Office for National Statistics

Traditional data aggregates in consumer prices

How we collect, process and validate price data from traditional sources to produce elementary aggregates. Part of 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

Our inflation measures are separated into components called "aggregates" (or sometimes "strata"). These exist within a hierarchy, where "elementary aggregates" form the lowest level.

For most areas of the consumer prices basket, a sample of prices is manually collected through "traditional" sources in store, online and by phone.

In this article, we describe how we calculate elementary aggregates from these traditional data sources, including how we collect, process, and use these data 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 prices indices technical guidance
- Scope and coverage of consumer prices indices
- Special case aggregates in consumer prices
- Alternative data aggregates in consumer prices
- Rents and owner-occupier housing aggregates in consumer prices
- Higher-level aggregation and weights in consumer prices
- Calculating the Retail Prices Index article
- How we publish our consumer price outputs

2 . Data sources

There are several different ways that we sample and collect prices for our traditional sources of consumer prices indices data. These can be grouped into two high-level categories: local collection items and central items.

Local collection items

The "local collection items" category covers price collectors going to shops in various locations across the country to collect prices for items either in store, online or by phone. These are supplemented by prices collected by head office staff for shops with national pricing policies, and prices collected via phone, email or online for local services such as plumbers, taxi hire and hotels (known as the "regional services collection"). Following data collection, these prices are combined before further processing.

Central items

The "central items" category covers around 150 items for which the prices are collected centrally. These items make up approximately 30% of the Consumer Prices Index including owner occupiers' housing costs (CPIH) basket, and 50% of the Retail Prices Index (RPI) basket by expenditure weight. The elementary index calculation is carried out separately to the locally collected items.

3 . Sampling for the local collection

We introduce the concept of "consumption segments" and explain how individual representative "items" are chosen for the consumer basket in our Scope and coverage of consumer prices indices article. Consumption segments are broader in scope than individual items but are 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.

In areas of the basket where we do not use alternative data, a consumption segment matches one of our representative items exactly. The consumption segment (or item) may then be stratified by region and shop type.

In areas of the basket where we use alternative data – potentially together with traditional data – a consumption segment is represented by one or more items. For the alternative data part of the collection, we can use a near-census of products, to match the consumption segment. For the traditional data part of the collection, we can represent the consumption segment with the chosen items.

Constructing a fully accurate consumer price index would require recording the price of every version of every product and service in every physical and online retailer in the UK each month. Alternative data help us to progress towards this objective. However, price sampling remains essential, because this is not possible for many areas of spending.

This section describes the additional sampling stages used in local price collection, including:

- the selection of locations
- the selection of outlets within those locations
- the selection of product varieties ("price quotes") within each representative item

It also covers the approach for sampling shops with regional or national pricing policies and the regional services collection. For the rest of this section, we refer to representative items, rather than consumption segments, because they reflect the level of hierarchy used for our other sampling frames.

For local collection items, around 150,000 price quotes are obtained monthly by these sampling methods. As only a sample of prices are recorded, there is inevitably some sampling error in measuring consumer price inflation. Our ongoing work to review the quality and relevance of our consumer price statistics, including our sampling methodology and measures of accuracy, is described in our Consumer prices development plan article series.

For more information about historical reviews of sampling arrangements, please see Section 5.6 in our Consumer Prices Indices Technical Manual, 2019.

Sampling of locations

Producing a location boundary

Since 2015, we have used the "location-allocation" method to identify and define location boundaries around areas of high retail activity, known as hotspots. This forms a sampling frame from which we select locations for field collection. A location boundary sampling frame, based on retail data, is used to create locations that are representative of both the retail turnover and geographic areas of the UK. Similar to previous methods, location-allocation uses geographic information systems (GIS) software and the following steps:

1. The UK is split into 500 square metre grids and the latest Inter-Departmental Business Register (IDBR) data on number of outlets, employees, expenditure, and retail turnover are assigned to each grid.

2. Hotspots are used as the centre of the new location boundaries; these are identified as areas with high retail turnover.

3. Outlets on the IDBR that fall within an impedance distance of each hotspot are mapped; the impedance distance is set at 3.5 miles across the UK and 0.5 miles in London, and reflects ease of travel, rather than a simple straight-line radius.

4. Each outlet is linked only to its nearest hotspot, to avoid any overlap in coverage.

5. Polygons are grown around the outlets for each hotspot, forming outlines of the location boundaries.

6. The boundaries are adjusted to fit real-world features (such as roads, railways and waterways) using Ordnance Survey map information, maintaining the retail turnover and number of outlets in each location, while reducing the amount of space that has no outlets.

Further details of the sampling procedures used before 2015 can be found in Section 4 of our Consumer Price Indices Technical Manual, 2014.

Location selection

Location selection takes place separately within each region. We use probability proportional to size (PPS) systematic random sampling, with a size measure that is relative to locations' retail sector activity. PPS systematic random sampling is a type of interval sampling. The number of locations to be selected in each region is determined as the proportion of national expenditure taking place in that region, multiplied by the total number of locations to be visited nationally (141 locations in total).

We first list all shopping locations within each region. This forms the basic sampling frame. We then modify this to ensure that a full shopping basket (all the items in the sample) can be collected in each location.

We exclude locations with too few outlets and locations where field auditor experience suggests that it is not possible to collect a complete basket of goods. Locations that are not large enough to support the collection of a full basket on their own are provisionally paired with a nearby excluded location. These locations can be merged to form a single collection area, from which it will be possible to collect prices for a complete basket of goods.

Using this list of locations, PPS systematic random sampling is then performed. We first calculate an interval value using the cumulative total of the size measure, divided by the number of sampled locations required. We then calculate a starting point by generating a random number between 0 and 1 and multiplying this by the interval value. The location whose size variable contains the starting value is selected as the first location.

We then generate a second random number by adding the interval value to the starting point. This is used to select the second location, by using the location whose size variable contains the second random number. We repeat the process of adding the interval value to the previous random number, and selecting the corresponding location, until the requisite number of locations has been sampled. This process is shown in Table 1 using the following worked example, with turnover used as the size variable.

Table 1: How we carry out probability proportional to size systematic random sampling
Probability proportional to size (PPS) systematic random sampling procedure, including example sampled locations, number of outlets, turnover and cumulative totals

Location name	Number of outlets	Turnover	Cumulative total
Location A	607	5,377	5,377
Location B	306	2,486	7,863
Location C	264	2,265	10,128
Location D	449	4,006	14,134
Location E	322	2,589	16,723

Source: Consumer Prices Indices from the Office for National Statistics

Worked example of location selection

We want to select two locations. PPS systematic random sampling gives each location a chance in proportion to its turnover.

