

Alternative data aggregates in consumer prices

How we use alternative data sources to produce elementary aggregates. Part of our technical guidance on consumer prices indices.

Contact:
Consumer Price Inflation team
cpi@ons.gov.uk
+44 1633 456900

Release date:
25 March 2026

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [Data sources](#)
3. [Data processing](#)
4. [Calculating elementary aggregate indices](#)
5. [Methods of the GEKS-Törnqvist](#)
6. [Definitions](#)
7. [Related links](#)
8. [Cite this methodology article](#)

1 . Overview

In the early 2020s, we introduced several large, automatically collected "alternative" data sources. These allow us to broaden our product coverage, use more representative prices and use weights at the product level.

Our inflation measures are separated into components called "aggregates" (or sometimes "strata") as described in our [Scope and coverage of consumer prices indices methodology article](#). These exist within a hierarchy where "elementary aggregates" form the lowest level.

In this section of the [Consumer price indices technical guidance](#), we explain how we compile "elementary aggregates" using these alternative data sources. We describe these data, how we process them, and how we use them to calculate price indices.

How we compile measures of inflation

This article is part of a set explaining how consumer price inflation and associated indices are compiled. Other related guidance articles include:

- [Consumer price indices technical guidance](#)
- [Scope and coverage of consumer prices indices](#)
- [Traditional data aggregates in consumer prices](#)
- [Special case aggregates in consumer prices](#)
- [Private rents and owner-occupier housing aggregates in consumer prices](#)
- [Higher-level aggregation and weights in consumer prices](#)
- [Calculating the Retail Price Index](#)
- [How we publish our consumer price outputs](#)

This set of related articles replaces components of our [Consumer Prices Indices Technical Manual, 2019 methodology](#).

2 . Data sources

We have introduced several alternative data sources into the calculation of our consumer price indices. These include:

- Grocery scanner data were introduced from March 2026
- Second-hand cars were introduced from March 2024
- Rail fares were introduced from March 2023

These datasets do not contain information that can identify individual consumers. Instead, they reflect aggregated sales (or listings) across the retailers covered by the datasets.

The following subsections describe each dataset in more detail.

Grocery scanner data

Grocery scanner data are created by retailers from customer transactions recorded at in-store tills or online. Each row in the dataset gives the total expenditure and quantity sold for a product at a specific outlet over a defined timeframe (typically a week, but sometimes a day).

Table 1 shows a simplified synthetic example. Real scanner datasets contain billions of rows and additional features such as product sizes, descriptions and hierarchy information.

Table 1. Examples of grocery scanner data

Week start	Outlet	Product ID	Expenditure (£)	Units sold
7 Sept 2025	Pencoed	2512616	2211	2521
7 Sept 2025	Glastonbury	2525165	321	1010
14 Sept 2025	Glastonbury	2525165	452	1256
21 Sept 2025	Glasgow	5252511	5265	3257

Source: Office for National Statistics

More comprehensive information on the grocery data and methods can be found in our [Introducing grocery scanner data into consumer price statistics article](#).

Second-hand cars

Second-hand car data are provided by the company Auto Trader. The data contain hundreds of thousands of unique car listings each month, from thousands of car dealerships across the UK.

Each listing includes detailed characteristics on the advertised vehicle, such as its make, model, age and mileage. The dataset also records the advertised price for each day a vehicle is listed.

Car listings may appear for several weeks before the car is sold and removed from the website. If a car is removed from the website, then relisted later, this may indicate a sale that has fallen through. We infer a sale occurs if a listing has been removed and does not re-appear within the following four days. The final listing price is then taken as the sale price.

More comprehensive information on the second-hand cars data and methods can be found in our [Using Auto Trader car listings data to transform consumer price statistics, UK: July 2023 article](#).

Rail fares

Rail fare data for Great Britain are provided by the Rail Delivery Group and are sourced from the rail industry's Latest Earnings Networked Nationally Overnight (LENNON) ticket revenue system. We receive tens of millions of rows of data each month. Each row within the dataset relates to a transaction – this provides us with information on the expenditure and quantities of purchased tickets, along with relevant data on the route and ticket bought.

