

Article

Introducing alternative data into consumer price statistics: aggregation and weights

New aggregation structure and weights methods, combining alternative and traditional data to measure consumer price inflation statistics.

Contact:
Consumer Prices Methods
Transformation
cpi@ons.gov.uk

Release date:
29 April 2025

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [New aggregation structure](#)
3. [Weight sources and design](#)
4. [Case studies of aggregation structures](#)
5. [Future developments](#)
6. [Related links](#)
7. [Cite this article](#)

1 . Overview

The Office for National Statistics (ONS) introduced a new aggregation hierarchy containing a "consumption segment" level from 2025, which will allow inflation to be measured from a combined use of traditional and alternative data. This article updates the [previous release](#), published on 9 November 2021, to reflect the latest position on aggregation of traditional and alternative data. The important principles behind these changes are still valid.

Consumer price inflation statistics have been going through a series of transformations to improve the way inflation is measured, which includes introducing new data sources. These new data sources have advantages compared with traditional sources, allowing us to calculate more accurate representative prices, use product-level weights, and vastly increase our product coverage. In this article, we will discuss how we will aggregate these new data sources along with our traditionally collected data sources, to ensure both are represented within our inflation statistics.

2 . New aggregation structure

When calculating inflation statistics, we use prices of a sample of similar products to measure the price change of an "elementary aggregate". We then use a weighted arithmetic average of a group of elementary aggregates to combine them into a higher-level aggregate. This higher-level aggregate can in turn be combined with others, into a higher-level aggregate one further level higher. This process is repeated and creates a pyramid structure, where aggregates in a lower level can be aggregated into an aggregate one level higher. We call this pyramid structure our "aggregation hierarchy".

As described in our [Impact analysis on transformation of UK consumer price statistics: January 2025 article](#), we introduced a new aggregation structure into our inflation statistics in 2025. The new structure gives us the flexibility to use traditional data sources, alternative data sources, or a dual collection of both types of data, for different areas of the basket of goods and services.

Our new aggregation structure will feature the following levels:

- five levels of Classification of Individual Consumption According to Purpose (COICOP) (Consumer Prices Index (CPI), Consumer Prices Index including owner occupiers' housing (CPIH)), or four levels of the Retail Price Index (RPI) structure
- one consumption segment level
- one region level
- one retailer (type) level
- for alternative data; optionally, one or more extra strata levels
- for traditional data; one item level

In Case study A: Groceries (rice) in [Section 4: Case studies of aggregation structures](#), we show several case studies for different areas of the basket of goods and services to demonstrate the flexibility we have with this aggregation structure. As shown in the groceries case study, this change will be particularly helpful for introducing grocery scanner data into consumer price statistics from 2026.

Classification of Individual Consumption According to Purpose

Classification of Individual Consumption According to Purpose (COICOP) is an international standard used for the five highest levels of the aggregation structure in the UK CPI, CPIH and Household Costs Indices (HCIs). In order, these levels are described as headline, division, group, class and subclass.

RPI does not use COICOP, but rather a bespoke UK aggregation structure featuring four levels comprising headline, broad groups, groups and sections. Both COICOP and the RPI higher-level aggregation structure are discussed in more detail in sections 3.2 and 12.2 of the [Consumer Prices Indices Technical Manual, 2019 methodology](#), respectively.

COICOP has been revised in [Classification of Individual Consumption According to Purpose \(COICOP\) 2018 by the Department of Economic and Social Affairs \(PDF, 2.46MB\)](#). Our new aggregation structure has been designed to allow for the introduction of this revised COICOP 2018; for example, by aligning our consumption segments with the high-detail structure for food products where possible.

Consumption segment

Consumption segments are a UK-defined extra level of granularity lower than COICOP, introduced in 2025. For areas where we are introducing alternative data sources (particularly groceries), consumption segments are defined to be broader than items but still relatively homogeneous. An example may be "rice", which encompasses various types of dry rice, microwavable rice, and rice snacks such as rice cakes. We can then maximise use of alternative data by incorporating a near-census of all rice variants within lower-level aggregates. By contrast, for traditional data, a consumption segment will either match a single item or be broader than and represented by one or more items.

Region

Region-based stratification is possible in our new aggregation structure. Note that there is some flexibility in this. Some areas of the basket of goods and services may be stratified by region, with all 12 regions represented in this level of the structure. Other areas of the basket may consist only of prices collected centrally, without geography specified. In this case, we may have a singular "UK" stratum at the region level. An example of this is shown in Case study B: Second-hand cars (petrol cars) of [Section 4: Case studies of aggregation structures](#).

