisted methods

- Test methods against a range of synthetic datasets with different pricing behaviours.

### Step 3: Choose the appropriate method

- Assess pricing behaviour of unique item over given time series.
- Determine the characteristics of the data.
- Assess whether the highest ranked method is suitable given the pricing behaviour and characteristics of unique item.
- If the highest ranked method is unsuitable, choose appropriate alternative from the shortlist.

To limit the number of methods available for use, we have produced a shortlist of methods based on a quality framework of pre-determined criteria. A large number of methods is undesirable as it is both impractical to implement and complex to explain. While a large number of methods is undesirable, a single method may also not be suitable for all data sources and categories of spending. We have therefore produced two shortlists of appropriate methods: one shortlist for use when expenditure weights, or approximations, are available and one shortlist for when this information cannot be obtained.

The framework for assessing index number methods has been discussed with, and informed by, both our [Technical](#) and [Stakeholder](#) Advisory Panels on Consumer Prices (APCPs). The framework will be periodically reviewed and updated in line with our own and international research and guidance as well as with any emerging price index number methods.

There are five criteria that we use to produce our shortlist. Table 3 provides the criteria (with reference to the [European Statistical System's \(ESS's\) quality dimensions](#)) and their respective weights within our framework. Detailed information about the framework, criteria weights and how the index number methods performed can be found in [The winning formula? A framework for choosing an appropriate index method for use on web-scraped and scanner data](#), presented to the Technical APCP in January 2020.

Table 3: Criteria weights within the Index Number Methods Framework

Criterion	Weight
a. Theoretical properties (accuracy and reliability)	55%
b. Resource (timeliness and frequency)	20%
c. Interpretability (accessibility and clarity)	15%
d. Flexibility (relevance)	10%
e. Coherence (coherence and comparability)	0% (used as a secondary filter)

Source: Office for National Statistics

New methods will be assessed against our framework and ranked against existing methods as they emerge. Scores for our existing methods will be reviewed periodically to ensure that they account for the most recent research and developments in the international literature.

Prior to assessing methods against our framework, two primary filters are applied. First, if the information-processing requirements are unmanageable, then the method is excluded as we do not want to hinder the timeliness or frequency of the [consumer price inflation publication](#). Secondly, if the price movements are not intuitive to those producing or using the data, then the method is excluded as we believe any price movements should be understandable to both producers and users.

After applying these primary filters, each method is assessed against each criterion. The final scores are used to rank the methods and produce the shortlists of appropriate index methods for UK consumer price statistics. In cases of equal scores between methods, the cohesion criterion is used as a secondary filter to separate methods in the rankings. For example, if two methods received the same score in the rankings, any method in use by other National Statistics Institutes (NSIs) would take precedence in the shortlist.

Following discussions with our Technical APCP and other index number method experts, we have made some small alterations to the framework scores and resulting shortlists. Our current shortlists for index number methods when weighting information (or approximates thereof) are available and for when weighting information (or approximates thereof) are unavailable are shown in Tables 4 and 5 respectively.

Table 4: Shortlist 1, favoured index number methods when weighting information is available

Rank	Method
1	Quality adjusted Geary Khamis (QU-GK) using a Fixed Base Monthly Expanding window (FMBE)
2	GEKS-Törnqvist using a Movement Splice
3	GEKS-Fisher using a Movement Splice
4	GEKS-Jevons using a Movement Splice
5	GEKS-Törnqvist using a Geometric Mean Splice
6	GEKS-Fisher using a Geometric Mean Splice
7	GEKS-Törnqvist using a Window Splice
8	GEKS-Fisher using a Window Splice
9	GEKS-Jevons using a Geometric Mean Splice
10	GEKS-Jevons using a Window Splice

Source: Office for National Statistics



Table 5: Shortlist 2, favoured index number methods when weighting information is unavailable

**Rank Method**

4	GEKS-Jevons using a Movement Splice
9	GEKS-Jevons using a Geometric Mean Splice
10	GEKS-Jevons using a Window Splice
33	Chained Jevons (CJ)
44	Fixed base Jevons (FBJ)

Source: Office for National Statistics

Tables 4 and 5 show that the multilateral methods consistently outperform the bilateral methods in both shortlists. Our shortlist in Table 4 shows that our unweighted multilateral methods (GEKS-Jevons) outperforms our weighted bilateral methods. Only three unweighted methods ranked within the top 10 methods in our shortlists, as seen in Table 5.

While bilateral methods did not rank well in the framework, we have chosen to include a fixed-base and chained Jevons index in our second shortlist (in Table 5) to provide a comparison to the multilateral methods and ensure that, were we to find that the GEKS methodology was not suitable for a dataset, we could revert to more traditional index number methods were we to deem them appropriate for the dataset in consideration. In the future, we may also consider the hedonic approaches for use when expenditure information are unavailable, but as these methods did not rank within our top 10 they have not currently been assessed as part of this research.

## 5 . Stress-testing shortlisted index number methods

To assess the potential suitability of our shortlisted methods in the production of consumer price indices, we have produced a range of synthetic datasets demonstrating isolated pricing behaviours to stress-test each method's performance. The behaviours we have isolated and include in this section are high attrition rates and product churn, product obsolescence, high variance in prices, and high quantities of product sold.

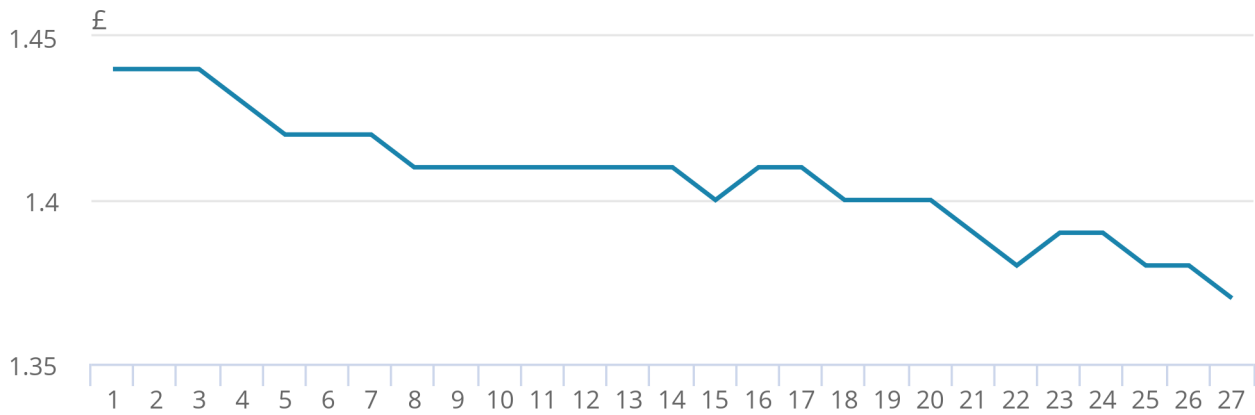
The synthetic datasets were produced through modification of an open source dataset known as [Dominick's Finer Food data \(Dominick's\)](#). These data have been provided by the James M. Kilts Center, University of Chicago Booth School of Business. The data were restricted to a single store, and values with an absent price or quantity were removed before taking a random sample of the data. A model was then fitted to the data to understand the relationship between price and quantity, and this model was subsequently used to build a sythetic dataset. Once the base data had been created, behaviours could be added into the dataset in isolation to see their impact on the resulting indices.

A simple base dataset was initially produced to understand the differences in each method's index values when the dataset shows a static set of products, so all products are available in all time periods, with no changes in the underlying quality of the sample. There are relatively small changes in prices and quantities throughout the 27-month period studied, as shown by the reduction in mean prices between Periods 1 and 27 in Figure 1.



**Figure 1: Mean price across time for the base synthetic data**

Figure 1: Mean price across time for the base synthetic data

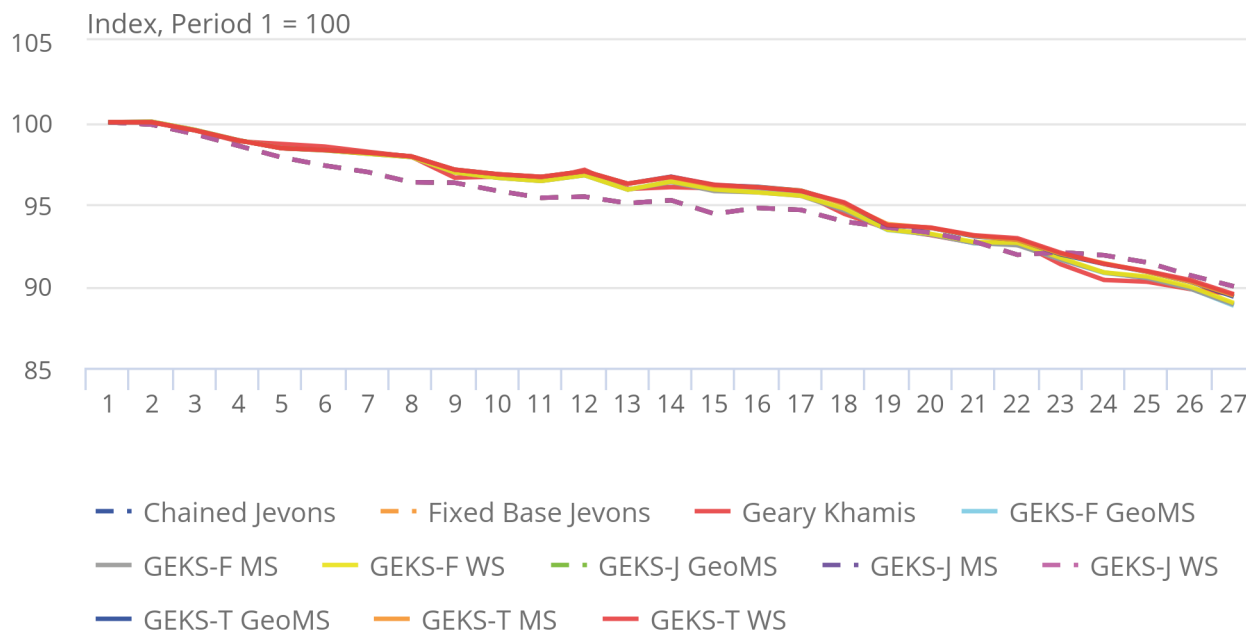


**Source: Office for National Statistics**

Figure 2 provides the index values for our shortlisted methods (listed in Tables 4 and 5) using these base data. For the purposes of this article, all indices that use an extension method use the expanding window for the first 13 months and then begin to use the extension methods thereafter. Fixed-base methods and the Quality-adjusted Geary-Khamis using the Fixed-Base Monthly Expanding-window method use an annual chain-link to maintain a representative basket, in line with our traditional methodology. This is how new data sources for items will often be introduced into consumer price indices were there to be no historical data available to us.<sup>1</sup>

**Figure 2: Indices produced for shortlisted methods using base synthetic dataset**

Figure 2: Indices produced for shortlisted methods using base synthetic dataset



**Source: Office for National Statistics – New index number methods in consumer price statistics**

The indices produced for the unweighted methods (GEKS-Jevons with a range of extension methods, fixed-base and chained-Jevons) are identical because all products are available in all time periods. When the availability of products on the market is static, multilateral methods should be exactly equal to their bilateral counterparts.

The difference between the weighted and unweighted indices is greater, on average, than the difference observed within the weighted methods, highlighting that the presence of weighting information at the product level is arguably more important than the choice of weighted index number method itself.

All of the extension methods give extremely similar outputs for the GEKS methods, something that we found throughout our analysis on these particular synthetic datasets. So, for ease of interpretation, we have decided to present findings using only the movement splice (our top-ranked extension method for the GEKS methods) for the remainder of this article. Note that this finding is in contrast to other countries (for example, [Australian Bureau of Statistics \(2016\)](#), [Statbel \(2018\)](#) and [Statistics Netherlands \(2019\)](#)). [Section 6: Case study using real-world data](#) considers differences in these extension methods over longer periods of real-world scanner data.

Given our extension methods over this periodicity of data show little difference in their results, for the remainder of this section we present six methods to be stress-tested against different pricing behaviours, the Quality-adjusted Geary Khamis (QU-GK), GEKS-Törnqvist (GEKS-T), GEKS-Fisher (GEKS-F), GEKS-Jevons (GEKS-J), chained-Jevons (CJ) and fixed-base Jevons (FBJ). The QU-GK method is presented using the Fixed-Base Monthly Expanding window (FBME) method, the GEKS methods are presented using the Movement Splice (MS) extension method. We have included other methods for comparison where they aid interpretation of the results.

## High attrition rate and product churn

One feature that is prominent in web-scraped and scanner data is a high number of products entering and leaving the market driven, for example, by episodes of fast-fashion (particularly in clothing markets) or rapidly advancing technology (for technology products such as PCs, laptops, tablets and smartphones).

Products can leave or enter the sample in many ways, including:

- the product goes out of stock and temporarily leaves the sample
- the product is restocked and re-enters the sample
- the product is discontinued and permanently leaves the sample
- the product is new to the market
- the product is rebranded or relaunched as a new product

The [International Labor Organization \(ILO\) – Section 7.153 \(2004\)](#) recommended high-frequency chaining when dealing with high product churn because the set of seasonal commodities that overlap during two consecutive months is likely to be much larger than the set obtained by comparing the prices of any given month with a fixed-base month. Therefore, the comparisons made using chained indices will be more comprehensive than those made using a fixed base. A fixed-base method will also fail if there is no product overlap between the current period and the chosen base period. A practical example is women’s coats, where at the end of each “season” the whole stock can be phased out in certain retailers and replaced with the new season's stock (see [Analysis of product turnover in web scraped clothing data, and its impact on methods for compiling price indices](#)).

A simple example of this fixed-base flaw using a basket of four products (A:D) can be seen in Table 6 and Figure 3.