More comprehensive information on rail fares data and methods can be found in our [Using transaction-level rail fares data to transform consumer price statistics, UK article](#).

3 . Data processing

Data validation and staging

Our data validation process ensures that the raw alternative data we receive from external suppliers are accurate, complete, and suitable for further analyses. Data validation of our alternative data sources is separate to our validation of traditional data. See our [Traditional data aggregates in consumer prices methodology article](#) for more information.

We first perform file-level checking, ensuring the files we receive follow correct file formatting. We perform the following types of checks:

- confirm files follow expected naming conventions
- ensure the correct number of files has been received
- validate that file size is within expected ranges

Once data passes file-level assurance, the second type of checking is done at a row-level. Detailed checks are applied to verify the quality of the raw data. We perform row-level checks, including:

- checking that the number of rows is within acceptable thresholds
- calculating the sum of the main variables and comparing them with thresholds
- ensuring the null rates for important variables are below the allowed limits
- verifying that categorical values match expected categories
- ensuring that data formats (such as dates and codes) are correct
- confirming that values fall within expected ranges
- checking that unique values match expected counts (such as the number of stores)
- ensuring that validating data covers the correct reporting period
- making sure that no duplicate rows or identifiers exist
- confirming that all required variables are present, with no extras

When any of our checks fail, we have follow-up procedures that are consistent with the size of the issue. Any file-level check failures result in a request for data redelivery, as these files cannot be used by our processing pipelines. By contrast, some row-level checks may result in desk research to improve our understanding of the issue. For example, an increase in row count may be explained because additional market activity was created during a bank holiday. Other row-level checks may prompt us to discuss the matter further with the data supplier.

After we are satisfied that the data are of a sufficient quality, we "stage" the data. This involves applying data processing to create new datasets that are easier to analyse than the initial "raw" datasets we receive. The processing includes:

- creating consistent variable names for the same concepts across all datasets
- standardising product size unit of measurements (for example, converting from kilogrammes into grammes)
- combining multiple features to create a single product definition variable

Staging allows us to reuse the same data processing code across the various data supplies we have.

Defining strata

In our [Scope and coverage of consumer prices indices methodology article](#) we described the aggregation structure we use within our consumer price statistics.

Our alternative data sources need to fit within this structure. This means we need to make decisions on how to split the data into these strata. We will need to decide:

- the consumption segments we will use to cover the data source
- whether to apply a regional and retailer stratification
- whether to create any additional custom "extra" strata for the category

We will now describe the decisions we have made on stratification for each category.

Groceries

- Consumption segment – approximately 150 consumption segments spanning food, drinks (both alcoholic and non-alcoholic) and tobacco
- Region – 12 UK regions
- Retailer – each data supplier uses a separate retailer stratum
- Extra strata – no further alternative data extra strata

For groceries, we use a dual collection method that involves using both alternative and traditional data to form our indices; on top of the alternative data retailer strata, we also have strata (and items) covering the traditional data portion of the collection. This is described in more detail in our [Traditional data aggregates in consumer prices methodology article](#). The methods described in the remainder of this methodology apply only to calculating the alternative data elementary aggregates.

Second-hand cars

- Consumption segments – 2 covering petrol and diesel cars
- Region – no region stratification
- Retailer – no retailer stratification
- Extra strata – age groups and car manufacturers

Rail fares

- Consumption segments – 5 covering different ticket types
- Region – 11 GB regions
- Retailer – no retailer stratification
- Extra strata – no further extra strata

For rail fares, the LENNON data only cover Great Britain, so we also have a separate consumption segment that captures rail fares in Northern Ireland.

Classification

Now that strata have been defined, we need to align observations within our data with these definitions. This process is known as classification. Each observation is assigned to a consumption segment, and where relevant, a region, retailer and any extra strata. Classifications at the United Nations' [Classification of Individual Consumption According to Purpose \(COICOP\) levels](#), which are ranked higher than the consumption segment level within our aggregation hierarchy, are automatically inherited through the consumption segment.