Step 1 is to calculate the interval value:

- total turnover is 16,723; we need two selections
- interval value equals the total turnover divided by the number of selections
- interval value is 16,723, divided by 2, which equals 8,361.50

Step 2 is to generate the random starting point:

- choose a random number between 0 and 1 (0.33905)
- random starting point is the interval value multiplied by the random number
- random starting point is 8,361.50 multiplied by 0.39905, which equals 3,336.66
- 3,336.66 is Selection Point 1

Step 3 is to calculate the Selection Point 2:

- Selection Point 1 is 3,336.66
- Selection Point 2 is Selection Point 1, plus the interval value
- Selection Point 2 is 3,336.66, plus 8,361.50, which equals 11,698.16

Step 4 is to identify which two locations contain these values:

- Selection Point 1 (3,336.66) is between 0 and 5,377, so the selected location is A
- Selection Point 2 (11,698.16) is between 10,128 and 14,134, so the selected location is D

Location rotation and re-enumeration

It is not feasible to select and list (enumerate) all outlets for an updated set of locations every year. However, maintaining a fixed sample of locations (141 locations across the UK) and enumerating only once reduces the total number of locations used for price measurement. This would result in enumeration lists that would:

- contain outlets that are no longer operating
- omit outlets that had opened since the enumeration
- miss regional shifts in consumer expenditure

Our practical approach is to update a sample of around 30 locations each year. We either exclude a location and replace it with a new one (rotation) or refresh the list of outlets in the existing location (re-enumeration). Locations are enumerated in the year that they are sampled and then introduced into the collection the following year, while the basket is updated (as described in our Scope and coverage of consumer prices indices article). They should remain in the sample for four or five years, so that each location is refreshed either through rotation or re-enumeration once in a five-year period cycle.

Sampling of outlets

Enumeration of the selected locations is carried out by price collectors. They visit the postcodes in each location and note details of all retail outlets found (up to a limit of 1,500 outlets per location) to produce a sampling frame. The details noted for each outlet include:

- the outlet address
- the outlet postcode
- the range of items sold
- its size, if the outlet is a shop
- whether it is independent or part of a multiple chain, if the outlet is a shop (multiple chains are defined as having more than 10 outlets across the whole country, as identified in the IDBR data)

Shops with national pricing policies are excluded from the enumeration, as outlined in the subsection on Sampling of shops with regional or national pricing policies within this section.

Turnover is the ideal size measure of an outlet for PPS sampling. However, this is not readily available, so the net retail floor space (estimated by the outlet enumerators) is used as a proxy. For department stores and other shops selling a wide variety of goods, the floor space devoted to each commodity group is measured.

The appropriate code indicating what each shop sells is assigned to each outlet, based on the appropriate classification. Using this, outlets are classified by commodity group and, where appropriate, by shop type (multiple or independent). This is not a true stratification; an outlet may be in more than one stratum if it sells items from more than one commodity group.

For each commodity group, the required number of outlets, plus some reserves (used if an outlet closes down), are drawn from the sampling frame. PPS sampling is used where we know there is a wide range of store sizes, and therefore a wide range of turnover, such as for do-it-yourself (DIY) stores, which may be superstores or local shops.

The following example shows how this works for meat. Items are grouped into commodity groups, so fresh beef and lamb, cooked meats, and fresh bacon, pork and chicken are grouped together. These meat items are sold in butchers, supermarkets and some department stores, which can be multiple or independent shops. There should be one price collected for meat in each location and store type: one from a butcher and one from either a supermarket or a department store that sells food.

A shop holding a closing down sale is treated as already closed and is excluded from the sampling frame; we then find new outlets within the location to replace them. This is because its prices will neither be comparable with previous ones, nor available in the future. Shops selling only second-hand goods are also excluded. Some exceptions to this rule exist where large high street chains have closed down; these are treated on a case-by-case basis.

Selection of products and varieties (price quotes)

For most items, collectors choose one variety that is representative of what people buy in their area, from all products matching the specification of each item to be priced within that outlet. They ask the retailer what the most popular brands are and which of those are stocked regularly. It is important that the same product is priced each month. Collectors must record enough detail of the product, such as make and model, to ensure that it is uniquely identifiable. The number of varieties (quotes) required in each location depends on the typical variance in price change for that item.

The chosen products are reviewed each January to ensure that what is being priced still reflects these criteria. This is because January prices ("base prices") are used to compare prices against for the rest of the index year. If the product being priced is not available for January, one that is available must be chosen so that there is a valid base price for the forthcoming year (see Section 12: Calculating elementary aggregate indices). Prices are collected in January for both the old (if possible) and new products, and for old and new items where these change, to allow for chain-linking.

Local probability sampling

Between January 2004 and January 2014, local probability sampling (sometimes referred to as remote sampling) was used for individual models within outlets for several household appliances, such as washing machines. Further details of how this method was applied can be found in Section 4.5 of our Consumer Price Indices Technical Manual, 2014. We stopped using this method for all items in 2014, because of data quality concerns and because no suitable alternative data source was available. Instead, the guidelines for collecting comparable replacements were improved and are updated annually.

Sampling of shops with regional or national pricing policies

Our local collection also involves collecting prices from major chains of shops with national pricing policies set by head office staff ("central shops"). Branches of these chains are excluded from the outlet enumeration. Some chains enter price data on spreadsheets by email. More frequently, the data are obtained from the company's website. These prices are replicated across regions and then combined with prices for the same items from the in-store collection.

For chains with regional pricing policies ("regional central shops"), collectors visit only a few of their outlets and assume that each outlet reflects their pricing policy within a given region. For these regional central shops, one collection is carried out where the retailer operates in Wales, Scotland, Northern Ireland, and the nine regions in England. This means that a maximum of 12 price quotes will be collected for each item in each retailer. These prices are then combined with prices for the same items from the in-store collection in other outlets.

For both shop types, the collected prices are given extra weight to reflect their market share (see Section 11: Replication factors).

Sampling of service providers and price quotes for regional services collection

Items collected under the regional services collection differ from other local collection items because they are generally not purchased in retail centres, for example electricians' fees, childminder fees, and driving lesson fees. These prices are obtained directly from the relevant businesses or organisations, either by phone, text, email, or by consulting information published on their websites.