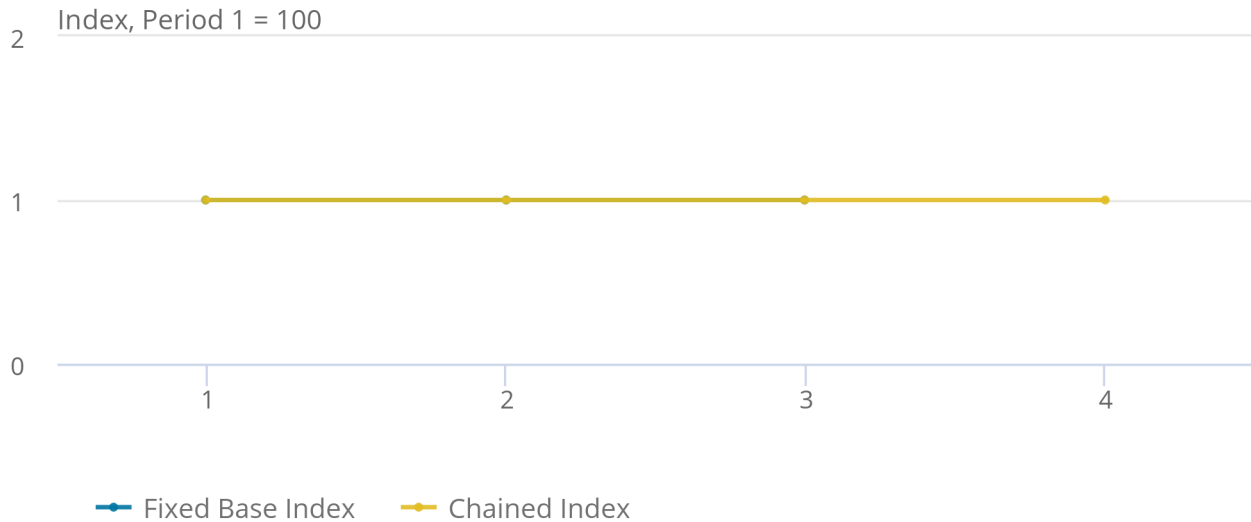
Table 6: Price data for four products in four periods

Period	Product A	Product B	Product C	Product D
January (1)	£1.00	£1.50		
February (2)	£1.00	£1.50	£1.20	
March (3)		£1.50	£1.20	£2.00
April (4)			£1.20	£2.00

Source: Office for National Statistics

**Figure 3: Comparison of fixed base and chained approach**

Figure 3: Comparison of fixed base and chained approach



**Source: Office for National Statistics – New index number methods in consumer price statistics**

In Period 4, we are unable to calculate an index using the fixed-base method because all the products available in the base period (products A and B) have now left the market. This means that fixed-based methods become highly problematic in datasets with high to complete product churn, as the sample of products that can be used to calculate the price index reduces throughout the year. This is one of the reasons that the bilateral fixed-base methods do not perform well in our framework compared to the multilateral methods.

The chained method continues, provided that there is some product overlap between the current period and the previous period. However, high-frequency chaining can have severe limitations. The approach can create drift in an index series owing to prices and quantities bouncing arising from sales, which is why chained bilateral methods were also outperformed by multilateral methods in our framework.

[Diewert and Fox \(2017\)](#) provide the example of chain drift that we have included in Table 7.

Table 7: Price and quantity data for two products for four periods  
Office for National Statistics - Diewert and Fox (2017)

Period	A price	A quantity	B price	B quantity
January (1)	£1.00	10	£1.00	100
February (2)	£0.50	5000	£1.00	100
March (3)	£1.00	1	£1.00	100
April (4)	£1.00	10	£1.00	100

Product A is subject to periodic sales, where a 50% price reduction is seen in February. As the price for Product A is reduced, the quantity purchased increases dramatically, from 10 to 5,000 units. In March, the product returns to its previous price, and the number of units dips to below-normal levels as consumers have already “stockpiled” the product in February. In April, typical pricing and purchasing behaviour resumes. The price and quantities of Product B remain stable across the four periods. Given that both prices and quantities purchased are the same in January and April, we would naturally assume that there has been no lasting inflation between these months and would therefore expect an index value of one (prices are the same as their original value).

Table 8 shows that the fixed-base superlative Törnqvist and Fisher price indices,  $I_{T(FB)}$ ,  $I_{F(FB)}$ , behave as expected and return to their period one index value in this simple scenario. The same is true of the multilateral GEKS versions of Törnqvist and Fisher,  $I_{GEKS-T}$ ,  $I_{GEKS-F}$ . However, the chained approaches,  $I_{T(C)}$ ,  $I_{F(C)}$ , show chain drift with 3% and 2% downward bias. If these data were chained monthly throughout a year, the overall chain drift bias would be significant.

Table 8: Indices (I) for price and quantity data in Table 7

Period	I T(FB)	I F(FB)	I GEKS-T	I GEKS-F	I T(C)	I F(C)
January (1)	1	1	1	1	1	1
February (2)	0.69	0.7	0.7	0.7	0.69	0.7
March (3)	1	1	0.99	0.99	0.97	0.98
April (4)	1	1	1	1	0.97	0.98

Source: Office for National Statistics

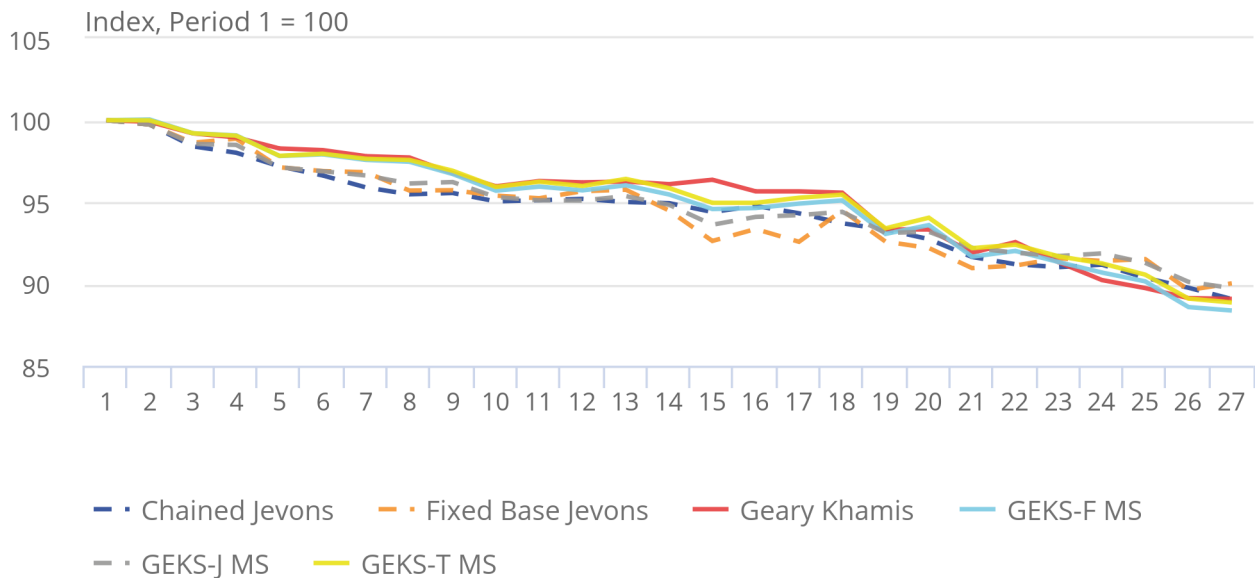
To test our shortlisted methods on a larger scale, high product churn was reproduced in our synthetic data by applying the following conditions to the base data:

- some products drop out of the sample permanently at various points
- some products enter the sample for the first time at various points
- some products temporarily drop out at random across the whole sample

The prices and quantities used in this high-churn dataset remain the same as in the base dataset. The only differences arise because of products being introduced as new products and products that are temporarily missing from the dataset. We therefore assess the performance of each method by comparing it to its base data index and seeing the resulting differences. We expect some small differences resulting from the third feature, as some price quotes will be temporarily missing at certain points across the time period assessed.

**Figure 4: Indices produced for top shortlisted methods using a high churn synthetic dataset**

Figure 4: Indices produced for top shortlisted methods using a high churn synthetic dataset

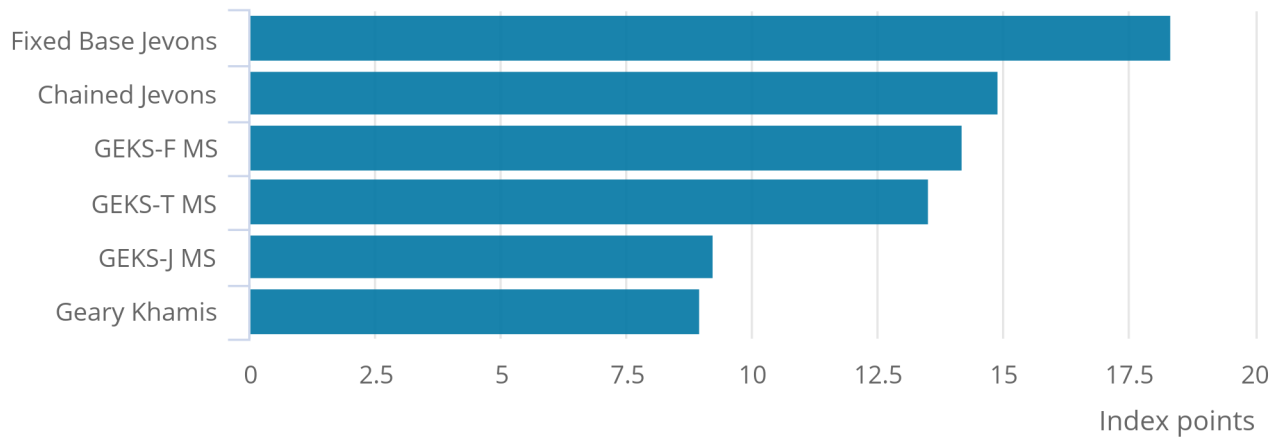


**Source: Office for National Statistics – New index number methods in consumer price statistics**

Figure 4 shows that all methods fall in a similar way to the fall that we observed in the base data (Figure 2), although the variation between the different index number methods is now greater. To assess the performance of each method, we compare the sum of the absolute differences between the index values based on the simple base data and the index values produced using the dataset with high churn added. Results from this comparison are displayed in Figure 5.

**Figure 5: Sum of absolute differences between high churn and base index values across 27 periods**

Figure 5: Sum of absolute differences between high churn and base index values across 27 periods



**Source: Office for National Statistics – New index number methods in consumer price statistics**

Figure 5 shows that the difference between the sum of absolute differences for the index number methods investigated is relatively small, but this simulation would need to be run a number of times to see if this is a recurring feature across all high-churn datasets. While some methods may match closely in the high churn and base scenarios, that does not necessarily mean either values are representing a true closeness to inflation. For example, while the GEKS-Jevons has a small difference between the high-churn and base datasets, it differed from the weighted methods in the base datasets because it does not use product-level weighting information.

For our dataset, when churn is high the fixed-base Jevons and chained Jevons vary further from their base indices than the multilateral methods. As already noted, the fixed-base Jevons method will use an ever-decreasing proportion of the available data each month of a year, so it becomes less representative of the item than the chained and multilateral methods over time. Furthermore, chained bilateral methods typically suffer from chain drift so, over time, we would expect them to become increasingly different from the baseline dataset. Given the properties of the methods, we would therefore expect the sum of absolute differences to increase for the fixed-base and chained Jevons compared to the multilateral methods over time.

## Product obsolescence

Another common feature of consumer prices data is product obsolescence. Obsolescence occurs when a product ceases to be functionally competitive with other products. This may occur when a retailer stops producing or marketing a product because the product becomes undesirable to the consumer. High-tech products such as smartphones often display this feature, as when new models are released the older models will often be withdrawn from the market.

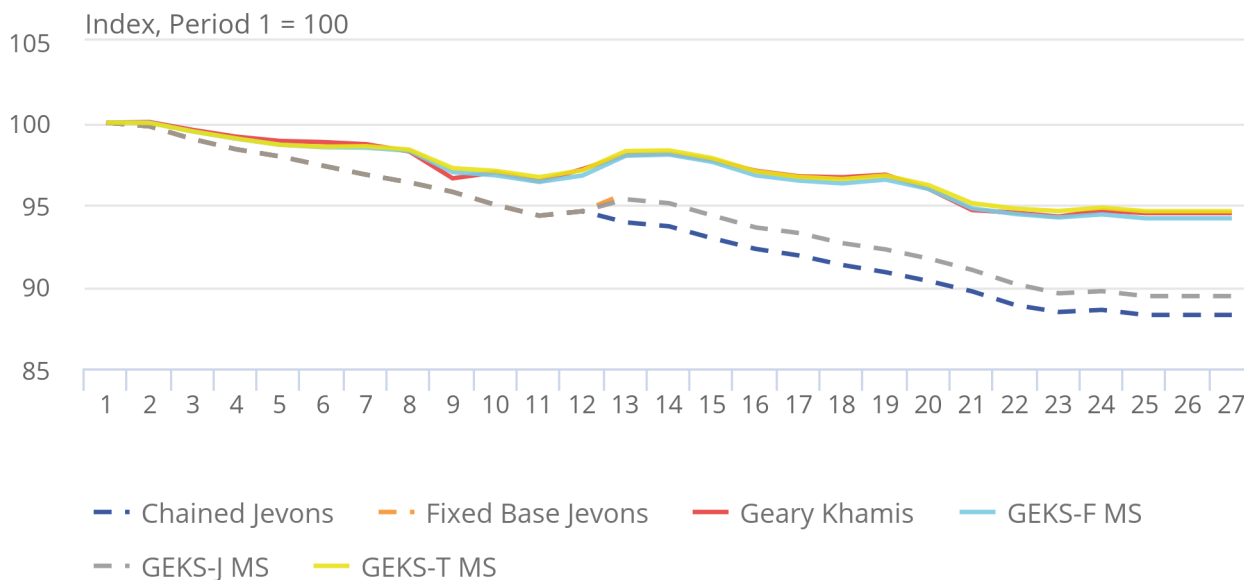


An obsolescence scenario was created in our synthetic data by reducing product prices and quantities in the basket over time, until all products become obsolete and are no longer sold. These products are subsequently replaced by new products that have a definite launch period where the price is the same value as the corresponding product that it replaces had at its launch. There is a window of four periods where the initial products leave the market while new products are launched. There may be some overlap here as some products will be launched before the old product has left the market.