For example, for rail fares, in line with the strata definitions we have described, we may classify an observation as:

- Consumption segment: advance single
- Region: Wales
- No retailer stratification
- No extra stratification

Across our alternative datasets, we classify most observations automatically using pre-existing variables in the data or by deriving new variables. For instance, for grocery scanner data, we can use postcodes to assign observations found within store outlets to their corresponding region.

Classifying grocery scanner data to consumption segment is the one exception where more work is required. This classification is currently done through manual assignment. Head office staff use a labelling application to manually assign each product (and therefore all underlying observations) to a consumption segment. They have access to:

- a recommendation system, offering a shortlist of probable options
- a search feature to help navigate the labelling options
- guidance on how to handle edge cases

They will make the final decision on which consumption segment the product is assigned to.

Defining a month

For grocery scanner data, we only use data from the first three full weeks of a month. This is because grocery scanner data are typically provided weekly-aggregated – we cannot assign weeks that overlap two months, to either month. Using a three-week definition across all our grocery scanner retailers gives us a consistent definition of a month and gives us sufficient time to scrutinise such a big area of the basket.

For other alternative data sources where our data are daily-aggregated, such as rail fares and second-hand cars, we use expenditures and quantities from across the entire month.

Calculating unit values for products

There are often multiple rows per product each month in an alternative dataset. For example, in grocery scanner data, there may be 60 rows in a monthly dataset covering three weeks and 20 outlets the product is sold in. We call these rows "observed expenditure and quantities", representing sales figures observed at a particular time and outlet.

Our index method requires each product to be represented by a single price and quantity each month. Therefore, for each product, we calculate:

- a "representative price" (or "unit value") by dividing the product's total observed expenditure by its total observed quantity
- a "representative quantity" by totalling the observed quantities sold

These calculations are shown in Formula 1.

Formula 1. The formulae for calculating representative prices and quantities from observed expenditures and quantities

$$p_i = \frac{\sum_j e_{i,j}}{\sum_j q_{i,j}}$$

$$q_i = \sum_j q_{i,j}$$

Where:

- p_i and q_i are the representative price and quantity for product i
- $e_{i,j}$ and $q_{i,j}$ are the j-th observed expenditures and quantities for product i

A product may be sold at different prices in different places and at different times of the month. A representative price will generally not match any of these individual prices, but instead represents the average price paid by consumers for the month within the stratum.

For groceries, we go a step further and perform a size-adjustment by calculating the price per unit of a product. This is shown in Formula 2.

Formula 2. The formulae for calculating size-adjusted representative prices and quantities (modified from Formula 1)

$$p_i = \frac{\sum_j e_{i,j}}{\sum_j (q_{i,j} \times s_{i,j})}$$

$$q_i = \sum_j (q_{i,j} \times s_{i,j})$$

Where:

- $s_{i,j}$ is the j-th observed size for product i

To give an example, without size-adjustment we would track the price of a can of beans, whereas with size-adjustment we track its price-per-gramme. This quality adjustment means we will capture the influence of both price and size changes on price inflation.

Some products are sold loose by a weight chosen by the consumer. For instance, instead of buying a bag of carrots, a consumer may choose to put several loose carrots in their basket, with the total weight (and therefore price) of those carrots calculated at the till. Where this happens, we divide the product's total expenditure (across all customers) by the total weight sold, to similarly obtain a price-per-gramme. If the total weight sold is not available, the product is excluded from our aggregates as we are not able to calculate a meaningful price for the product.

Defining a product

As we have explained, derivation of representative prices (unit values) involves averaging many observations into a single product price. It is important that when we perform this averaging, we do it over observations that are "homogeneous in quality".

For example, suppose we combined observations for a 60-pence Pink Lady apple and a 40-pence Granny Smith apple into a single "combined apple product". Even if prices remain unchanged for each individual apple, the representative price for the "combined apple product" may vary because of the proportion of people buying each apple type. This shows unit value bias, where price movements are affected by compositional effects rather than pure price change. The two apples differ in quality (so are not homogeneous), so their representative prices should be tracked separately.