The regional services collection sample is designed to provide the equivalent of one or two quotes per item per location. Collectors are instructed to ensure that the selected businesses or organisations meet the required sample counts for each region, which is why these are referred to as regional collections.

Unlike the standard fixed sampling approach, the regional services sample is updated throughout the year to account for the high rate of business entry, exit and restructuring seen in many servicesector activities. When a business becomes permanently unavailable, or requests removal from the sample, a replacement is recruited and a base price is imputed using the price movements of the remaining businesses for that item (see our subsection on Imputation of a new base price in Section 9: Product substitution, quality adjustments and imputation).

Most of these items are collected periodically (see our subsection on Frequency of collection in Section 4: Price collection for the local collection). This balances the need to minimise respondent burden with the need to promptly monitor price changes. Items with more frequent price fluctuations, like theatre admissions, are collected monthly. Other items, like school meals, are collected only at specific points in the year, usually at the start of each school term.

4 . Price collection for the local collection

General procedure

Most local price collection is completed by an external collection agency on a contractual basis, operated to open competition-tendering procedures. Performance indicators are set (for example target collection rates).

For this collection, prices are recorded on hand-held collection devices, which speeds up data processing and transfer, and means that prices are validated interactively as they are entered. This also reduces the number of queries that need to be dealt with when the data are processed in the head office (see subsection on Local collection checks using hand-held collection devices in Section 7: Validation of local collection).

For shops with national pricing policies, and the regional services collection, prices are collected by head office staff.

Choice of index day

The Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI) are intended to reflect average prices over the month. For practical reasons, the local price collection collects prices over at least one working week at, or near, the middle of each month ("index week"). Collectors aim to provide month-to-month consistency by collecting the majority of prices on the Tuesday of index week, known as the "index day".

In February 2018, a second collection day (the Friday before index week) was also introduced for fresh fruit and vegetables, as the prices for these groceries can also change throughout the month. However, these prices do not contribute to the calculation of the RPI (see our Calculating the Retail Prices Index article).

The choice of collection days and the number of weeks between them depends on operational considerations, particularly the timing of bank holidays. Collection days will never fall in an index week that includes a bank holiday Monday, because some prices will need to be collected on this Monday, when outlets may be closed or may charge abnormal prices. The collection dates are not published in advance because of the hypothetical risk that service providers or retailers may change their prices in order to influence consumer price inflation statistics.

Frequency of collection

Price collectors try to collect all prices every month, except for seasonal items when they are not in season (see our Special case aggregates in consumer prices article) and periodic prices, which are only collected every three or four months in each location.

For periodic items, each location is allocated into one of four groups at random. Prices are then collected according to the following timetable:

- Group 1 – January, May and September
- Group 2 – February, June and October
- Group 3 – March, July and November
- Group 4 – April, August and December

For those periodic items that are outsourced to the external collection agency, an additional January collection is also carried out for quotes in groups 2, 3 and 4.

In the months when periodic items are not collected in a location, the previous month's prices are carried forward. Items collected periodically are mainly services in the household and leisure groups, and their prices are known to change relatively infrequently compared with locally collected goods and services.

Methods of payment

The price usually used is that for a cash transaction. This means that charges are usually ignored for:

- paying by instalments
- use of credit cards
- discounts for paying by direct debit

However, this is not always the case: some centrally calculated indices, such as electricity charges, measure the price of several different forms of payment. Discounts for paying by cash should also be allowed for. Value Added Tax (VAT) and compulsory service charges are included, but delivery charges are not. Delivery charges are collected as a separate item.

Indicator codes

Collectors are required to note if there are any special features in the prices recorded. The following codes are used:

- S – sale or special offer (typically explains a reduction in price)
- R – recovery from S (typically explains a price jump); this is not necessarily the same price as before the sale
- N – this is where the original product is no longer available, and the collector selects a new non-comparable product or variety to represent an item (implying that the original product's or variety's base price is not suitable for comparison); for more information, see subsection on Imputation of a new base price in Section 9: Product substitution, quality adjustments and imputation
- C – this is where the original product is no longer available, and the collector selects a new comparable product that has changed but is not substantially different from the old one (C for comparable, implying that the original base price is suitable for comparison); for more information, see subsection on Direct comparison in Section 9: Product substitution, quality adjustments and imputation
- T – temporarily out of stock
- M – item missing from outlet and not likely to be stocked again in the near future
- Q – a special note has been made (Q for query) by the collector for head office staff to examine and respond as required
- W – weight or size change, for example, manufacturer has made a permanent change to the weight of a product – for more information, see subsection on Direct quality adjustment in Section 9: Product substitution, quality adjustments and imputation
- X – comparable item introduced that is on sale
- Z – non-comparable item introduced that is on sale
- P – current price unavailable for the period – this code is only used for the periodic items (see subsection on Frequency of collection within this section) to ensure that, where a price is not available in the selected month, the collectors try to collect the price again in the subsequent month instead of waiting until the following rotation

If the price entered fails a validation check carried out by the hand-held collection device, collectors must enter a message explaining why. These messages and indicator codes are used in the head office at a later stage of the validation process (see subsection on Internal data consistency checks in Section 7: Validation of local collection).

A price should be recorded only if the exact product being priced is on display or in stock at the outlet. For some items, such as furniture, which normally must be ordered, it is acceptable to record the price if the item is available to order.

Unavailable items

If a chosen product is temporarily out of stock when the collector visits, no price is recorded and a T code is used. If it is out of stock for three consecutive months, the collector should choose a replacement product that matches the item description, using an N, C, X or Z code, as appropriate, to inform head office staff carrying out subsequent validation on the replacement. If a replacement product cannot be found, the collector should use an M code. There is an exception to this rule for fruit and vegetables, where collectors will replace out-of-stock items in the first month because of frequent changes in products, for example, country of origin.

Obtaining a price per unit

Some food items, such as cheese, are sold in packs of variable weight, so it may not be possible to find the identical weight each time. In this case, a price per unit weight is collected. If it is not marked, it is calculated from the displayed price and weight. Each month, a pack of roughly the same weight is used, as a lower price-per-unit weight may be charged for larger packs.

If a single good, such as one bar of chocolate, is specified, and it is only available as a multi-pack in January, the price of one bar is computed from that of the multi-pack. The same multi-pack is used in subsequent months. If price collectors are forced to calculate a single good price from a multi-pack price, they are instructed to use the smallest multi-pack (for example, using a two-pack rather than a three-pack).