For example, if a new phone entered the market at £800, its price and the number of units sold would gradually decrease over the period until a newer model is released, at which point the number of the initial product sold diminishes to zero. In this scenario, the new model would be released at £800, within a four-month window of the old model ceasing sales. Figure 6 shows the index values that are produced when using this synthetic data on our shortlisted methods.

**Figure 6: Indices produced for shortlisted methods using the product obsolescence synthetic dataset**

Figure 6: Indices produced for shortlisted methods using the product obsolescence synthetic dataset



**Source: Office for National Statistics – New index number methods in consumer price statistics**

The products observed in the first period are all obsolete by period 13, where replacement products are launched at the same price as the products they replaced were at their launch in period 1. New products are launched over 4 periods, so some prices may have already begun to decrease by the time all the new products have been launched. Therefore, we do not expect the index values to return completely to their original levels at any point.

The problem of using a fixed-base bilateral method is highlighted in Figure 6, as after period 13 there are no longer any matching products in the set of the current and base periods; therefore, no more index values can be produced, and the time series ends. This is resolved in traditional methods through annual chain-linking of fixed-base indices or through product replacements when products become obsolete within-year (with appropriate quality adjustments made to account for any changes in specification). Annual chain-linking would not solve the problem were the products to become obsolete mid-year. The process of directly replacing products when they become obsolete is very manual and would be extremely time consuming to use with the large data sources that we are currently investigating.

The unweighted methods result in indices significantly lower than the weighted methods in this scenario. This is unsurprising as they do not use the expenditure information, instead giving each product in the basket equal importance. The weighted methods identify that, in this obsolescence dataset, products with increased popularity do not reduce in price as quickly as the products that are leaving the market, again showing the importance of having sales information, or approximates of sales values, at the product level.

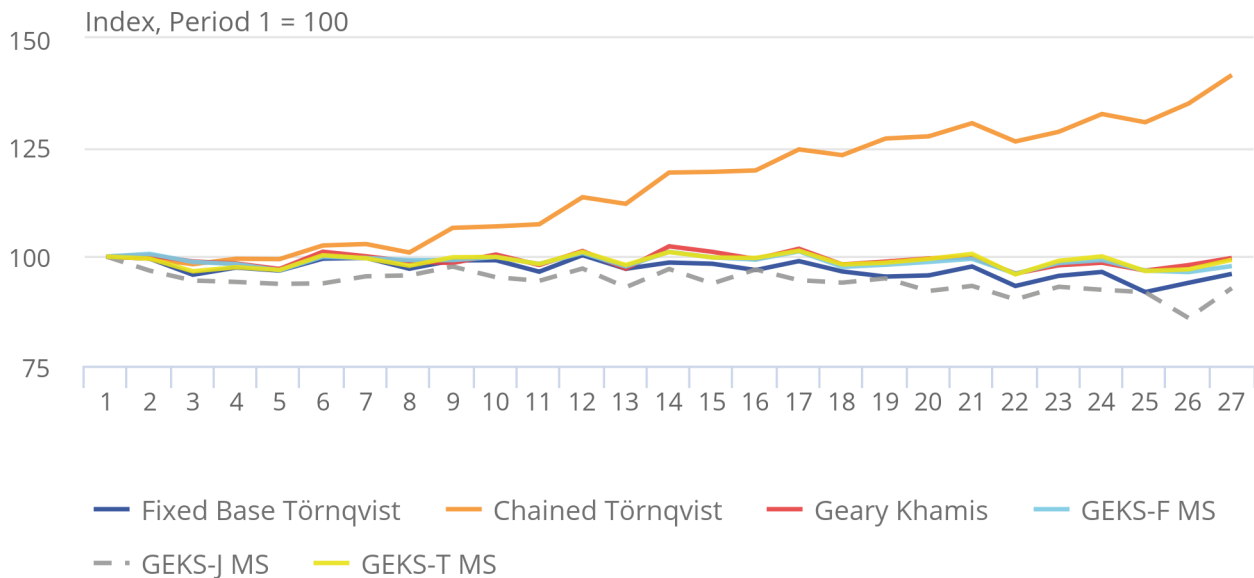
## High price variance

The price movements of products in some commodity groups may have a higher variance than those observed in our base data. To increase the variance of price movements, a synthetic dataset was created that multiplied each product's monthly price movement in the base dataset by a factor of five. This feature was isolated in our modifications of the base data; therefore, there is no churn and our unweighted methods all give identical results. We therefore only include unweighted index values for GEKS-J MS as well as the other weighted indices. As we are basing our results on a static set of products, we compare our shortlisted methods against a fixed-base Törnqvist index.

The resulting indices, found in Figure 7, show that our shortlisted methods track the increased upward and downwards price movements of the fixed-base Törnqvist relatively well in the initial 13-month window; however, after this initial window, the QU-GK, GEKS-T and GEKS-F all show similar upward bias of up to 5%. For comparison, the chained Törnqvist method has been included, once again justifying the choice to favour a multilateral method, as by the 27th period the chained Törnqvist method shows an upward bias of approximately 45%.

**Figure 7: Indices produced for some shortlisted methods using the high price variance synthetic dataset**

Figure 7: Indices produced for some shortlisted methods using the high price variance synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

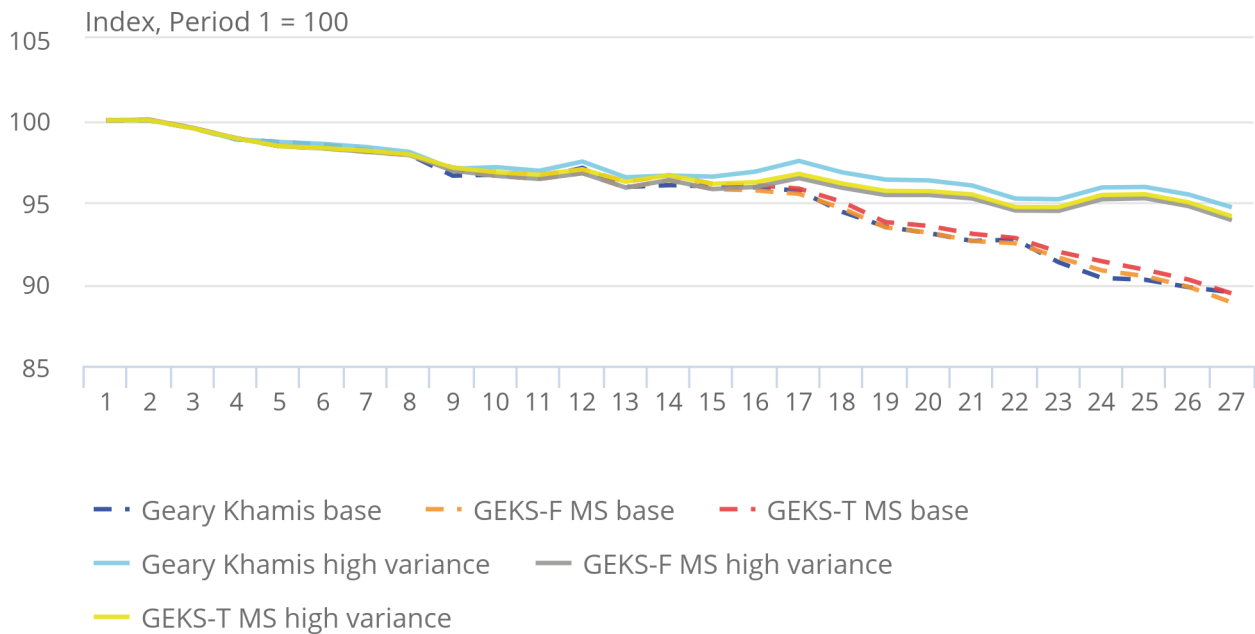
## High quantities of products sold

Price indices should reflect consumer spending, but unweighted methods do not make use of quantity information so do not alter in periods where there is an increased number of sales for a particular product or range of products. Weighted methods are more comprehensive, as if sales are higher for certain products in a sample then these products have more of an impact on the resulting price indices. To produce a synthetic dataset showing this feature, we increased the number of products sold within a proportion of our sample where prices are decreasing less rapidly.

Figure 8 shows that each of the QU-GK, GEKS-T and GEKS-F represent this high sales feature well, with greater index values from their counterpart values on the base data; this is because the increased sales quantities were in a proportion of the sample where prices were decreasing by less.

**Figure 8: Indices produced for weighted shortlisted methods using the high sales synthetic dataset**

Figure 8: Indices produced for weighted shortlisted methods using the high sales synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

We also note that the QU-GK appears to be slightly more responsive to the items with increased sales values than the GEKS-T and GEKS-F. As unweighted indices are unaffected by changes in quantities purchased, their results are the same as they were for the base data and are therefore not presented in this subsection.

### Other features assessed

Table 9 shows additional features that have been isolated in our synthetic datasets and used to further stress-test our shortlisted methods. The stress-testing against these behaviours did not result in any further conclusions and therefore have not been presented here, but corresponding charts can be found in [Annex B](#).

Table 9: Other synthetic data sets for testing index number methods

Feature	Description
Dump prices	When a product ceases to be produced, it is possible that retailers will apply a “dump” price on their remaining stock of the product to encourage sales.
External shocks	Any external shocks in supply or demand can be replicated in the synthetic data to see how well the index number methods account for these shocks.
Level shift	This dataset will replicate any level shifts that may be seen in pricing structures, such as an increase in VAT.
Flat line	If there are no prices changes for any products between two periods we would want a price index to remain the same as the previous month.
Seasonality	Seasonality is a characteristic of a time series in which the product experiences regular and predictable changes in price and/or sales that recur at consistent intervals
Short series	A method should be able to cope should a data series come to an abrupt ending. E.g. if all products in an EA cease to be sold, the method should not continue producing index values for future periods.

Source: Office for National Statistics

## Computational run time of the methods

Each method selected for use in production of consumer price indices must be integrated into our [new system for calculating consumer price statistics](#). Ahead of a conclusive decision on the chosen methods for use on the available alternative data sources, computational run times for each method on four datasets were calculated. These run-times are displayed in Table 10.

Table 10: Comparison of computational run times for some methods in ONS consumer prices pipeline

Data Set	Number of unique items	Number of rows of data	Window length of method	Computational run time for method (Seconds, 2 d.p.)				
				QU-GK FBME	GEKS-T MS	GEKS-F MS	GEKS-J MS	Chained Jevons
A	827	226278	13	66.08	16.76	9.8	12.37	6.55
B	827	452556	13	449.22	22.41	11.75	15.87	7.56
C	863	465799	27	478.63	37.01	25.9	35.65	8.19
D	863	931598	27	570.26	27.1	22.21	35.9	8.07

Source: Office for National Statistics

### Notes

1. These methods are still undergoing optimisation and it is hoped that efficiency can be further improved.

Interestingly, the run times for some of our GEKS methods reduce as the number of rows of data increase. This is to do with the systematisation of the methods in our pipeline that runs in PySpark, which is the Python interface to Spark. More details on our processing pipeline can be found in our article [Using alternative data sources in consumer price indices](#). PySpark is optimised for large datasets through processing them in parallel and batch systems. Because of the sequential nature of the QU-GK method, the parallel processes used by Spark are less successful, meaning that the QU-GK method appears to take a much longer time to run. This run time increases as the data scale.

While the QU-GK method takes a substantially longer time to run than the other shortlisted methods, the calculation of indices for 863 elementary aggregates in less than 10 minutes is currently still deemed acceptable within our framework as it is unlikely to hinder the timeliness or frequency of our consumer price statistics. More research is required to understand at what point the QU-GK method becomes unacceptably slow, but it is clear that this threshold could be reached for the QU-GK method before any of the GEKS methods or the chained Jevons within our current infrastructure.

## Notes for: Stress-testing shortlisted index number methods

1. Historical data can be provided by retailers, and we are also building up a database of web-scraped data. Therefore, by the time we implement scanner and web-scraped data, from 2023, we will be able to use historical data to calculate the index using a full window of data rather than expanding the window for the first year. This is still under consideration, so for the purposes of this article we have chosen to replicate the methods that we would use were there no historical data available to us.