To avoid unit value bias, we need to define our products to include observations of (approximately) consistent quality. To do this, we use a combination of variables which, when taken in conjunction, group together transactions that are relatively homogeneous.

As well as homogeneity, we need to ensure that products are stable. If we are too restrictive in our product definition, there is a risk we will have products which lack observations in some months, making us unable to calculate the price relatives we use in our measures.

To define a unique product for groceries we use the following variables:

- Stock Keeping Unit – often abbreviated as SKU, this is a unique product identifier used by retailers to mark a product
- Store type
- Unit of measurement

Units of measurement are standardised when we stage the data. This means that if the retailer changes their designation of the unit of measurement, then we could still track the product. So, if a retailer changes a product from 1,000 grammes to one kilogramme, then we would convert the product back to 1,000 grammes. Since the product then retains a consistent unit of measurement, we can continue to track price changes within this product.

For second-hand cars we use:

- Car model
- Car mark
- Engine size
- Mileage bin
- Transmission type
- Body type

For rail fares we use:

- Origin station
- Destination station
- Route
- Product name
- Discount type
- Fare product group

Note that, since the data are divided into strata then defined by product, there may be instances where the same product identifier is found in more than one stratum. For example, since second-hand cars are stratified by fuel type, age and make, the same petrol car identifier may be found in the different age strata based on the age of that car. Therefore, strata also form an implicit part of the product definition.

In our Pink Lady and Granny Smith example each product would be defined by different SKUs, and SKU is used within the product definition for grocery scanner data, ensuring that the two apple types are grouped into distinct representative prices.

Discounts and refunds

Historically, some discounts have been difficult to account for when measuring product price change in our locally collected data sources; this is because of a lack of information on discount take-up rates. For example, a box of grapes may cost two British pounds (£) in January. In February, it may continue to cost £2 but may be available at a lower price of £1 for members of a loyalty card scheme.

Without knowing the proportion of consumers who are members of the loyalty scheme, it is not possible to know the true average price paid by the consumer in February. Under traditional data practices, we would therefore compare only pre-discounted prices, where no price change is observed. See our [Traditional data aggregates in consumer prices methodology article](#) for more detail.

In scanner data, the unit values calculated group together both consumers who bought the product at a discounted price, and consumers who bought it at full price. For example, if one person bought the box of grapes at the full price of £2, and one person bought the box of grapes at the discounted price of £1, then our data would reflect this as £3 of expenditure and two units sold. The unit value price would be £1.50 meaning that a true average price paid by the consumer is measured.

This means that scanner data typically accounts for price promotions, multibuy offers and loyalty scheme discounts, expanding the range of discounting behaviour we can account for compared with traditional data methods.

When a product has not been consumed and is then refunded, the product is therefore not considered in scope for inflation measurement. Some retailers provide refunds as separate rows, and where this is true, we can aggregate sales with refunds to remove refunded products, provided that the original sale and refund occur in the same month.

Relaunch linking – grocery scanner data only

Sometimes manufacturers remove a product from the market and then "relaunch" it. A relaunched product may differ to the original variant in price, quality, or both. The product may be relaunched for a variety of reasons, such as, because of recipe changes, changes to packaging, or a change in weight. Products can increase in weight (especially because of promotional offers) or decrease in weight, perhaps as part of a cost-cutting strategy (often described as "shrinkflation").

When a relaunch occurs, some retailers group together the original and relaunched product under the same SKU. This results in both the original and relaunched product having the same product ID, allowing us to capture any quality-adjusted price changes in the product (because of us using size-adjusted prices). However, some retailers assign different SKUs to the original and relaunch products and, if this is not corrected, our indices would not capture any (quality-adjusted) price change associated with the relaunch.