Subsidies and discounts

For traditional data sources, discounted and subsidised prices are only recorded if they are available to anyone with no conditions of sale, otherwise the non-discounted or unsubsidised price is recorded. This is discussed in our Scope and coverage of consumer prices indices article. Money-off coupons and loyalty cards are excluded. If there is a discount for multiple purchases, only the price of a single purchase is recorded. Where a price reduction on one product is associated with the purchase of another product, this reduction is excluded. However, a reduction associated with a given level of total spending on purchases is included where the cost of the single item being priced lies above that level (for example, the discount "10% off for purchases over £500" would be deducted for a bed priced at more than £500).

Sale prices are recorded if they are:

- temporary reductions on products that are likely to be available again at normal prices
- end-of-season reductions

Prices are not recorded for special purchases of products that are:

- end of range
- damaged
- shop-soiled
- defective

This is because they are deemed not to be of the same quality as, or comparable with, products previously priced, or those likely to be available in future.

Free gifts are not included, this includes:

- plastic toys in cereal boxes
- offers such as "send in 20 tokens for a free pen"
- trading stamps

This is because they are regarded as extras that may not be wanted by consumers.

Prices for items temporarily bearing extra quantities (for example, 20% extra free) are not adjusted to account for the increased quantity.

5 . Sampling and price collection for central items

There are about 150 central items for which the prices are collected centrally. The index calculation for these items is carried out separately from the processing of the local collection data. Selecting this type of collection and calculation is usually dependent on one or more of the following considerations:

- data sources
- data presentation
- frequency of price changes
- the possibility of future fundamental changes to pricing methods

For most of these items, the method of collection and calculation is based on standard methods. Exceptions to this are outlined in our Special case aggregates in consumer prices article and our Rents and owner-occupier housing aggregates in consumer prices article, and some items that specifically contribute to the Retail Prices Index (RPI) are outlined in our Calculating the Retail Prices Index article.

Where feasible, price data are collected over the internet. If this is not possible, price data are collected from one central source (for example, from trade associations and government departments) whenever possible. Data may be requested in writing, by telephone or by email, or may come automatically from a provider's mailing list. Providers may send either a full price list or a tariff sheet from which the relevant prices will be extracted. Some travel fares data are provided in the form of price indices.

Frequency of collection varies across the range of items and depends on when prices are known or are expected to change. The most common frequencies are monthly, quarterly or annual. However, there are some collections that are run three times a year (for example, school meals) and two times a year (for example, some travel fares) that are also included in the timetable.

Some items also have more frequent price collection within a given month. The prices for petrol and oil, which can change regularly throughout the month, are averaged over the month, based on the current prices each Monday.

In February 2026, we also introduced a second pricing day for computer games and overnight hotel accommodation. Computer games have previously been priced during index week, but prices are now also being collected in the following week. Overnight hotel accommodation was priced around six weeks in advance for a stay on collection day, but from February 2026, prices are also being collected at the same time for accommodation on the Thursday after index week. The additional collections were introduced to reduce the volatility resulting from changes in the composition of the best seller charts used in pricing computer games, and from short-term demand pressures on hotel prices.

Reduced prices for payment by direct debit are included in the calculation of some central items, such as electricity charges.

6 . Data processing

The data are then processed ready to be used in price index calculation. First, the data are validated (see Section 7: Validation of local collection and Section 8: Validation of central items). Data validation of our traditional data sources is separate to our validation of alternative data. Further details can be found in our Alternative data aggregates in consumer prices article.

Adjustments, such as base-price imputation, are then made to account for instances of product churn – this is where products often disappear or are replaced with non-comparable versions of a different quality or specification (see Section 9: Product substitution, quality adjustments and imputation).

Products are then assigned to groups known as "strata" (see Section 10: Defining strata for local collection and central items). Finally, replication factors are applied to ensure the data are representative of the population (see Section 11: Replication factors).

7 . Validation of local collection

Local collection checks using hand-held collection devices

Several checks are carried out on data collected on hand-held collection devices to ensure that indicator codes and price values have been entered correctly.

Key performance indicators (KPIs) are agreed with the external price collection agency to ensure that the expected quality and quantity of the price quotes sent to the Office for National Statistics (ONS) are met.

The price collectors are prompted to validate their input on the hand-held collection device when:

- the price entered is outside the minimum and maximum price range (see our subsection on Minimum-maximum checks within this section) and/or is outside the price change check range (see our subsection on Price change checks within this section) and/or is not accompanied by an appropriate comment
- a "C", "N", "Q", "X" or "Z" code has been entered without providing an appropriate comment
- a "C", "N", "Q", "X" or "Z" code has been entered without any amendments to the product description
- a "T" or "M" code has been entered alongside a genuine price, as opposed to a zero price (£0.00)
- an "S" code has been entered when the price has not decreased from the previous month
- an "R" code has been entered when there has not been either an "S", "X" or "Z" code the previous month, or the accompanying price is less than, or the same as, the price from the previous month
- a "W" code has been entered without an accompanying volume or weight change, or a comment has not been provided to describe this change
- letters or special symbols have been incorrectly entered in the price field or prices have been rounded inappropriately (to more than two decimal places)

Minimum-maximum checks

Whenever a collected price quote is below the minimum value or exceeds the maximum value set for that item, a warning message appears on the hand-held device. The minimum-maximum ("min-max") range for each item is derived from valid, non-zero price quotes from the previous month. The lowest and highest valid price quotes across all locations and shops are reduced or increased, respectively, by a set percentage (the price-range percentage) to form the min-max range. The price-range percentage is set by item groups, except for some specific food items.

The agreed percentages are:

- home-killed lamb – 50%
- fresh fruit and vegetables – 100%
- clothing and footwear – 40%
- all other items – 33%

For example, tea bags are included in the "all other items" group, so the price-range percentage is 33%. The cheapest packet of tea bags collected in May was £1 and the most expensive was £8. If a packet of tea bags that costs 70 pence was priced in June, then the min-max check would compute the following range for the quote:

$$minimum = min_{May} \times (1 - food\ price\ range\ percentage)$$
$$minimum = £1 \times (1 - 0.33)$$
$$minimum = £0.67$$
$$maximum = max_{May} \times (1 + food\ price\ range\ percentage)$$
$$maximum = £8 \times (1 + 0.33)$$
$$maximum = £10.64$$

The value collected in June (70 pence) is less than the lowest price collected in May (£1). However, this price quote passes the check because it is between the adjusted minimum and maximum range.