## 6 . Case study using real-world data

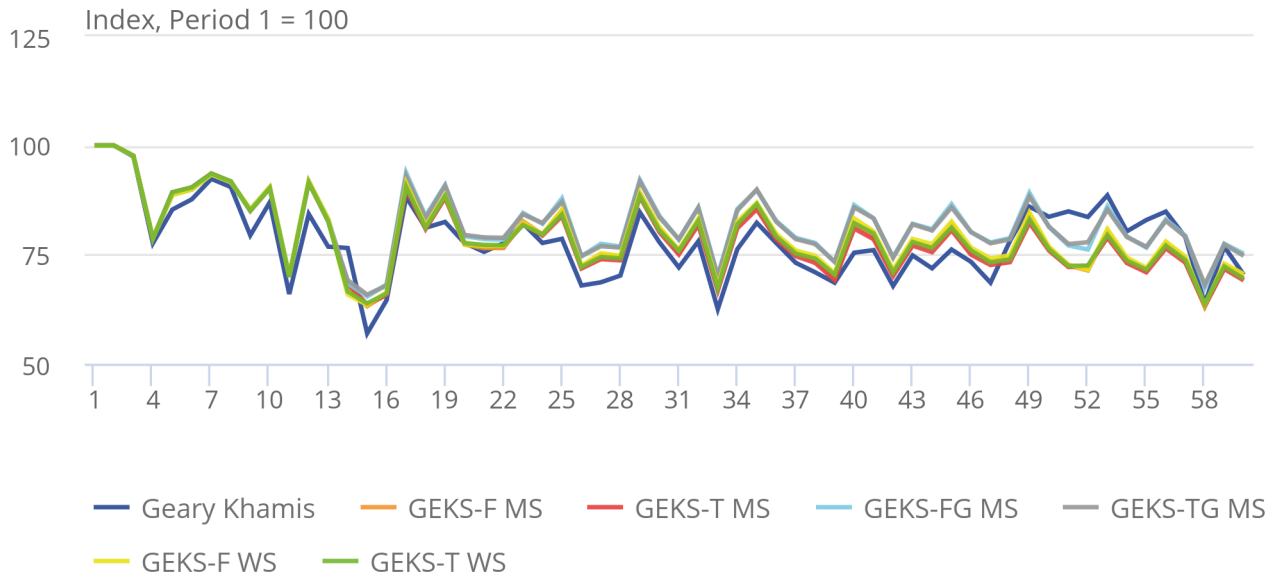
While our stress-testing of the methods isolated features to understand how our index number methods would respond, we are aware that a number of these features could be simultaneously present in a real-world dataset. To test our methods on a real dataset, we have used the open source [Dominick's Finer Food data \(Dominick's\)](#). These data have been provided by the James M. Kilts Center, University of Chicago Booth School of Business.

The Dominick's data cover store-level transactions data collected at Dominick's Finer Foods over a period of more than seven years, from 1989 to 1997, across 29 different categories throughout all stores in this US chain with over 100 stores. From these 29 categories, five were selected as case studies: beer; cereals; laundry detergents; soft drinks; and toothpaste. The first 60 periods were taken for the highest expenditure store in each of the selected categories, and indices were calculated for each of our shortlisted methods.

To compare the extension methods over a longer time period than was used in [Section 5: Stress-testing shortlisted index number methods](#), Figure 9 shows the indices produced using our weighted shortlisted methods for the Soft Drinks category. To compare weighted and unweighted indices using the full sample, Figure 10 shows weighted indices (QU-GK, GEKS-F MS and GEKS-T MS) compared with the unweighted indices (GEKS-J MS, GEKS-J WS and GEKS-J GMS). Similar charts for the remaining case-study spending categories are provided in [Annex C](#).

**Figure 9: Indices produced using weighted shortlisted methods for soft drinks**

Figure 9: Indices produced using weighted shortlisted methods for soft drinks

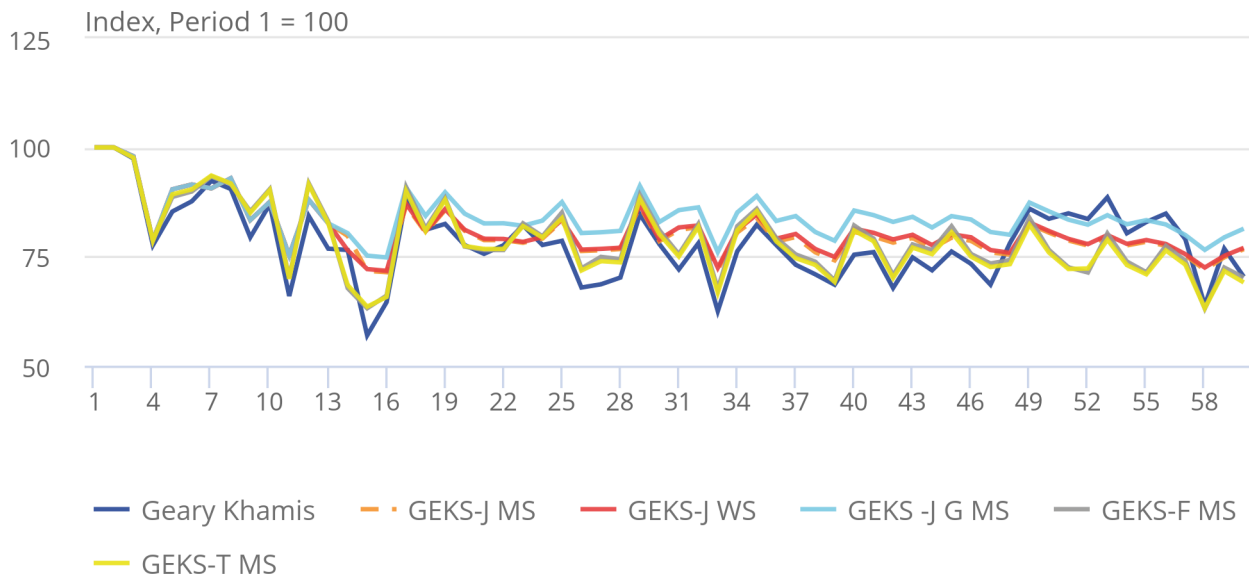


Source: Office for National Statistics – Dominick’s



**Figure 10: Indices produced using a range of shortlisted methods for soft drinks**

Figure 10: Indices produced using a range of shortlisted methods for soft drinks



**Source: Office for National Statistics – Dominick’s**

Two main observations can be made from the observations of these case studies. First, differences between each method’s and each extension methods’ indices are more substantial in practice than we observed when stress-testing the methods in [Section 5: Stress-testing shortlisted index number methods](#); this is likely impacted by a longer period of data as well as seeing many features occurring simultaneously rather than in isolation. Secondly, there is an apparent upward bias from the GEKS-J methods in comparison to the weighted methods; this is likely because consumers substitute towards products that are on sale and this is not accounted for when using unweighted methods. This again highlights that having information on sales values, or approximates thereof, is arguably more important than the choice between weighted index number methods themselves.

It is clear for each of the five categories that the different extension methods cause the indices to differ more significantly in the latter period of the time series, a result that replicates the findings of other National Statistical Institutes’ (NSIs’) studies.

In Table 11, we look at the average difference between our top-shortlisted weighted and unweighted methods for the five food categories considered. The lowest average difference in methods are presented in bold.

Table 11: Average difference between some unweighted and weighted methods for five Dominick's food categories, index points

Category	Weighted method	GEKS-J MS	GEKS-J WS	GEKS-J GMS
Beer	QU-GK	3.07	3.05	3.05
	GEKS-T MS	1.21	1.19	1.19
	GEKS-F MS	1.27	1.26	1.26
Cereal	QU-GK	2.14	2.19	1.76
	GEKS-T MS	0.42	0.47	0.04
	GEKS-F MS	0.53	0.58	0.15
Laundry detergent	QU-GK	0.96	0.99	0.91
	GEKS-T MS	1.49	1.52	1.44
	GEKS-F MS	1.48	1.51	1.43
Soft drinks	QU-GK	2.49	2.79	5.94
	GEKS-T MS	2.01	2.32	5.47
	GEKS-F MS	1.53	1.83	4.99
Toothpaste	QU-GK	0.15	0.09	0.08
	GEKS-T MS	0.77	0.71	0.7
	GEKS-F MS	0.87	0.82	0.8

Source: Office for National Statistics

Table 11 shows that the unweighted methods (GEKS-J MS, GEKS-J WS and GEKS-J GMS) are upwardly biased relative to the weighted methods for all five of the categories assessed from the Dominick's data. It is particularly striking that the averages are positive across all permutations. This is likely because under normal economic conditions, consumers will substitute towards products that are on sale, and this substitution is not accounted for in unweighted indices.

For beer, cereal, laundry detergent and toothpaste, the GMS extension method causes the least deviation from the weighted methods; however, for soft drinks, the movement splice deviates the least. Time series comparing the indices for the methods in Table 11 for the five categories can be found in Figures 21 to 25 in [Annex C](#).

This apparent upward bias of the GEKS-J methods could be problematic for web-scraped data. For a dataset that contains all available products without any detail of the number of each product sold (such as for a web-scraped dataset), the use of GEKS-J could be biased compared with the same data where quantity data would be available (such as for a scanner dataset). As our traditional collection samples products that are considered representative of consumer purchases, we would not expect to see the same scale of bias from using unweighted indices for these sources. Our ongoing research into [approximating sales values](#) should allow us to weight products by order of their economic importance, reducing the need to use unweighted methods in our indices produced using scanner and web-scraped data in the future.

## 7 . Conclusions and future work

Our framework and subsequent testing have shown that comparisons made using chained or multilateral indices will be more comprehensive than those made using fixed-base methods, as they account for new and disappearing products without the need for manual intervention. A fixed-base method becomes less representative of the current market over time and, more importantly, will fail in scenarios where there is high or complete churn of a market within a year. Chained bilateral methods were also outperformed by multilateral methods in both our framework and synthetic data testing. The chained method continues, provided that there is some product overlap between the current period and the previous period, but high-frequency chaining can create substantial chain-drift in an index series.

The QU-GK, GEKS-T and GEKS-F adequately represent each of the features tested and rarely diverged by more than a single percentage point over the 27 periods tested in our isolated feature analysis. The deviation between methods was greater when tested on a real scanner dataset, but this difference was mostly still smaller than the difference between our weighted and unweighted methods. This highlights the importance of using information on, or approximates of, sales values at the product level for web-scraped and scanner data and that this research is arguably more valuable than the choice between weighted index number methods and extension methods themselves.

The choice of extension method when using GEKS had minimal effect on our stress-testing scenarios, with the WS, MS and GMS all closely approximating each other's results when used on this periodicity of data. However, we did see greater differences between the extension methods when used on a longer periodicity of data in [Section 6: Case study using real-world data](#), a finding that has also been observed by other National Statistical Institutes (NSIs) (for example, [Australian Bureau of Statistics \(2016\)](#), [Statbel \(2018\)](#) and [Statistics Netherlands \(2019\)](#)). We will continue to investigate the optimal extension method as well as the impact of changing the window length in future research.

As QU-GK was the highest-scoring method against the frameworks criteria, coupled with satisfactory results under the stress-testing and in our real-world case studies and the (currently) satisfactory computational run time, the QU-GK method is our current proposed method to use with scanner and web-scraped data when expenditure information, or approximations of expenditure, are available. In the absence of expenditure information, we propose the GEKS-Jevons to be a suitable alternative, although we continue to stress the importance of research into approximating product-level expenditures as a more suitable approach – provided suitable expenditure approximates can be made.

We also note that the GEKS-T and GEKS-F methods also performed well when stress-tested against the isolated feature datasets and only scored slightly lower against the framework criteria and have a quicker computational run time than the QU-GK. Should increase of data collection ever reach a stage where QU-GK's slower run time becomes burdensome, then a move to GEKS-T or GEKS-F would be appropriate.

The move to use scanner and web-scraped data for consumer price statistics is a current point of interest among many other NSIs, and research is being undertaken internationally into the selection of index method. Should international consensus on an optimal method be reached, then a switch to ease international comparison could be sensible. We will continue to produce indices for our favoured methods in parallel while research, both in the UK and internationally, progresses.

## 8 . Related links

[Research and developments in the transformation of UK consumer price statistics: September 2020](#)

Article | Released 1 September 2020

The first in a series of biannual articles to update users on our research to modernise the measurement of consumer price inflation in the UK.

[Automated classification of web-scraped clothing data in consumer price statistics](#)

Article | Released 1 September 2020

Research into using supervised machine learning algorithms to efficiently classify web-scraped clothing data, for use in consumer price statistics.

[Using statistical distributions to estimate weights for web-scraped price quotes in consumer price statistics](#)

Article | Released 1 September 2020

Feasibility of predicting sales quantities from product ranks, for potential use with web-scraped data in consumer price statistics.

## 9 . Annex A: Technical descriptions of index number methods

For each product  $i$  the representative price  $p$  and quantity  $q$  sold in month  $t$  are described as  $p_i^t$  and  $q_i^t$  respectively.

A price relative for product  $i$  between two months measures the percentage change in the product's price. This is shown as:

$$\frac{p_i^{month\ 2}}{p_i^{month\ 1}}$$

Where relevant,  $n$  will be used to describe the number of products covered by the index method.

Expenditure is measured by price multiplied by quantity,  $p_i^t q_i^t$ . An expenditure share is the proportion of expenditure for product or item  $i$  measured at a specific time  $t$ . This is often known as an expenditure weight or simply a weight  $w$ . The calculation of an expenditure share is given as:

$$w_i^t = \frac{p_i^t q_i^t}{\sum_{i=1}^n p_i^t q_i^t}$$

The value for an index method measuring price change from month one to month two will be described as:

$$P_{Index\ method}^{(Month\ 1, Month\ 2)}$$

Note that often we use zero for the base month, which index values are benchmarked against, and  $t$  for the current month that we are trying to measure. The following would represent the price change between the base month and the current month using the Jevons index:

$$P_{Jevons}^{(0,t)}$$

Bilateral index methods are calculated using two time periods, the base period and the period of measurement. Multilateral methods, on the other hand, work over a time window of multiple consecutive months. We therefore define a window of length  $T$  as a set such that:

$$W = \{month\ 0\ (base\ month),\ month\ 1,\ \dots,\ month\ T - 1\}$$

$$length(W) = T$$

## A.1. Unweighted bilateral indices

### Jevons

The Jevons index is calculated as a geometric mean of price relatives between the time period of measurement and a base time period, for a consistent set of products. The Jevons formula is currently the most commonly used index number method at the lowest level of aggregation in UK consumer price statistics, but it cannot be used when prices fall to exactly zero (for more information, see [Paul Johnson review, UK consumer price statistics – Chapter 10, 2015](#)).