Relaunch linking describes our process of linking together the original product SKU and the relaunch SKU. To do this, we compare the product name of the new product against previous products and shortlist a handful of closely matching names. Head office staff will then decide whether the new product is a relaunch of any products on the shortlist.

In theory, indices can be both upwardly and downwardly biased by not accounting for weight changes. However, empirical evidence shown in our [Research into the use of scanner data for constructing UK consumer price statistics article](#) suggests indices tend to be biased down by not accounting for relaunching behaviour.

Defining invalid strata

Since there are thousands of scanner data strata because of the various permutations of consumption segment, region, and retailer, some strata would not produce reliable indices for a full production year because of low product count. To avoid an overreliance on imputing these strata, we choose to remove them.

To do this, we run a test to check whether the stratum could have been calculated over the two years leading up to the production year. We calculate a 25-month GEKS-Törnqvist window covering the period from the January two years previously to the current January. If a lack of product matches means we cannot form a mathematically valid index in any one of the 25 months within the window, we describe it as an "invalid stratum".

See [Section 5: Methods of the GEKS-Törnqvist](#) for more information on how the GEKS-Törnqvist is calculated.

Invalid strata are removed from our aggregation structure and weights are redistributed to other strata. This is done during aggregation and is explained in greater detail in our [Higher-level aggregation and weights in consumer prices methodology article](#).

Ordinarily, if a retailer is represented by alternative data sources, then we do not include price quotes for this retailer in our traditional strata to avoid duplication. However, when an alternative data retailer stratum is set as invalid – resulting in not using alternative data for that retailer – then we may then use price quotes from this retailer within the traditional stratum instead, if available. This ensures the retailer is being captured in at least one part of the collection.

This invalid strata process is automatic. However, we can also manually re-assign some strata as invalid as a contingency option. This can happen in situations where a data issue has arisen within the stratum and would likely bias the index, and where we have been unable to resolve this through other means (such as a data resupply).

In instances where the retailer is no longer able to deliver any data, we have separate contingency options as referred to in our [Higher-level aggregation and weights in consumer prices methodology article](#).

Data cleaning

The data cleaning process involves three main levels of data cleaning.

Firstly, any rows pertaining to "invalid strata" are removed from the data.

Secondly, we remove out-of-scope observations (which we sometimes refer to as "junk filtering"). The filtering we perform depends on the category type. For example:

- for groceries we remove observations lacking size information
- for second-hand cars we remove motorbikes and heavy goods vehicles
- for rail fares we remove underground fares, as prices for these are captured elsewhere

Thirdly, we perform two types of outlier detection. The first is used for all data sources and removes extreme price changes, as in representative prices that have gone up or down by some threshold. This gives enough room to account for promotional offers, without implausible price changes distorting our indices.

The second outlier detection method is used for grocery scanner data and accounts for dump pricing behaviour. Dump prices occur when a large price drop corresponds with a large drop in sales and they are usually associated with end-of-lifecycle products. This is where remaining stock is sold at extremely low prices, but the average customer cannot benefit from this sale because only limited stock are sold at these clearance prices. We identify and filter out dump prices when a product price is reduced by more than 50% and the quantity sold falls by more than 90%.

4 . Calculating elementary aggregate indices

As we have prepared the data using the methods outlined in [Section 3: Data processing](#), we can now calculate elementary aggregate indices for each of our "valid strata".

This section describes two main concepts:

- the 25-month GEKS-Törnqvist window
- the splicing procedure used to extend our index series

After introducing these concepts, we explain how they are applied to calculate our final indices.

The GEKS-Törnqvist window

A 25-month GEKS-Törnqvist window is a 25-month index series where each month measures price change relative to the first (base) month in the window.

To show this, consider this synthetic 25-month GEKS-Törnqvist index series running from March 2024 to March 2026:

100	99.4	100.1	101	102.4
104	104.9	104.1	104.9	105.5
105.6	107.1	107.4	108.5	108.1
108	108.2	108.2	108.6	110
110.1	111	111.8	113.2	113.6

In this example, February 2026 (the 24th month) has an index of 113.2, meaning that prices rose by 13.2% between March 2024 and February 2026.