The scaling factor is applied with the aim of reducing the amount of genuine price quotes that fail validation. This helps to account for seasonal sales. For example, many clothing items undergo large price reductions during the January sales, which would automatically fail validation if the scaling factor was not applied.

If the collected price exceeds the min-max range, the collector is asked to confirm the price they have entered is correct or to correct the price if it has been recorded incorrectly. Additionally, the collector can add a "Q" code and an associated comment to accompany the price for review.

Price-change checks

Every price quote that is collected is compared with the price for the same item, in the same shop, that was collected in the previous month. A warning message appears if the month-on-month change exceeds an agreed percentage-change range for that item. Percentage-change ranges are calculated by applying the price-range percentage (as outlined in the previous subsection on Minimum-maximum checks) to the price collected in the previous month. This generates upper and lower bounds for the price in the current month.

For example, if the price of a packet of tea bags in a specific shop in London increases from £8 in May to £9 in June, the hand-held device will compute the following ranges for the quote:

$$lower\ price\ change\ value_{June} = price_{May} \times (1 - food\ price\ range\ percentage)$$
$$lower\ price\ change\ value_{June} = £8 \times (1 - 0.33)$$
$$lower\ price\ change\ value_{June} = £5.36$$
$$upper\ price\ change\ value_{June} = price_{May} \times (1 + food\ price\ range\ percentage)$$
$$upper\ price\ change\ value_{June} = £8 \times (1 + 0.33)$$
$$upper\ price\ change\ value_{June} = £10.64$$

As the June price (£9) is between the lower and upper values of the price change check range, so it passes this check.

If a valid price for the previous month is not found, for example, because the item was out of stock, the check is made against the price two or three months ago. If there is no valid price for the previous consecutive three months, the check is not carried out. If the product had been recorded with an indicator code, then the checks are not carried out.

Internal data consistency checks

Data collected on hand-held devices are transmitted to the head office. They are then combined with data collected for shops with national pricing policies (central shops) and the regional services collection, and put through another series of checks.

Initial checks are carried out to ensure that the data are complete and correct. For instance, checks are run to remove unexpected duplicate prices (for the same item, in the same shop, in the same location) and to ensure that the location, outlet and item identifier codes that accompany each price exist and are valid. If any prices fail these checks, they are returned to the external collection agency for clarification and, if necessary, are corrected and re-submitted.

Once the price data are correct and complete, the quotes are run through a series of validation checks, outlined in the following subsections. Staff within the head office then review all price quotes that fail these checks, along with any indicator codes, the history of the quote, the quote description, and any messages provided by the collectors.

With the information available for each failed price, staff make one of the following decisions:

- accept the price and metadata
- accept the price, but as a new, non-comparable product, and calculate a new base price (see our subsection on Imputation of a new base price in Section 9: Product substitution, quality adjustments and imputation)
- accept the price, but as a new comparable product, and retain the same base price
- change the price, if a price correction is confirmed by the external collection agency or another source (for example, on the retailer's website)
- change the indicator code (for example, to highlight the product is on sale or was previously on sale)
- reject the price (for example, if the price cannot be verified or the product does not meet the item definition)

Quotes rejected in the previous month

Quotes are rejected as part of the internal scrutiny process if:

- they do not meet the specified criteria outlined in the product guidance
- if the price or response indicator cannot be appropriately verified based on the information provided

In the following month, any quotes that had been rejected in the previous month are then flagged for review. This ensures that the price collector has taken appropriate action.

Price indicator checks

Quotes may include indicator codes, for example, "T" (temporarily unavailable) or "M" (missing). Price indicator checks ensure that the price quote should have a corresponding £0 price if one of these indicator codes is used. A quote with any other indicator code should have a price greater than £0.

Message line checks

Message line checks ensure that any quotes with indicator codes that require further information in the metadata that are provided by the external collection agency contain the information required.

Weight changes checks

Any quotes that have been marked with an indicator code "W" indicate there has been a weight or size change in the product. These are flagged for review by head office staff, so that the weight adjustment is appropriately applied (see our subsection on Direct quality adjustment in Section 9: Product substitution, quality adjustments and imputation). These checks would flag if there were any discrepancies in units of measurement that needed to be accounted for, for example, if a weight is recorded in grams one month and in kilograms the next month.

Price relative checks

Any quotes showing extreme price changes either since the base month, or within the previous three months if the product was temporarily unavailable in the month before, are flagged for additional verification by head office.

The current thresholds flag any quote that is more than 1.8 times or less than 0.6 times the January base price. For example, if a chocolate bar cost £1 in January, any price above £1.80 or below £0.60 would be flagged for validation (unless they meet the criteria for auto-acceptance, outlined in our subsection on Exceptions and automatic validation within this section).

Minimum-maximum and price-change checks

The minimum-maximum (min-max) and price-change checks described in the Local collection checks using hand-held collection devices subsection, within this section, are applied again. Some price-range percentages that are used differ from those used on the hand-held devices. These ranges are:

- home-killed lamb – 50%
- fresh fruit and vegetables – test not applied because of volatility
- food – 35%
- clothing and footwear – 40%
- all other items – 33%

Unlike the validation checks on the handheld devices, the only indicator codes that preclude the price change check from being carried out are "N" or "Z" codes (non-comparable products).

Tukey check

An outlier detection process, known as the Tukey algorithm, is used to identify additional outliers, only using the prices that were not already flagged as potentially spurious or those with "N" or "Z" codes.

The Tukey algorithm has been used in the production of our consumer prices statistics since 1987. It produces limits that are:

- intuitively reasonable
- consistent from month to month
- robust in the presence of outliers (so that adding in one or two rogue observations does not substantially affect the limits set by the algorithm)
- robust as data volume changes (that is, limits calculated from a subset of the data do not vary substantially from those calculated on the full dataset)

The Tukey algorithm identifies and invalidates price movements that differ substantially from the norm for a particular item. For seasonal items with erratic price movements, the algorithm looks at price level, rather than price change. It operates according to three parameters, which are set uniformly over all items (though this is not essential).