In its fixed-base variant, the Jevons is measured as:

$$P_{FB-Jevons}^{0,t} = \sqrt[n]{\prod_{i=1}^n \frac{p_i^t}{p_i^0}}$$

In its chained variant, the Jevons is measured as:

$$P_{Chained\ Jevons}^{0,t} = P_{FB-Jevons}^{0,1} \times P_{FB-Jevons}^{1,2} \times \dots \times P_{FB-Jevons}^{t-1,t}$$

Note that when a consistent set of products is used over all time periods, the fixed-base Jevons and the chained Jevons are equivalent. Under traditional data collection methods, product linkage (where outgoing products are linked to replacement products) ensures that this holds. However, when using alternative data sources, such as scanner and web-scraped data, the product composition generally changes over time and so the chained and fixed-base Jevons can give different results.

### Carli

The Carli is similar to the Jevons except instead of a geometric average across price relatives, it applies an arithmetic average to the price relatives of a consistent set of products:

$$P_{FB-Carli}^{0,t} = \left(\frac{1}{n}\right) \times \left(\sum_{i=1}^n \frac{p_i^t}{p_i^0}\right)$$

Note, however, that the Carli is not transitive. Whereas, under a consistent set of products, the chained and fixed-base variants will be identical using the Jevons:

$$P_{FB-Jevons}^{0,t} = P_{FB-Jevons}^{a,b} \times P_{FB-Jevons}^{b,c}$$

The same is not always true for the Carli.

The chained Carli (and all other chained bilateral indices) follow the equivalent Carli formulation to the chained Jevons.

## Dutot

Where the Jevons (and Carli) are an average of price relatives, the Dutot is seen as a relative of average prices for a consistent set of products (i):

$$P_{FB-Dutot}^{0,t} = \frac{\left(\frac{1}{n}\right) \times \sum_{i=1}^n p_i^t}{\left(\frac{1}{n}\right) \times \sum_{i=1}^n p_i^0} = \frac{\sum_{i=1}^n p_i^t}{\sum_{i=1}^n p_i^0}$$

Note that the Dutot places more importance on higher-priced products. An expensive product increasing in price by 10% would have more of an effect than a cheaper product increasing by 10%. Therefore, the Dutot is better used when the item covers products of a similar price.

## A.2. Weighted bilateral indices

### Laspeyres (and Lowe)

The Laspeyres index uses quantities in the base period to weight each product or item /price. The formula is given as:

$$P_{FB-Laspeyres}^{0,t} = \frac{\sum_{i=1}^n p_i^t q_i^0}{\sum_{i=1}^n p_i^0 q_i^0}$$

An equivalent alternative formulation of the Laspeyres index can be expressed in terms of price relatives multiplied by their expenditure shares rather than quantities:

$$P_{FB-Laspeyres}^{0,t} = \sum_{i=1}^n \left( w_i^0 \times \frac{p_i^t}{p_i^0} \right)$$

This is useful since instead of the formula being based on exact quantities, they are based on expenditure shares, which are more readily estimated using the available data sources.

However, note that the base month is typically in January or December of the year of measurement and the latest available expenditure share information comes from sources collecting the data prior to the base month. These sources are then price updated to the base month. This gives rise to the Lowe index, often described as a "Laspeyres-type" index for its similarity in construction:

$$P_{FB-Lowe}^{0,t} = \sum_{i=1}^n \left( \hat{w}_i^{before-0} \times \frac{p_i^t}{p_i^0} \right)$$

Note that an equally weighted Laspeyres index is equivalent to a Carli index. Similarly, an equally weighted geometric Laspeyres index is equivalent to a Jevons index:

$$P_{FB-Geo\ Laspeyres}^{0,t} = \prod_{i=1}^n \frac{(p_i^t)^{w_i^0}}{(p_i^0)^{w_i^0}}$$

## Paasche

The Paasche index is similar in construction to the Laspeyres index. However, prices are quantity adjusted for the current period rather than the base period:

$$P_{FB-Paasche}^{0,t} = \frac{\sum_{i=1}^n p_i^t q_i^t}{\sum_{i=1}^n p_i^0 q_i^t}$$

Like the Laspeyres index, the Paasche index has both arithmetic and geometric forms.

## Fisher

The Fisher index is calculated as a geometric average of the Laspeyres and the Paasche indices:

$$P_{FB-Fisher}^{0,t} = \sqrt{P_{FB-Laspeyres}^{0,t} \times P_{FB-Paasche}^{0,t}}$$

As described in [Section 3: Overview of index number methods](#), since the Laspeyres is a base-weighted index method and the Paasche is a current-period weighted index method, taking the geometric average of the two places equal importance on base and current period weights, making the Fisher a superlative index method.

## Törnqvist

The Törnqvist index is calculated as the geometric average of the price relatives, weighted by the average expenditure shares from the two periods.

$$P_{FB-Törnqvist}^{0,t} = \prod_{i=1}^n \left( \frac{p_i^t}{p_i^0} \right)^{\frac{1}{2}(w_i^0 + w_i^t)}$$

Similar to the Fisher, the Törnqvist is a superlative index.

## A.3. Multilateral indices

### GEKS

The GEKS index was developed as a spatial index for Purchasing Power Parities. To make it suitable for temporal price indices, we compare price differences over time, rather than between countries. Pairing the GEKS method with different bilateral price indices gives rise to different varieties of GEKS indices. We specifically focus on three: GEKS-Törnqvist, GEKS-Fisher and GEKS-Jevons.

The GEKS uses a link month  $l$  that iterates through time window  $W$ . For each  $l$ , we calculate a paired set of bilateral indices from the base period 0 to  $l$  and from  $l$  to the month of measurement  $t$ . We then take the geometric value of all the bilateral pairs. Therefore, the generic form for the GEKS index is:

$$P_{GEKS}^{(0,t)} = \sqrt[T]{\prod_{l=0}^{T-1} \frac{P^{(l,t)}}{P^{(l,0)}}} \text{ (where } l \in W \text{)}$$

The Jevons, Fisher and Törnqvist bilateral indices all pass the time reversal test



$$P^{(0,t)} = \frac{1}{P^{(l,0)}}$$

Therefore, an alternative formulation is possible. To use the GEKS-Jevons as an example:

$$P_{GEKS-Jevons}^{(0,t)} = \sqrt[T]{\prod_{l=0}^{T-1} \frac{P_{Jevons}^{(l,t)}}{P_{Jevons}^{(l,0)}}} = \sqrt[T]{\prod_{l=0}^{T-1} (P_{Jevons}^{(0,l)} \times P_{Jevons}^{(l,t)})} \text{ (where } l \in W \text{)}$$

Therefore, we can understand the GEKS-Jevons as the geometric average of all possible paired bilateral Jevons indices from the base period 0 to the current period  $t$  through a link period  $l$  that iterates through the window  $W$ . Similar formulas hold for the GEKS-Törnqvist and GEKS-Fisher. Note, however, that some indices (such as the Carli) do not pass the time reversal test, and therefore the earlier formula should be used to ensure the index remains transitive.

## Geary-Khamis

The Geary-Khamis (GK) index was also developed as a spatial index for Purchasing Power Parities, but unlike the GEKS, which compares each period to each other, the GK index compares each period to a base period. It is an implicit price index that divides a value index by a weighted quantity index. Using notation like [Chessa \(2016\)](#), it is defined as:

$$P_{Geary Khamis}^{(0,t)} = \frac{\frac{\sum_{i \in S^*} p_i^t q_i^t}{\sum_{i \in S^*} p_i^0 q_i^0}}{\frac{\sum_{i \in S^*} v_i q_i^t}{\sum_{i \in S^*} v_i q_i^0}}$$

where the weights  $v_i$  are as follows:

$$v_i = \sum_{z \in T} \phi_i^z \frac{p_i^z}{P_{Geary Khamis}^z}$$

$$\phi_i^z = \frac{q_i^z}{\sum_{s \in T} q_i^s}$$

In simple terms, we consider the QU-GK in the consumer price indices context as a quality-adjusted unit-value index. For more information on this method, we recommend referring to [Chessa \(2016\)](#).

## Time product and time product dummy

The time product dummy (TPD) aims to decompose the price of a product into how much of the price comes from being that specific product and how much comes from it being observed in a specific time period. The TPD method uses a regression approach that is like those of hedonic-based methods – it uses the statistical relationship between prices, products and time to estimate the decomposition.

The TPD model is expressed as:

$$\ln p_i^t = \alpha + \sum_{t=1}^T \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t$$

where,

$\ln p_i^t$  = log of price for item  $i$  in period  $t$

$\alpha$  = intercept term

$\delta^t$  = time parameter corresponding to time period  $t$

$D_i^t$  = time dummy variable, equal to 1 if the price observation was observed to period  $t$  and 0 otherwise

$\gamma_i$  = product parameter corresponding to product  $i$

$D_i$  = product dummy variable, equal to 1 if the price observation  $p_i^t$  pertains to item  $i$  and 0 otherwise

$\varepsilon_i^t$  = error term

The TPD hedonic method uses a similar formulation but includes a product's characteristics as part of the hedonic equation:

$$\ln p_i^t = \alpha + \sum_{t=1}^T \delta^t D_i^t + \sum_{k=1}^K \beta_k^t z_{ki}$$

where,

$\beta_k^t$  = the estimate of the effect characteristic  $k$  has on the price at time  $t$

$z_{ki}$  = the value of characteristic  $k$  for item  $i$

More details of these methods can be found in [Office for National Statistics \(ONS\) methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data](#).

## A.4. Extension methods for multilateral indices

### Direct extension

As mentioned in [Section 3: Overview of index number methods](#), all multilateral methods operate by forming indices over many time periods within a window. However, consider the following 13-month window:

$$W = \{January\ 2020, February\ 2020, \dots, January\ 2021\}$$

At the time of publishing a multilateral price index for March,  $P_{(January\ 2020, March\ 2020)}$ , we would lack the information requirements from the remainder of the window until all data were collected by the end of January 2021. The practical way in which we can extend our time series is known as an extension method.

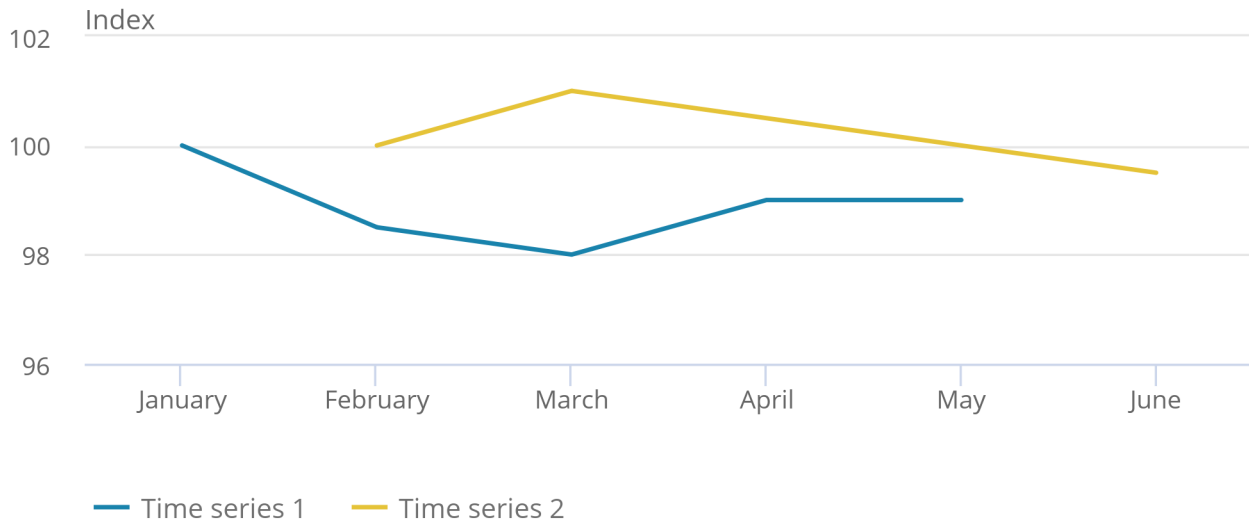
The direct extension method involves publishing the index based on the available months at the time of publication and then making revisions throughout the year as more data become available. This would mean revising the time series every month. Since UK consumer price statistics are [not typically revised](#), we deem this method unsuitable for UK consumer price inflation statistics.

### Splicing (window, half-window, movement and geometric)

Splicing involves “connecting” two time series through a specific month. For example, Figures 11a and 11b show two time series spliced through May, which extends the first time series into June.

**Figure 11a: Time series values for two windows of data before splicing**

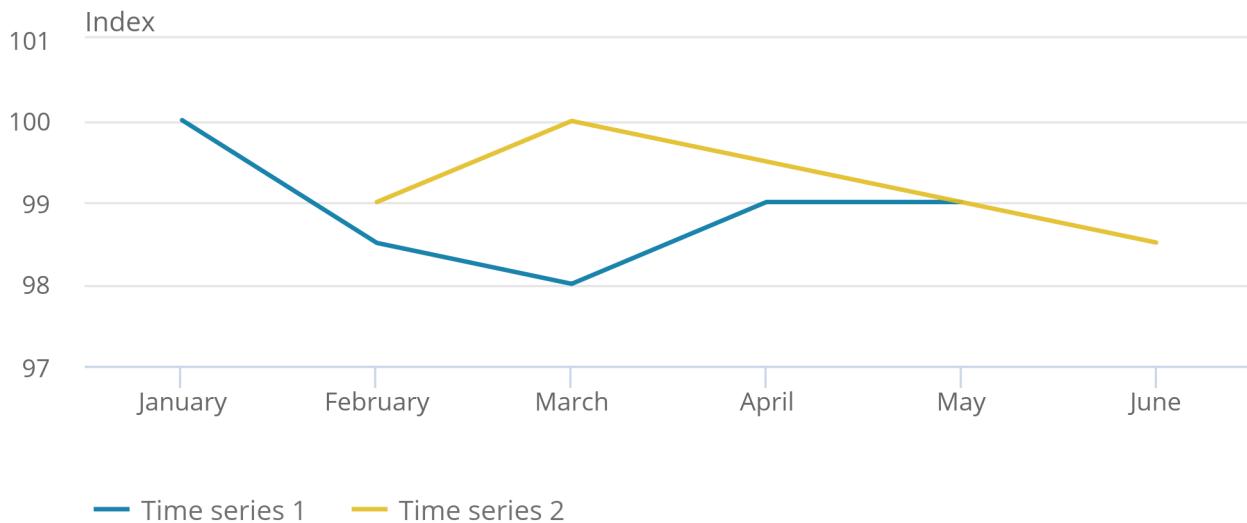
Figure 11a: Time series values for two windows of data before splicing



**Source: Office for National Statistics – New index number methods in consumer price statistics**

**Figure 11b: Time series two has been spliced onto time series one through May, extending the long-term series into June**

Figure 11b: Time series two has been spliced onto time series one through May, extending the long-term series into June



**Source: Office for National Statistics – New index number methods in consumer price statistics**

Splicing makes use of a rolling window. An example of a rolling window is shown:

$$W_1 = \{January\ 2020, February\ 2020, \dots, January\ 2021\}$$

$$W_2 = \{February\ 2020, March\ 2020, \dots, February\ 2021\}$$

$$W_3 = \{March\ 2020, April\ 2020, \dots, March\ 2021\}$$

Here, when publishing February 2021, we could splice the price movement between January and February 2021 using window  $W_2$  onto the January 2021 index formed using  $W_1$ . We could then continue to splice the price movement for March 2021, conducted over window  $W_3$ , onto the time series formed previously by  $W_1$  and  $W_2$ . The series continues to build in this way.