The GEKS-Törnqvist is described as a "multilateral" index method because it uses information from all months within the window, rather than comparing only two months at a time. This allows the index to capture the influence of price movements for products that appear or disappear within the window – movements that may not have been captured by comparing only two months alone.

We provide further detail on how GEKS-Törnqvist windows are calculated in [Section 5: Methods of the GEKS-Törnqvist](#).

Splicing

Because each GEKS-Törnqvist window covers only 25 months, we must extend the index series when a new month becomes available. We achieve this through splicing.

Splicing allows us to combine:

- an existing index series which runs up to month t , and
- a newly calculated index series which includes month t plus 1

We do this by using the overlapping indices between the two series to adjust the new index series, so that its base period is consistent with the first window before appending the new adjusted month.

For example:

- Index series 1 – January 2024 to May 2026
- Index series 2 – June 2024 to June 2026

The differences between the two series in the overlapping period – June 2024 to May 2026 – are used to align the two index series. The June 2026 index is then appended, creating a combined series running from January 2024 to June 2026.

In practice, we use the "mean splice on published" approach to splicing. See [Section 5: Methods of the GEKS-Törnqvist](#) for further technical details.

How we calculate elementary aggregates

With the GEKS-Törnqvist and splicing concepts established, we can outline our monthly production process. Each month, for each "valid stratum", we:

- roll the 25-month window forward, calculating a new GEKS-Törnqvist window
- extend the overall index series by splicing the new index on

For example, for the 2026 production year, the process would work as follows:

February

- Calculate the January 2024 to January 2026 GEKS-Törnqvist window
- This creates a 25-month index series – for reference call this "the publishing series"

March

- Roll forward, calculating the February 2024 to February 2026 GEKS-Törnqvist window
- Splice this onto the publishing series calculated in February
- This creates a 26-month series running from January 2024 to February 2026, thus extending the publishing series into February 2026

April

- Roll forward, calculating the March 2024 to March 2026 GEKS-Törnqvist window
- Splice this onto the publishing series calculated in March
- This creates a 27-month series running from January 2024 to March 2026, extending the publishing series into March 2026

We can repeat this process for all remaining months in the 2026 production round.

Each month, after splicing, the publishing series remains referenced to January 2024. For publication, however, indices for February 2026 onwards must represent price change relative to January 2026. We therefore re-reference the entire series to January 2026. This is done in the aggregation pipeline and is discussed in more detail in our [Higher-level aggregation and weights in consumer prices methodology article](#).

Imputation

Each time we calculate one of our GEKS-Törnqvist windows, there is a risk that one or more of the indices within the window are incalculable because of a lack of product matches. We use imputation techniques to fill these gaps.

We only calculate elementary aggregates for valid strata – see the "Defining invalid strata" subsection of [Section 3: Data processing](#) for more information. Valid strata must have a full set of indices for the initial GEKS-Törnqvist window, so no imputation is required for the window calculated in February.

Each GEKS-Törnqvist window following February contains 24 months which overlap with the publishing series, and one month which does not. If one of the overlapping index points has a missing index, then we use the month-on-month rate of the published series to adjust the previous index within the GEKS-Törnqvist series.

If the final month of the GEKS-Törnqvist series is missing, then this imputation cannot be done using the publishing series. Instead of extending the publishing series into the new month, we set the new month to a null value then use the imputation techniques described in our [Higher-level aggregation and weights in consumer prices methodology article](#) to impute this index during the aggregation process.

5 . Methods of the GEKS-Törnqvist

We have outlined the practical application of the GEKS-Törnqvist and mean splicing in producing our elementary aggregates; we have not explored the underlying mathematics in detail. This section summarises those methodologies. Further explanation and worked examples are available in our [Introducing multilateral index methods into consumer price statistics methodology](#).

The Törnqvist index

The Törnqvist index method is a symmetrically weighted geometric average of price relatives. The price relative is a ratio of the current and base month prices, while the weights are an arithmetic average of the base and current period expenditure shares. The formula is given in Formula 3.