The algorithm operates by:

- calculating the ratio of current price to previous valid price (the price relative) for each price quote; this stage is omitted for items that are tested by price level, rather than price change
- sorting the set of all ratios into ascending order for each item, and excluding ratios of one (unchanged prices); the prices themselves are sorted for items that are tested by price level, rather than price change
- removing the top and bottom 5% of the list (the figure of 5% is parameter one)
- calculating the trimmed mean, which is the mean of the remaining observations
- calculating the upper and lower "midmeans", which are the means of all observations above or below the trimmed mean
- calculating the upper (or lower) Tukey limit, which is the trimmed mean plus (or minus) 2.5 times the difference between the trimmed mean and the upper (or lower) midmean (the upper and lower 2.5 figures are parameters two and three; they can be set independently if desired, but are currently set to be equal)
- flagging price relatives, or price levels, outside the Tukey limits as invalid

Missing quote for three months or more checks

If a quote has been unavailable for three or more months, it is flagged for review in the month that it re-enters the collection. This ensures that the quote is equal to, or comparable with, the quote that was collected the last time it was available, or that it uses the appropriate indicator code.

New item, shop or location checks

Any quotes collected for the first time are also flagged for review. This may be because they are either a new basket item, or have been collected from a new shop in an existing location or a new shop in a new location. This ensures that the product that has been priced meets the item description specified by the central team.

Other indicator code checks

Other indicator codes are also routinely checked during the validation process. For example, quotes with a "Q" (query) code mean there has been some uncertainty during collection and therefore have been flagged for inspection by head office staff. Any quotes with "N" (non-comparable replacement) or "Z" (non-comparable replacement on sale) codes can also be reviewed, to ensure the product chosen to replace the originally sampled product is still representative of the specified item.

Final check

As a final check, the price relative check is re-run after all other quote validation checks have been completed. A report of all locally collected quotes is issued to senior price analysts for final approval. At this stage, the scrutiniser will seek confirmation that particularly high or low outliers have been checked and may withdraw them from the final calculation if they are not satisfied. Any quotes withdrawn (rejected) from elementary aggregation are checked with the external collector during the following index cycles.

Exceptions and automatic validation

Sequencing procedures mean that if a quote fails a specific check, it will not be included in further checks. For example, if a quote is flagged as having an "N" or "Z" indicator code, that quote will only contribute to the min-max check.

Quotes that initially fail the price relative, min-max or price-change checks are validated automatically when:

- the indicator code shows that the item is on sale in the current month but was neither on sale nor recovering from a sale in the previous month, and the price has fallen by less than 55%
- the item has recovered ("R") from a sale in the previous month, and there has been a price increase of less than 110%
- the price in the current month is the same as the (valid) price in the previous month

Quotes that initially fail the Tukey check are validated automatically only if the price in the current month is the same as the (valid) price in the previous month.

If there is no valid quote in the previous month, quotes from up to three months before the current month can be used for auto-acceptance.

8 . Validation of central items

Central items are validated on a case-by-case basis. Head office staff record price quotes and other descriptors for these items on assigned spreadsheets, some of which are programmed to flag up potentially anomalous observations. Supporting evidence for price quotes collected are stored and subsequently used to verify the entered price during checking. Typically, price quotes for centrally collected items are checked by two members of staff: an initial checker and a sign-off checker.

9 . Product substitution, quality adjustments and imputation

As a measure of price change alone, measures of consumer price inflation should reflect the cost of buying a fixed basket of goods and services of constant quality. For the processing of local collection data, this means that prices for a product in the current month should be compared with prices for the same product in the base month (see Section 12: Calculating elementary aggregate indices).

However, products are often discontinued or replaced with new versions of a different quality or specification, and brand-new products also become available. This requires methods to accurately measure and treat quality change owing to changing product specifications.

When this happens, one of the following methods is adopted:

- direct comparison
- direct quality adjustment
- imputation of a new base price

In all cases, a nominal price in the base month is needed for the new or replacement product; this nominal base price is used until the following January. If the retailer can supply the previous January price of the new product, this can be used as the new base price with no further adjustment.

These methods are used for both local collection and central items.

Direct comparison

If another product is directly comparable to the old product, then it is assumed to have the same base price. For example, if a new garment is identical to the old garment in all respects but colour, then the new garment directly replaces the old garment and its base price remains the same.

This process obtains a replacement product that may be treated as essentially identical to the previous product. This means that any difference in price level between the new and the old product is entirely owing to price change and not quality differences. These products are marked as "comparable" using the "C" indicator code, or the "X" indicator code, if the comparable product is on sale.

Direct quality adjustment

This is the preferred method for when a replacement product is of a different quality or specification. An attempt is made to quantify the value of the quality or specification difference, and the base price is adjusted accordingly.

This section discusses quantity adjustment, which is the simplest form of direct adjustment. Another method of direct quality adjustment is hedonic regression, which is only applied for certain items (see our Special case aggregates in consumer prices article).

Quantity adjustment is used when an item's size permanently changes, for example, when the size of a confectionery bar changes. This occurs most frequently with homogenous goods, such as food and drink. When this happens, the nearest equivalent new size of the product priced in an outlet is found in each outlet and a pro rata adjustment is made to the base price for the change in weight.

Formula 1: How base prices are updated when a product's weight has changed

$$New\ base\ price = Current\ base\ price\ \times \frac{New\ weight}{Old\ weight}$$

For example, if the base price of a chocolate bar is 50p and the weight decreases from 85g to 80g, the new base price is 47p. This is calculated as:

$$New\ base\ price = 0.50\ \times \frac{80}{85}$$
$$New\ base\ price = 0.471$$

Instances where this has occurred are marked using the "W" indicator code.

Imputation of a new base price

If the replacement product is of a different quality or specification, and no information is available to quantify the difference, assumptions must be made. These products are marked as "non-comparable" using the "N" indicator code, or the "Z" indicator code if the non-comparable product is on sale.

A base price is calculated for the new product by assuming that its price change from the base month up until that month equals the average price change for products within the same stratum (see Section 12: Calculating elementary aggregate indices). For example, if the price is £14.99 and the index for that item (which is calculated excluding the product in question) in that stratum is 108.34, the new base price is £14.99, divided by 108.34 multiplied by 100, which equals £13.84.

This procedure ensures that bringing in the new product has no effect on the elementary aggregate for that item in the month that it is introduced.

Base price imputation is also used if an outlet closes, or refuses to allow further price collection. In this case, all products priced there are dropped, a new outlet is selected in the same location, and new base prices are imputed for products priced in the new outlet. In other cases, it is not possible to collect a valid base price for a product in January (for example, the product was not in stock), so a new base price is imputed for this product when the product next becomes available.