Retailer (type)

The retailer (type) level is flexible depending on the data sources used for the area of the basket of goods and services being represented.

For alternative groceries data, the retailer level represents one of our data suppliers from which we are constructing indices. Sometimes this is a single retailer, as is the case in our groceries scanner data. In other cases it may be that a single data supplier collates several retailers, such as our second-hand cars data, where listings come from a variety of sellers.

For traditional data, where we stratify by retailer, we group retailers into "multiples" (retailers with 10 or more physical outlets) and "independents" (fewer than 10 physical outlets). Some manual adjustments are made to account for online retailers with high market shares. If we don't stratify by retailer, all quotes are combined into a single "other retailer" category.

We avoid double counting in areas where we have a dual collection of scanner and traditional data. For example, if a retailer is represented by scanner data, then it will not be reflected in its corresponding "multiple" grouping in traditional data, either in terms of price quotes, or aggregate weight.

Alternative data sources

For alternative data sources, we may optionally break the retailer level down by one or more levels of "extra strata". This allows us to stratify alternative data down in category-specific ways. For example, we currently break second-hand cars down by two additional levels reflecting the age and make of the vehicle. These are specific to second-hand cars and are not used for other data sources. This further breakdown can be used to improve the homogeneity of our aggregates and increase the granularity of our inflation statistics.

Traditional data sources

For traditional data sources, we will break the retailer level down by one or more representative items. Items are similar to how we have used them in the past. We will continue to select around 700 items to represent the basket of goods and services in our [Consumer price inflation basket of goods and services: 2025 article](#), which is updated annually.

We will continue to use a sample of prices to measure price change for each item. The main difference is that items were previously higher than the region and retailer levels in the aggregation hierarchy, but now form the lowest level. This means we will often not have a single item index, but rather many regional- and retailer-stratified indices. Consumption segments will now form the higher-level aggregated index.

3 . Weight sources and design

Item, consumption segment and COICOP weights

The items in the consumer price inflation basket of goods and services are weighted based on the ratio of total expenditure of each item to all goods and services expenditure in the UK. Item weights indicate the relative importance of each item and are measured in parts per 1,000. Item weights are discussed in more detail in section 8.4 of the [Consumer Prices Indices Technical Manual, 2019 methodology](#).

The weights at Classification of Individual Consumption According to Purpose (COICOP), or equivalently, the Retail Price Index (RPI) structure, and consumption segment levels are calculated as the sum of the weights of all the items that underpin the stratum. For example, if the "rice" consumption segment contains the items "microwavable rice", with a weight of 0.7, and "basmati rice", with a weight of 0.5, then the rice consumption segment would be given a weight of 1.2. This weight is then normalised when aggregating with other consumption segments, to ensure all weights sum to 1 at each level of aggregation. Since item weights sum to 1,000, this means that the weights at each of these levels also sum to 1,000.

Note that if a consumption segment does not contain any underlying items, then its weight will be 0 and therefore will not contribute to our inflation measures – it will be represented by other consumption segments in the same section. This ensures proportionate use of the alternative data source retailers, and that items continue to represent the basket in accordance with their weights.

Retailer weights

Retailer type weights are calculated from the Annual Business Survey as the proportion of expenditure which that retailer (or group of retailers) covers within the COICOP "class" level. These proportions are then used as weights for any consumption segment within the class. If no retailer stratification occurs (for example, where large retailers dominate the market), then there is only a single stratum with all the weight.

When we are using a dual collection of scanner and traditional data, the weight of a scanner retailer will only be counted once. For example, if multiples collectively have an 80% market share, and retailer X is a multiple with a 15% market share for which we have scanner data, then the retailer X stratum would have a weight of 15% and the multiples stratum would have its weight reduced to 65%.

Region weights

Regional weights are calculated by apportioning expenditure measured in the Living Cost and Food Survey (LCF) by region for broad product groups. Since the RPI relies on LCF data for its source of expenditure weights, we use RPI sections to define these broad product groups. For Consumer Prices Index (CPI) and Consumer Prices Index including owner occupiers' housing (CPIH), these weights are then mapped onto the CPI aggregation structure at consumption segment level. For items stratified by region, their item-level regional weights then use their corresponding parent consumption segment weights. Items that are not stratified by region are assigned a singular region with a weight of 1.