Formula 3. The formula for calculating a Törnqvist index.

$$Tq(a, b) = \prod_i \left(\frac{p_i^b}{p_i^a} \right)^{0.5 \times (s_i^a + s_i^b)}$$

Where:

- p_i^t is the price of product i at time t
- s_i^t is the expenditure share of product i at time t

The GEKS-Törnqvist

The GEKS-Törnqvist measures the price change of an aggregate using the prices and expenditure shares from many periods (over a “window” of months) and is therefore described as a “multilateral index method”. We can understand the GEKS-Törnqvist as a geometric average of many “pairs” of Törnqvists. This is where each pair measures a different chained Törnqvist from the base month to the current month. The formula for the GEKS-Törnqvist is given in Formula 4.

Formula 4. The formula for the GEKS-Törnqvist

$$GEKST(a, b, W) = \prod_{m \in W} (Tq(a, m) \times Tq(m, b))^{\frac{1}{\dim(W)}}$$

Where:

- a and b are the base and current months
- W is a window of months which contains a and b
- Every m is a month within window W
- $Tq(x, y)$ is the Törnqvist from month x to month y
- $\dim(W)$ is the number of months within the window W

To give an example, if we had a 25-month window from $t1$ to $t25$, then the GEKS-Törnqvist from $t1$ to an arbitrary month b within this window is:

$$GEKST(t1, b, \{t1, \dots, t25\}) = (Tq(t1, t1) \times Tq(t1, b))^{1/25} \times (Tq(t1, t2) \times Tq(t2, b))^{1/25} \times \dots (Tq(t1, t25) \times Tq(t25, b))^{1/25}$$

By replacing current month b with each month within the window $\{t1, \dots, t25\}$, we can calculate indices for every month within the window, all referencing base month $t1$.

Splicing multilateral indices

We used Formula 4 to calculate a 25-month GEKS-Törnqvist index series from $t1$ to $t25$. To extend this index series beyond $t25$, we will use splicing – specifically, mean splicing.

First, we calculate an initial publishing index series based on an initial GEKS-Törnqvist window. This is shown in Formula 5.

Formula 5. The formula to initialise the publishing series, based on an initial GEKS-Törnqvist window (for $t = t1..t25$)

$$Pub(t1, t) = GEKST(t1, t, \{t1, \dots, t25\})$$

In other words, for the first 25 months, the publishing index series exactly matches the first GEKS-Törnqvist window we calculate.

We will now “roll” the window on and calculate a second GEKS-Törnqvist window, this time between $t2$ and $t26$. We extend the publishing series into the new month by multiplying the final GEKS-Törnqvist index in this new window by a geometric average of ratios between the overlapping indices, between the new GEKS-Törnqvist window and the publishing series. This is shown in Formula 6.

Formula 6. The splicing formula to splice a new GEKS-Törnqvist index onto the existing publishing series

$$Pub(t1, t) = GEKST(t - 24, t, \{t - 24, \dots, t\}) \times \prod_{i=t-24}^{t-1} \left(\frac{Pub(t1, i)}{GEKST(t - 24, i, \{t - 24, \dots, t\})} \right)^{1/24}$$

This is best illustrated using this example – we have already calculated a publishing series from t1 to t26, and we want to extend it into t27. We have used a rolling window to calculate a new GEKS-Törnqvist between t3 and t27 and wish to splice this on. Synthetic data for this is given in Table 2.

Table 2. Synthetic data for a new GEKS-Törnqvist window that we want to splice onto the publishing series

t	t1	t2	t3	t4	...	t26	t27
Pub(t1, t)	100	101.2	102.3	102.4	...	111.3	
GEKS-T(t3, t, {t3, ..., t27})			100	102.1	...	110.4	111.2

Source: Office for National Statistics

Then to extend the publishing series into t27, we perform the following calculation:

$$Pub(t1, t27) = 111.2 \times \left(\frac{102.3}{100} \right)^{\frac{1}{24}} \times \left(\frac{102.4}{102.1} \right)^{\frac{1}{24}} \times \dots \times \left(\frac{111.3}{110.4} \right)^{\frac{1}{24}}$$

In Formula 6, we calculate a geometric average of many different ratios between the publishing series and the GEKS-Törnqvist window, which is then used as an adjustment factor. There are other splicing approaches (such as the window, movement and half splices), where only one of these ratios is used – however, we are not currently using these methods.