Splicing can occur in any period that is shared by the two windows. In Figures 11a and 11b, we chose to splice through May. This would be known as a movement splice, as we are splicing over one month prior to the month of measurement (t-1). Alternatively, we could splice over February, our earliest time period that is available in both windows one and two, giving a window splice. We could also splice over the midpoint of our current window, in this case April, known as a half-window splice. Generally, we could splice over any other month shared by the two windows. As a compromise, we could also take the geometric average of the values obtained by splicing on all the months, giving a geometric mean splice.

Formulae are given for the various splices. Note that when splicing,  $t$  always refers to the last month in the window, therefore  $t_0=t-(T-1)$  refers to the first month. Zero refers to the base month in the long-term series.  $M$  refers to a generic multilateral index method. A generic formula is given, where  $s$  represents the month that is being spliced over:

$$P_{M,s}^{(t_0-1,t)} = P^{(t_0,t)} \times \frac{P^{(t_0-1,s)}}{P^{(t_0,s)}}$$

We can splice over any  $i$  that is shared between the two windows. Some common choices of  $i$  are named:

$s = t_0$  gives the window splice.

$s = t - 1$  gives the movement splice.

$s = \frac{t_0+t}{2}$  gives the half – window splice.

Alternatively, we could splice on a geometric average of all possible months, giving the geometric mean splice:

$$P_{M,Geo\ mean\ splice}^{(0,t)} = P^{(t_0,t)} \times \sqrt{(T-1) \prod_{s=t_0}^{t-1} \frac{P^{(t_0-1,s)}}{P^{(t_0,s)}}$$

## Fixed-base monthly expanding window

As an alternative to splicing, which involves use of a rolling window, we could instead make use of an expanding window with a fixed-base:

$$W_1 = \{January\ 2020, February\ 2020\}$$

$$W_2 = \{January\ 2020, February\ 2020, March\ 2020\}$$

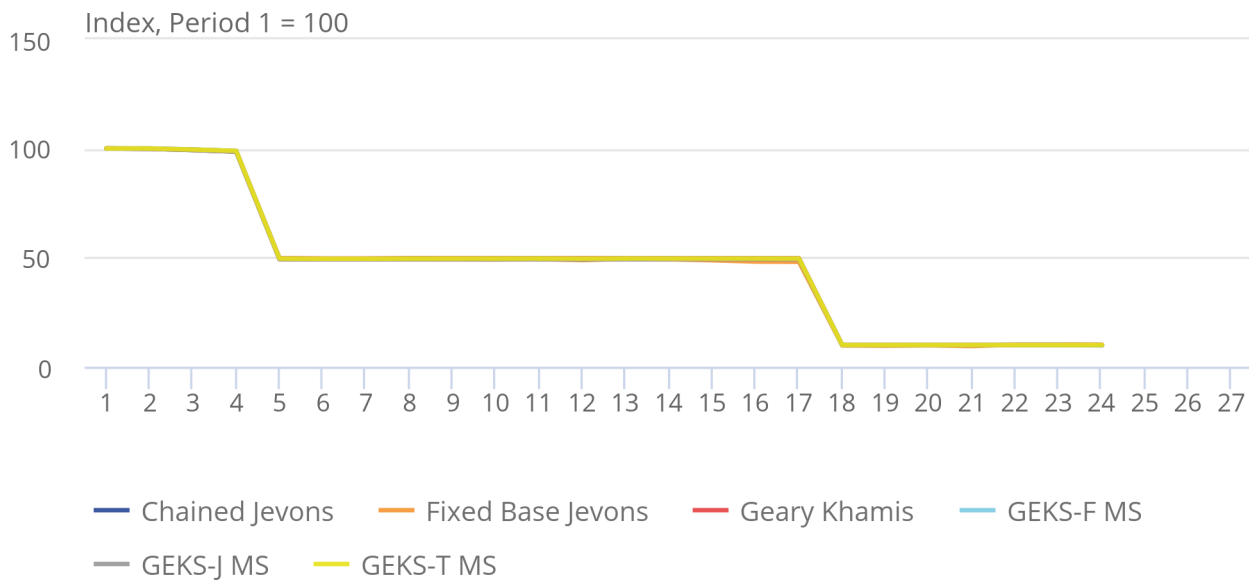
For example, to calculate the index from January to February 2020, we calculate the multilateral index over  $W_1$ , and to calculate the index from January to March 2020, we calculate the multilateral over  $W_2$

The fixed-base monthly expanding window can be seen as applying the direct extension method without making revisions as new data come in.

## 10 . Annex B: Stress-testing of additional features

Figure 12: Indices produced for shortlisted methods using the dump price synthetic dataset

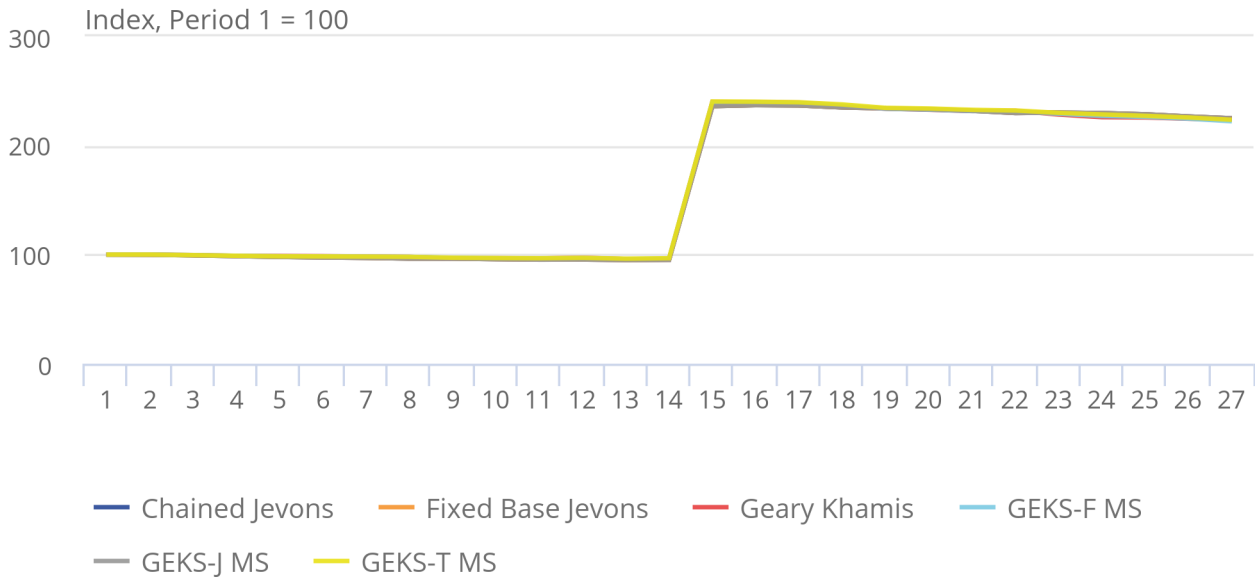
Figure 12: Indices produced for shortlisted methods using the dump price synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

**Figure 13: The external shock is seen in the indices in period 15, all indices show very similar results**

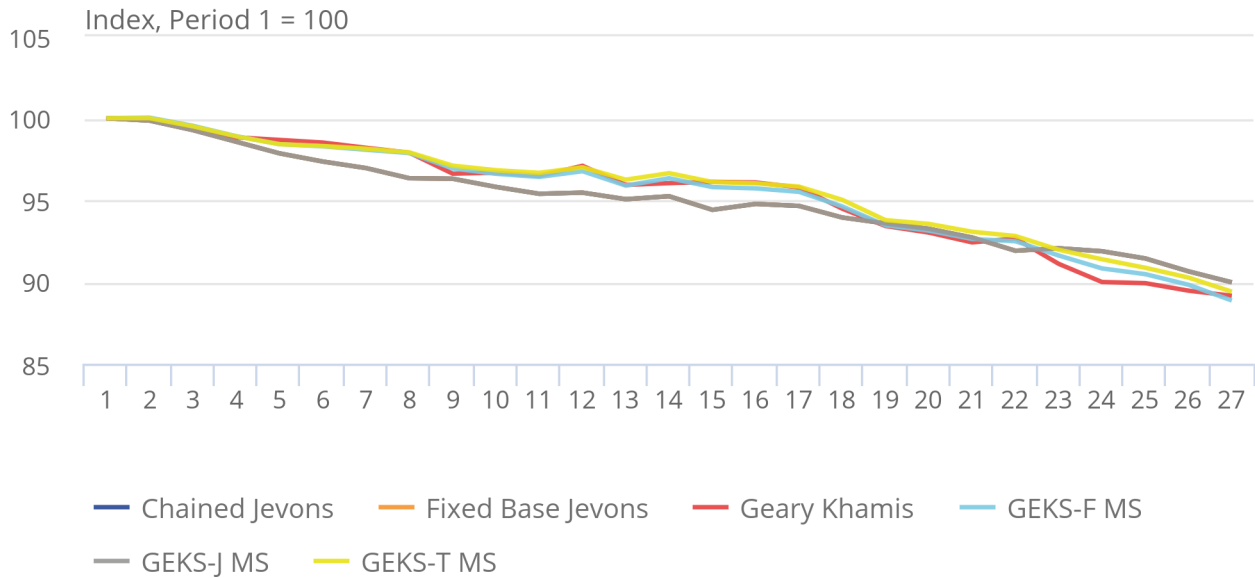
Figure 13: The external shock is seen in the indices in period 15, all indices show very similar results



Source: Office for National Statistics

**Figure 14: Indices produced for shortlisted methods using external shock quantity increase synthetic data**

Figure 14: Indices produced for shortlisted methods using external shock quantity increase synthetic data

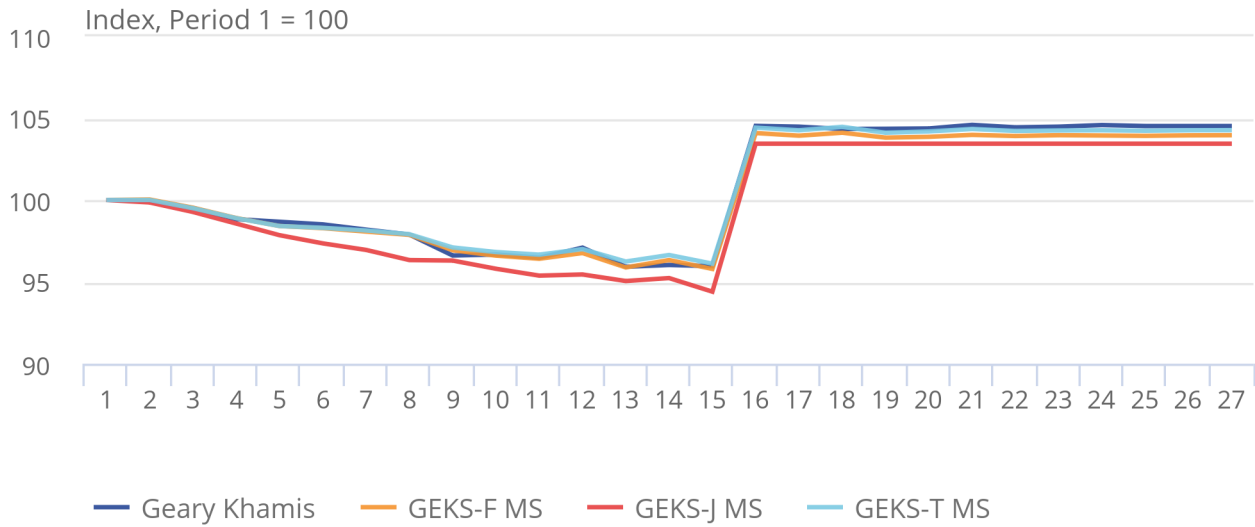


Source: Office for National Statistics – New index number methods in consumer price statistics