10 . Defining strata for local collection and central items

Stratification is the process of assigning quotes into lower-level "groups" or "strata" that can then be combined to form indices at elementary aggregate levels and higher, as defined in our Higher-level aggregation and weights in consumer prices article.

This initial stratification process assigns each quote to a group, which is made up of a combination of item, region, and retailer. For more information on the structure of UK consumer price indices, please see our Scope and coverage of consumer prices indices article.

For local collection items, the possible types of stratification are:

- stratification by item only
- stratification by item and region only
- stratification by item and shop type only
- stratification by item, region and shop type

For central items, the collection and processing methods allow for more flexibility in different stratification methods. In some cases, weights are available at the level of individual price quotes.

11 . Replication factors

For local collection items, some quotes are also assigned replication factors, otherwise known as "shop weights". These factors reflect the market share of chain shops and are used to appropriately represent price quotes from those shops with regional or national pricing strategies. They are not strictly weights; they are replication factors that give the number of times that each collected price should appear in each stratum.

The same replication factors are used in all measures of consumer price inflation. The market shares of the companies are calculated mainly from Living Costs and Food Survey (LCF) data, along with a variety of sources, such as market research reports. These are then broken down into individual replication factors for each item priced at that shop. Before the replication factors are estimated, the following factors are considered:

- the stratum weights
- the number of prices expected to be collected in each stratum cell
- the replication factors given to other regional or national chains

The replication factors for each company are broken down to regions, based upon the distribution of the company's shops.

Suppose that for item "X", which is stratified by shop type but not region, there is just one shop with a national pricing policy "Shopco", while all the other price data for this item are collected in store. Assume also that the following statistics relate to the collection of data for this item:

- item "X" is stratified by shop type (multiple versus independent shop types) only
- "Shopco" has 20% overall market share for item "X"
- on average, around 160 price observations are taken locally each month, of which 110 come from multiples and 50 from independent shops
- multiples in total have a 75% market share for item X

Then, the single-price observation from "Shopco" will be replicated 40 times in the multiple shop-type stratum cell. This means that of the 200 total price observations, 40 will be from "Shopco", giving it 20% of the market share. Overall, there will be 150 price observations in the multiple-shop stratum cell (110 locally plus 40 from "Shopco"), and 50 price observations in the independent-shop stratum cell (all collected locally). The two stratum indices can then be combined using stratum weights to produce an item index for item "X".

The formulas used to calculate the replication factors are:

Formula 2: How replication factors are calculated

$$R_t = \frac{\frac{M_t}{W} \times 100}{\left(100 - \left(\frac{M_t}{W} \times 100\right)\right)} \times L$$

$$R_s = R_t \times \frac{M_s}{M_t}$$

where:

- R_t equals total of all replication factors for that item
- R_s equals replication factor for shop "s"
- L equals expected number of prices to be locally collected for multiple shops for that item
- M_t equals market share for all shops for that item (as percentage)
- M_s equals market share for shop "s" for that item (as percentage)
- W equals shop-type stratum weight for multiple shops for that item (as percentage)

If the item is also stratified by region, then the replication will be split up so that the price is replicated within each region as well. The proportion of the replication factor put into each region depends on market information on total revenue by region for that shop. If this information is not available, the proportions are estimated by examining the total number of outlets for that shop in each region.

12 . Calculating elementary aggregate indices

Elementary aggregates are the set of indices calculated at the very first stage of aggregation. At this level, detailed expenditure weights are not available for the local collection data or for most central items. For example, the expenditure on "pink lady apples" bought in Cardiff from an independent shop is not known. Because this weighting information is missing, unweighted index formulas are used to construct these elementary aggregates. Different methods rely on different assumptions about the underlying expenditure shares, but in general the approach reflects the fact that the precise weights for individual products within a stratum are not available.

An elementary aggregate index can be constructed in different ways. The most commonly used unweighted index number formulas are:

- the Jevons Index, the geometric mean of price relatives
- the Dutot Index, the ratio of average prices
- the Carli Index, the arithmetic mean of price relatives

For the Consumer Prices Index (CPI)and Consumer Prices Index including owner occupiers' housing costs (CPIH), we use:

- the Jevons Index for local collection items
- a mixture of the Dutot and Jevons formulations for central items

For the Retail Prices Index (RPI), (see our Calculating the Retail Prices Index article), we use:

- a mixture of the Dutot and Carli formulations for local collection items and central items

Jevons Index

The Jevons Index can be expressed as follows:

Formula 3: How to calculate the Jevons Index

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

Where:

p_i^t is the price of product i at time t

This can be thought of as the geometric mean of the price relatives.

An alternative, and algebraically equivalent, way of thinking about this calculation is to express it as the ratio of the geometric average prices:

Formula 4: Alternative arrangement of the Jevons Index

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

Where:

p_i^t is the price of product i at time t

It is essential to use prices for matching comparable products. If, in any month, there is no price for an item corresponding to one in the base month, that price must be excluded from the calculations or a quality adjustment (for example, imputation of the base price) must be made.

Dutot Index

The Dutot Index can be expressed as follows:

Formula 5: How to calculate the Dutot Index

$$I_{Dutot}^{0,t} = \frac{\frac{1}{n}\sum_{i=1}^{n} p_i^t}{\frac{1}{n}\sum_{i=1}^{n} p_i^0}$$

Where:

p_i^t is the price of product i at time t

This is the ratio of the arithmetic means of the prices.

Carli Index

The Carli Index is only used in the RPI, where the Dutot Index cannot be used. It is used at the elementary aggregate level and can be expressed as follows:

Formula 6: How to calculate the Carli Index

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

Where:

p_i^t is the price of product i at time t

This is the arithmetic mean of the price relatives.

13 . Using elementary aggregate indices

In some circumstances, an upward bias is introduced to the overall price index when the Carli Index is combined with chain linking of within-year indices (see our Higher-level aggregation and weights in consumer prices article). The Jevons and Dutot formulas are not as susceptible to biases resulting from chaining. In the context of cross-country comparisons, they are also much less influenced by detailed differences in index and sample design in individual countries.

The International Labour Organisation (ILO) strongly encourages the use of the Jevons Price Index for calculating elementary indices, where weights are unavailable, as outlined in their Consumer Price Index Manual – Concepts and Methods. The ILO notes that the Dutot Price Index should only be used in cases where the sample of transactions is homogeneous with respect to base price levels or price trends. It also strongly discourages the use of the short-term Carli Price Index because of its known upward bias.