6 . Definitions

Aggregates

Aggregates (or "strata") are classifications into which the raw data can be separated. The strata "region" and "shop type" within item are generally used for the Consumer Prices Index including owner occupiers' housing costs (CPIH), Consumer Prices Index (CPI), Retail Prices Index (RPI) and the Household Costs Indices (HCIs). The data within each stratum are combined, and the resulting indices for each of the strata are then combined using stratum weights.

Alternative data

These are larger, automatically collected data sources. We have introduced several alternative data sources into the calculation of our consumer price indices since the early 2020s.

Basket

A convenient way to understand the nature of consumer price inflation statistics is to envisage a very large shopping basket comprising all the different kinds of goods and services bought by a typical household. As the prices of individual items in this basket vary, the total cost of the basket will also vary – consumer price statistics measure the change from month to month in this total cost.

Base prices

Our index methods measure price change between two months: the base month and the current month. Base prices are the prices that are used to represent the price of a product in the base month. This representative price may be a single sampled price, or an average of many different prices.

Consumption segment

A consumption segment is broader in scope than individual items but is still intended to be relatively homogenous, with respect to price change.

For example, the consumption segment "rice" includes various representative items, such as dry rice, microwaveable rice, and rice snacks (like rice cakes) from the traditional data collection. For alternative data sources, the consumption segment includes all rice products that have been sold.

In areas of the basket where we are not using alternative data, a consumption segment matches one of our representative items exactly.

Chain linking

A "chain link" is the mechanism we use for connecting indices with different baskets or weights. The calculation relies on a link period (December and January in CPI, CPIH and the HCIs). Subsequent index movements are "chained" to this link period by multiplication.

Current price

Our index methods measure price change between two months: the base month and the current month. Current prices are the prices that are used to represent the price of a product in the current month. This representative price may be a single sampled price, or an average of many different prices.

Elementary aggregates

The set of indices calculated at the very first stage of aggregation.

GEKS-Törnqvist

The GEKS-Törnqvist is described as a "multilateral" index method. It uses information from all months within a window, rather than comparing only two months at a time. This allows the index to capture the influence of price movements for products that appear or disappear within the window that may not have been captured by comparing only two months.

Price quotes

Individual prices collected through traditional data collection for specific products or varieties that households buy.

Products

Products, or "varieties", are the varieties of goods or services available within an item specification. For example, automatic washing machines with different specifications are produced by different firms, but they are all automatic washing machines.

Splice (splicing)

Splicing, like chaining, is a method for connecting indices based on different baskets or weights.

Strata (stratum)

Strata (or "aggregates") are classifications into which the raw data can be separated. The strata "region" and "shop type" within item are generally used for the CPIH, CPI, RPI and HCIs. The data within each stratum are combined, and the resulting indices for each of the strata are then combined using stratum weights.

Weight

A factor by which a component is multiplied to reflect the level of consumers' expenditure on that component.

7 . Related links

[Consumer price indices technical guidance](#)

Methodology article | Last revised 25 March 2026

How measures of consumer price inflation and associated indices are compiled.

[Consumer price inflation, UK](#)

Bulletin | Released monthly

Price indices, percentage changes, and weights for the different measures of consumer price inflation.

[Household Costs Indices for UK household groups](#)

Bulletin | Released quarterly

Household Costs Indices, 12-month growth rates, expenditure shares, and contributions for UK household groups and all households.

8 . Cite this methodology article

Office for National Statistics (ONS), published 25 March 2026, ONS website, methodology article, [Alternative data aggregates in consumer prices](#).