**Figure 15: Indices produced for some shortlisted methods using the price-level shift synthetic dataset**

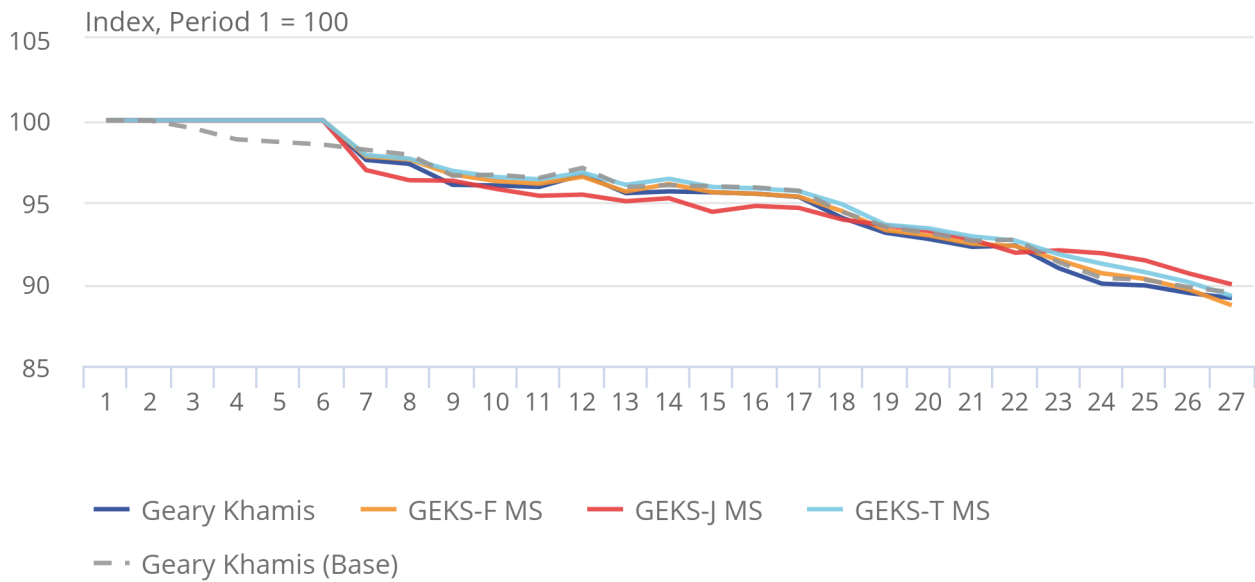
Figure 15: Indices produced for some shortlisted methods using the price-level shift synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

**Figure 16: Indices produced for shortlisted methods using the flat line synthetic dataset**

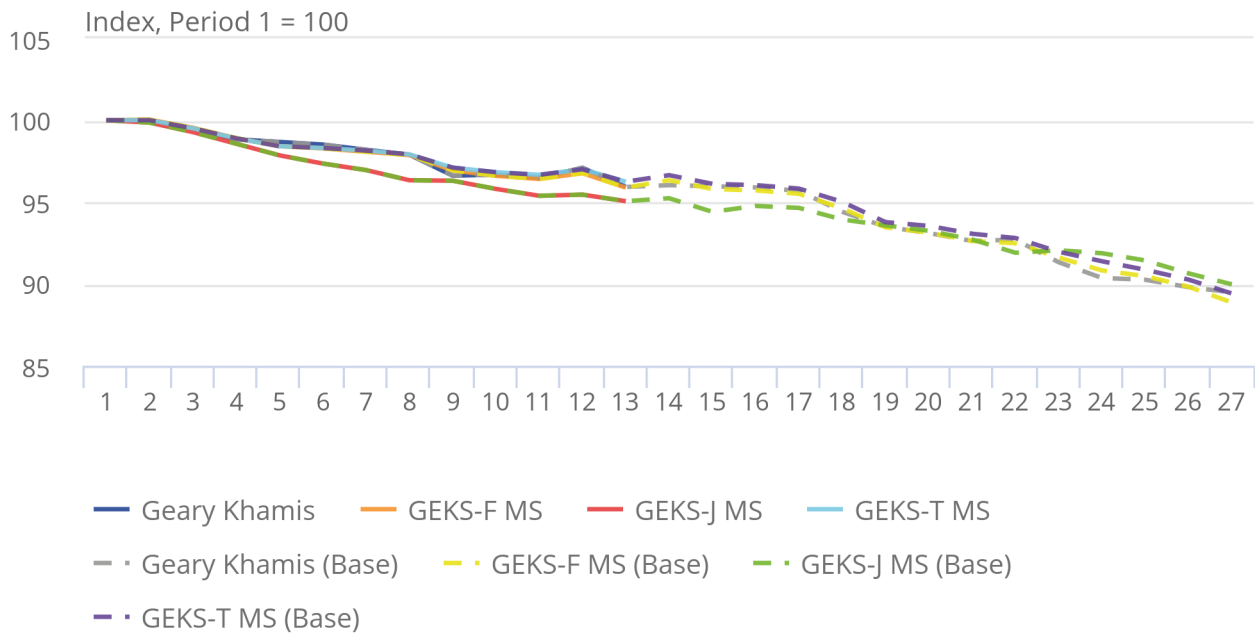
Figure 16: Indices produced for shortlisted methods using the flat line synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

**Figure 17: Indices produced for shortlisted methods using the short series synthetic dataset**

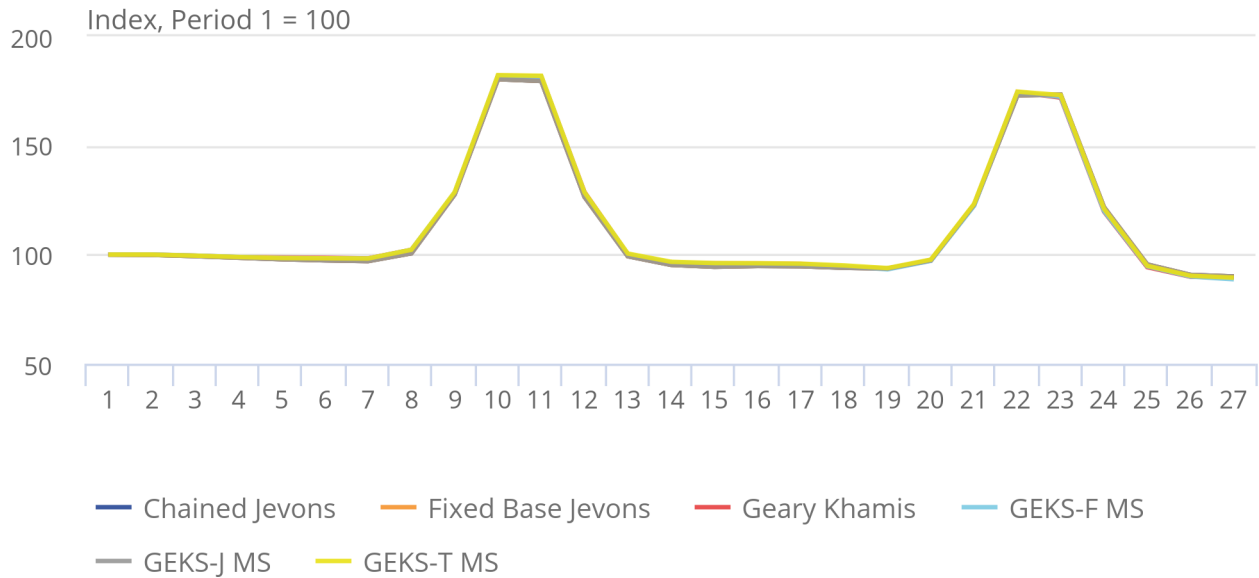
Figure 17: Indices produced for shortlisted methods using the short series synthetic dataset



Source: Office for National Statistics – New index number methods in consumer price statistics

**Figure 18: Indices produced for shortlisted methods using seasonal synthetic dataset**

Figure 18: Indices produced for shortlisted methods using seasonal synthetic dataset

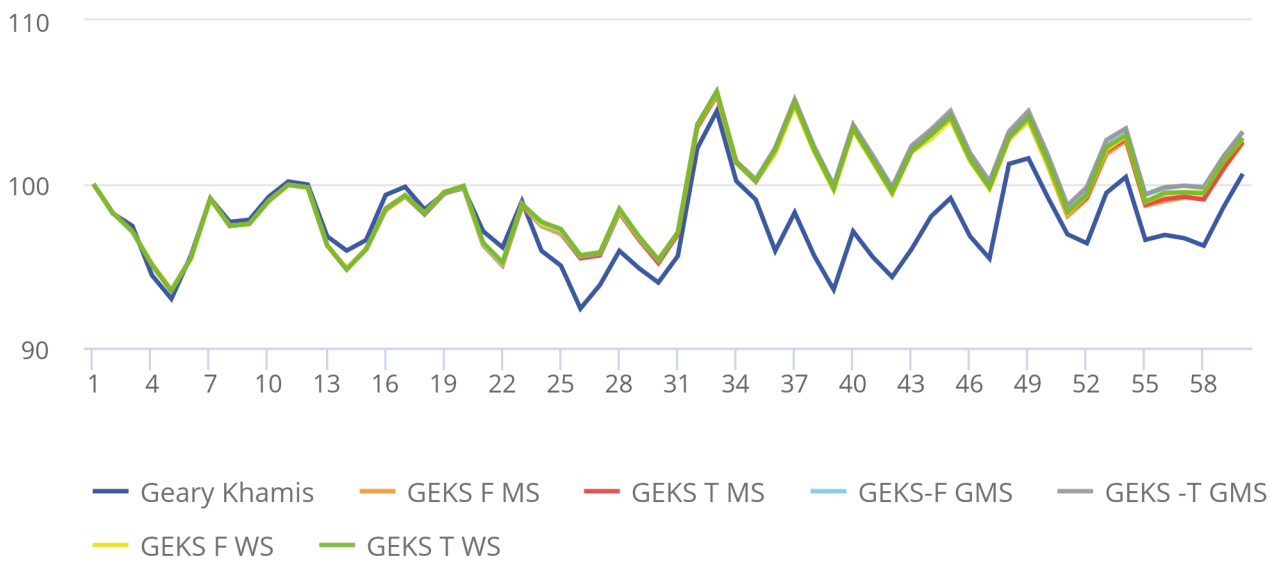


Source: Office for National Statistics – New index number methods in consumer price statistics

# 11 . Annex C: Case studies of indices produced using real-world data

Figure 19: Indices produced using shortlisted weighted methods for beer

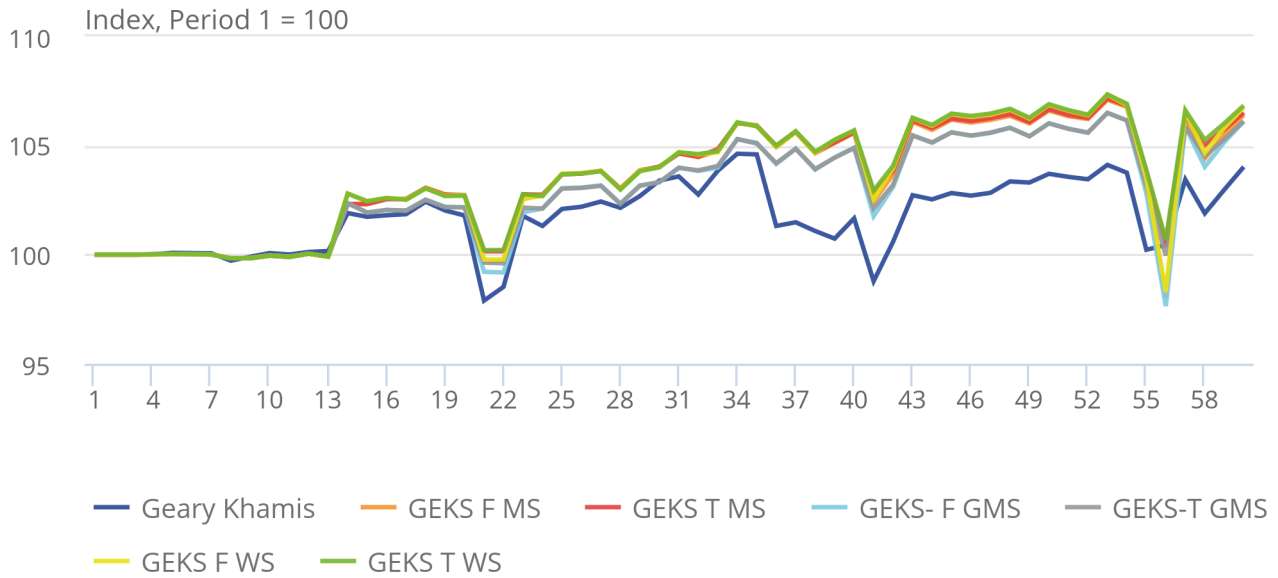
Figure 19: Indices produced using shortlisted weighted methods for beer



Source: Office for National Statistics – Dominick’s

**Figure 20: Indices produced using shortlisted weighted methods for cereal**

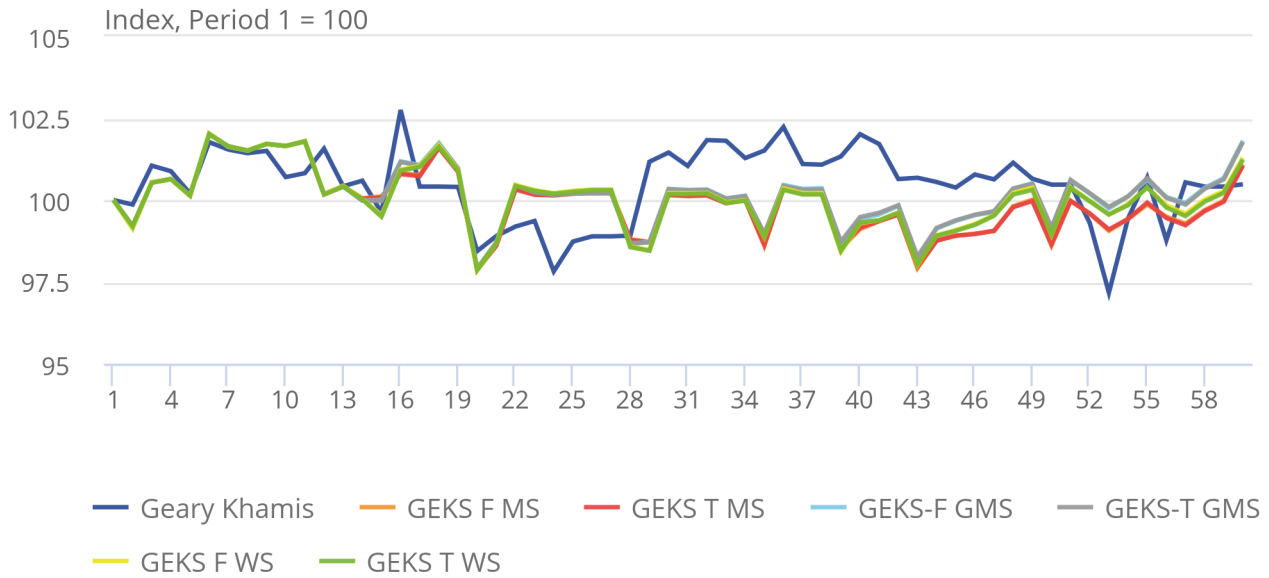
Figure 20: Indices produced using shortlisted weighted methods for cereal