Eurostat regulations allow the use of the Jevons and Dutot Indices, but effectively rule out the use of the Carli Index, unless it can be demonstrated to behave in a similar way to those preferred formulas. Though the UK is no longer required to follow Eurostat regulations, the associated guidance reflects widely recognised good statistical practice. Applying these principles helps ensure that differences in inflation rates across countries arise from genuine differences in price movements, rather than from variations in the basic indexnumber formulas used to aggregate price data.

In a 2012 survey of then EU member states, we found 17 countries used the Jevons Index in their national consumer price index:

- Austria
- Bulgaria
- Croatia
- Cyprus
- Denmark
- Finland
- France
- Greece
- Ireland
- Italy
- Luxembourg
- Poland
- Portugal
- Romania
- Slovenia
- Spain
- Sweden

We found that 8 countries used the Dutot Index:

- Belgium
- Czech Republic
- Estonia
- Germany
- Latvia
- Lithuania
- Malta
- Slovakia

We found that three countries used a mixture of Jevons and Dutot:

- Hungary
- Netherlands
- UK

Outside of Europe, Australia, Canada, New Zealand and the US mainly used Jevons in the calculation of the national consumer price index, and Japan used Dutot.

Additional information can be found in the following papers and reports:

- Elliott, D., O'Neill, R., Ralph, J. and Sanderson, R. (2012) Stochastic and sampling approaches to the choice of elementary aggregate formula, Office for National Statistics
- Winton, J., O'Neill, R. and Elliott, D. (2013) Elementary Aggregate Indices and Lower Level Substitution Bias, Office for National Statistics
- Clews, G., Dobson-McKittrick, A. and Winton, J. (2014) Comparing class-level chain-drift for different elementary aggregate formulae using locally collected CPI data, Office for National Statistics.
- UK Consumer Price Statistics: A Review

14 . 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.

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.

Carli Index

In line with international best practice, we consider the use of Carli to be inappropriate. The Carli index is an unweighted index number formula, which is the arithmetic mean of price relatives.

Central items

About 150 items for which the prices are collected centrally, where the index calculation is carried out separately from the processing of the local collection data.

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.

Class

In the CPIH, the CPI and the HCIs, all categories of expenditure on which significant amounts of money are spent are arranged into 12 divisions, which are subdivided into groups and then into classes. Examples of classes are bread and cereals, water supply, and transport insurance.

We publish price indices for each class.

Coverage

Those transactions that can be identified and measured in practice. This is determined by the expenditure categories for which weights are compiled.

Division

In the CPIH, CPI and HCIs, all categories of expenditure on which significant amounts of money are spent are arranged into 12 divisions, such as clothing and footwear, transport, and recreation and culture.

We publish price indices for each division.

Dutot Index

An unweighted index number formula, which is the ratio of average prices.

Elementary aggregates

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

Enumeration

Detailed listing of all outlets in a location, giving address, size, outlet type, and range of products sold.

Group

In the CPIH and CPI, all categories of expenditure on which significant amounts of money are spent are arranged into 12 divisions, which are subdivided into groups. Examples of groups are food, postal services and insurance.

In the Retail Prices Index (RPI), all categories of expenditure on which significant amounts of money are spent are arranged into 14 groups, such as food, housing and motoring costs.

We publish price indices for each CPIH/CPI and RPI group.

Index day

The specific Tuesday within index week when price collectors collect the majority of prices for traditional data sources. This maximises consistency in monthtomonth comparisons.

Index week

The designated week, at or near the middle of each month, during which local price collectors gather prices. Index day falls within Index week.

Indicator codes

Codes entered by price collectors into the handheld device if there are any special features in the prices recorded. For example, collectors enter an "S" if the item is on sale or special offer.

Inter-Departmental Business Register

The Inter-Departmental Business Register (IDBR) is a comprehensive list of UK businesses used by government for statistical purposes. The IDBR provides retailer outlet counts, which are used to determine whether outlet types are independents or multiples.

Items

Any type of consumer good or service that can be purchased, for example, apples. Several different varieties of that item may be available, for example, Granny Smith and Braeburn apples.

Local collection items

Prices for items collected either in-store, online or by phone by price collectors in various locations across the country. These are supplemented by prices collected by head office staff for shops with national pricing policies, and regional services collection prices collected by phone, email or online for local services, such as plumbers, taxi hire and hotels.

Locations

Locations are clusters of enumeration districts that broadly represent a central shopping area. Out-of-town shopping centres are also included.

Jevons Index

An unweighted index number formula, which is the geometric mean of price relatives.

Outlets

An outlet is anywhere from which goods or services can be purchased. For most items, it is usually a shop or market stall. However, for some items, outlets include restaurants, pubs, solicitors' offices, or a sole trader operating from home.

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.

Regional services collection

Items collected under the regional services collection are generally not purchased in retail centres, for example electricians' charges, childminder fees and driving lesson fees. This is different to the other local collection items. These prices are collected directly from the relevant businesses or organisations, either by phone, text, email, or from information published on their websites.

Replication factors

Replication factors (or "shop weights") reflect the market share of large chain shops. They are used to appropriately represent price quotes from those shops with regional or national pricing strategies.

Representative items

Representative items are those items that are in the basket of goods and services.

Sampling frame

A complete list of the objects to be sampled, together with sufficient information on each object to stratify, if required.

Scope

All transactions that one would ideally want to measure.

Shops with national pricing policies

Major chains of shops with national pricing policies (also known as "central shops"). Branches of these chains are excluded from outlet enumeration because their prices are collected by head office staff. Prices collected are given extra weight to reflect their market share (see the Glossary entry for Replication factors).

Shops with regional pricing policies

Chains of shops with regional pricing policies (also known as "regional central shops"). Collectors visit only a few of their outlets and assume that each outlet reflects their pricing policy within a given region. Prices collected are given extra weight to reflect their market share (see the Glossary entry for Replication factors).

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.

Tukey algorithm

The Tukey algorithm identifies and invalidates price movements that differ significantly from the norm.

15 . Related links

Consumer prices indices technical guidance Methodology | 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.

16 . Cite this methodology article

Office for National Statistics (ONS), released 25 March 2026, ONS website, article, Traditional data aggregates in consumer prices