Source: Office for National Statistics – Dominick’s

**Figure 21: Indices produced using shortlisted weighted methods for laundry detergent**

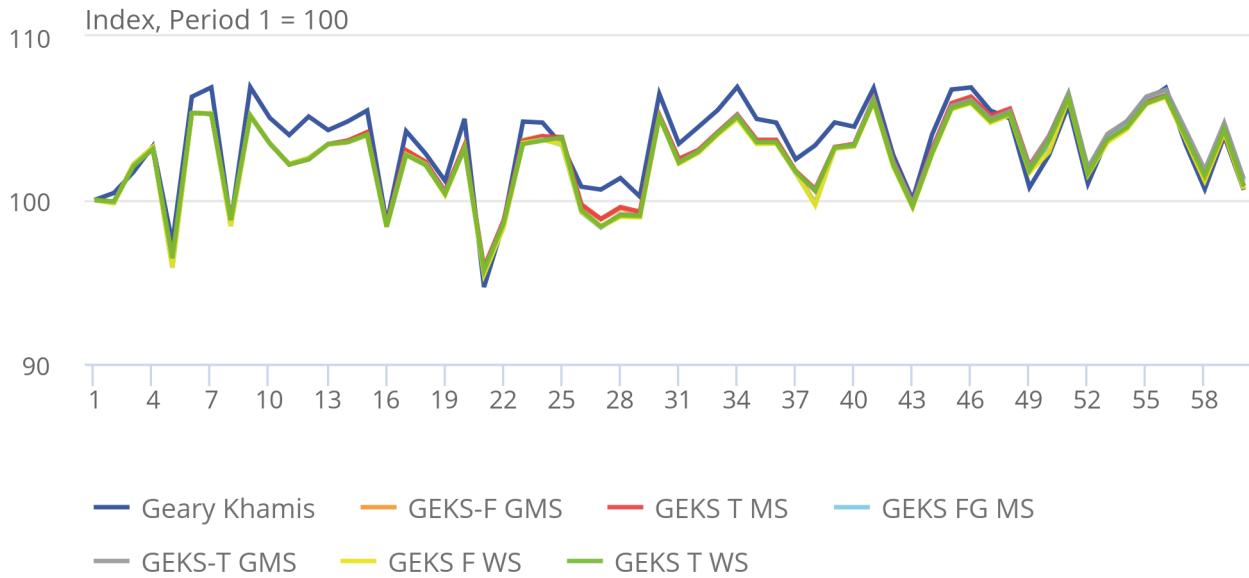
Figure 21: Indices produced using shortlisted weighted methods for laundry detergent



Source: Office for National Statistics – Dominick’s

**Figure 22: Indices produced using shortlisted weighted methods for toothpaste**

Figure 22: Indices produced using shortlisted weighted methods for toothpaste

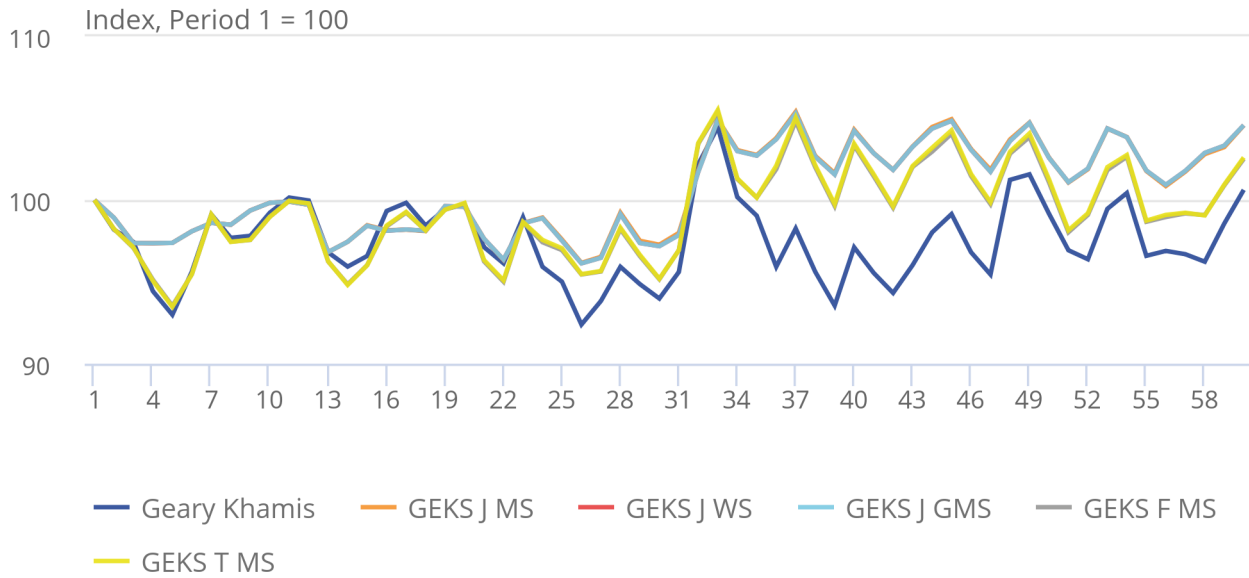


Source: Office for National Statistics – Dominick’s



**Figure 23: Indices produced using a range of weighted and unweighted shortlisted methods for beer**

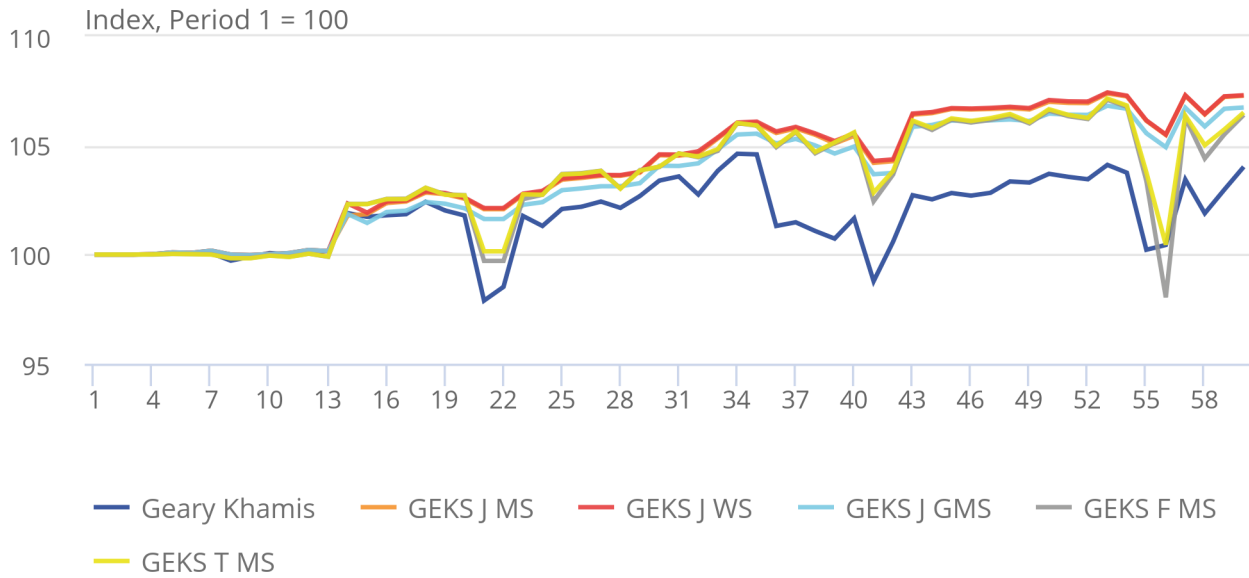
Figure 23: Indices produced using a range of weighted and unweighted shortlisted methods for beer



Source: Office for National Statistics – Dominick’s

**Figure 24: Indices produced using a range of weighted and unweighted shortlisted methods for cereal**

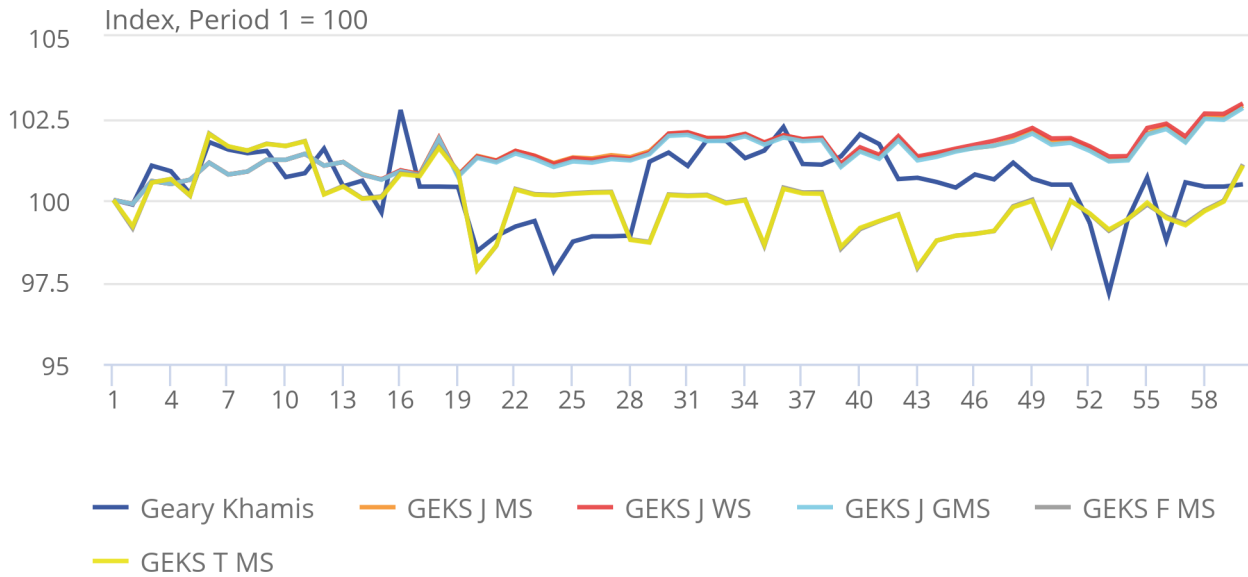
Figure 24: Indices produced using a range of weighted and unweighted shortlisted methods for cereal



Source: Office for National Statistics – Dominick’s

**Figure 25: Indices produced using a range of weighted and unweighted shortlisted methods for laundry detergent**

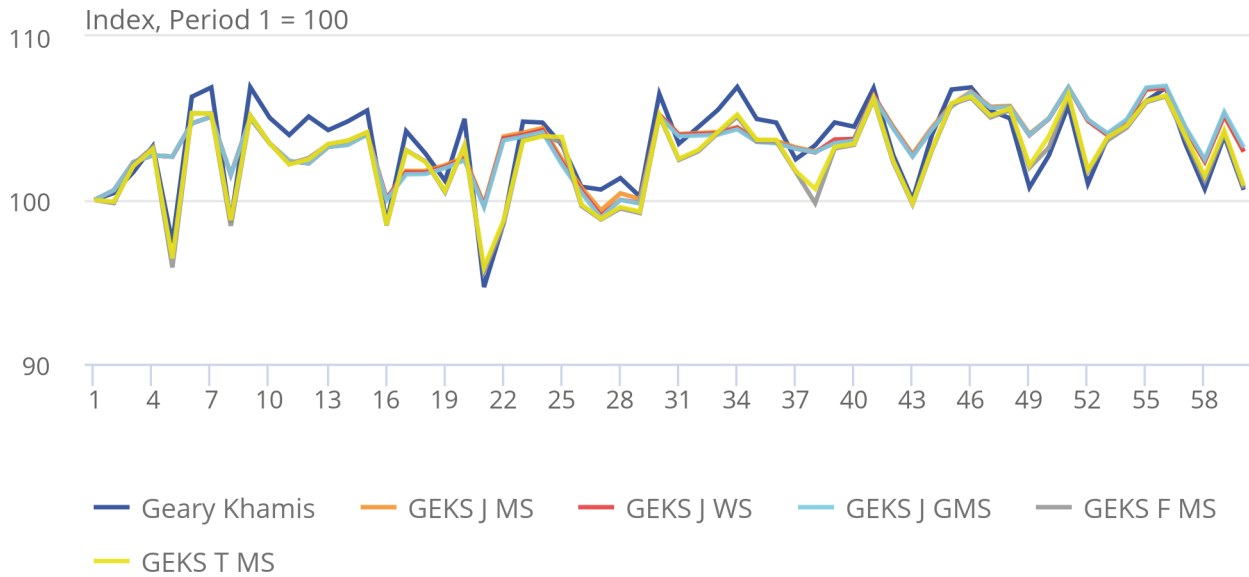
Figure 25: Indices produced using a range of weighted and unweighted shortlisted methods for laundry detergent



Source: Office for National Statistics – Dominick’s

**Figure 26: Indices produced using a range of weighted and unweighted shortlisted methods for toothpaste**

Figure 26: Indices produced using a range of weighted and unweighted shortlisted methods for toothpaste



Source: Office for National Statistics – Dominick’s