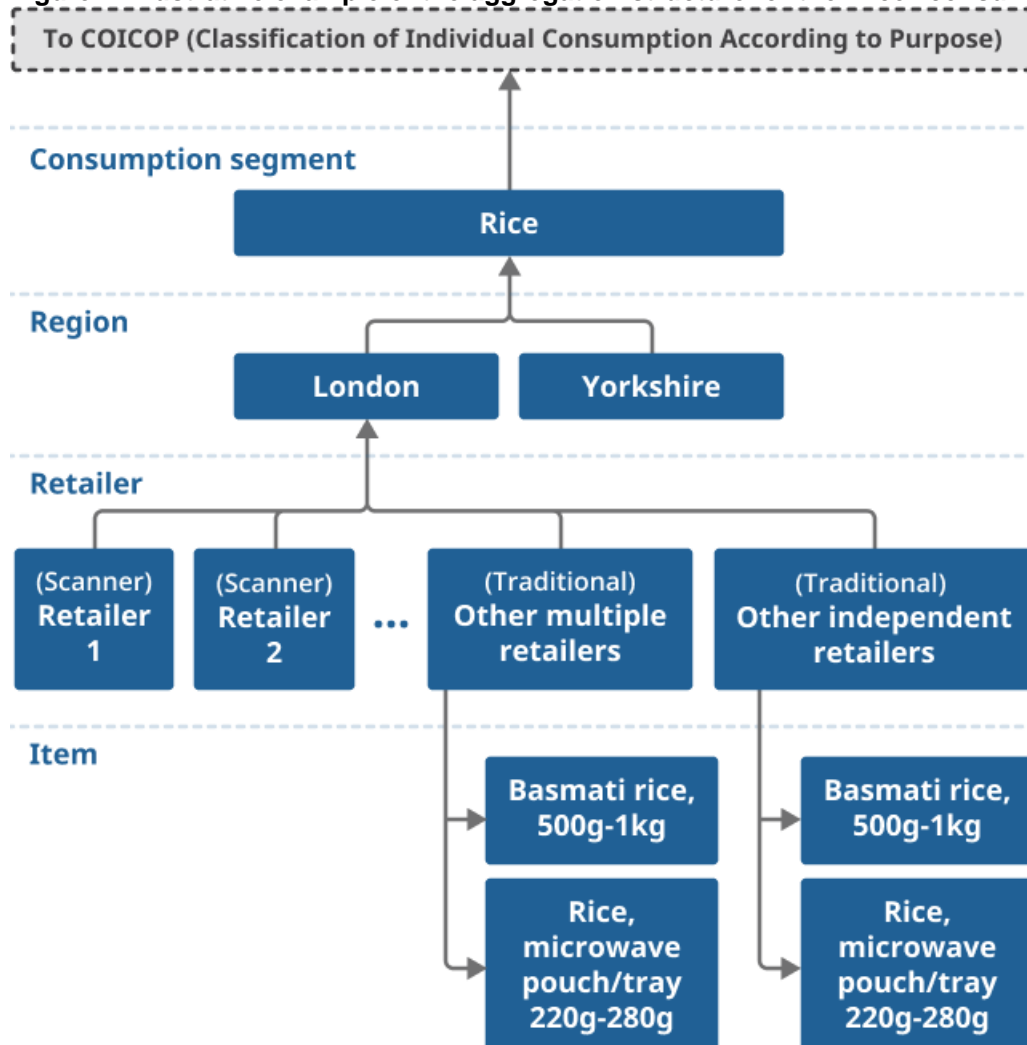
As explained in Section 3: Introduction of consumption segments of our previous article, [Introducing alternative data into consumer price statistics: aggregation and weights](#), we defined consumption segments as relatively homogeneous sets of products. This means that consumption segments can only contain items based on the same RPI sections and will therefore have the same regional weights as their underlying items. Additionally, consumption segments that do not contain regionally stratified items will be assigned a singular region with a weight of 1.

4 . Case studies of aggregation structures

Case study A: Groceries (rice)

Groceries will be dual collected, with both alternative scanner data and traditional locally collected data used to represent groceries. An example of the aggregation structure for the rice consumption segment can be found in Figure 1.

Figure 1: Illustrative example of the aggregation structure for the “rice” consumption segment



Source: Office for National Statistics

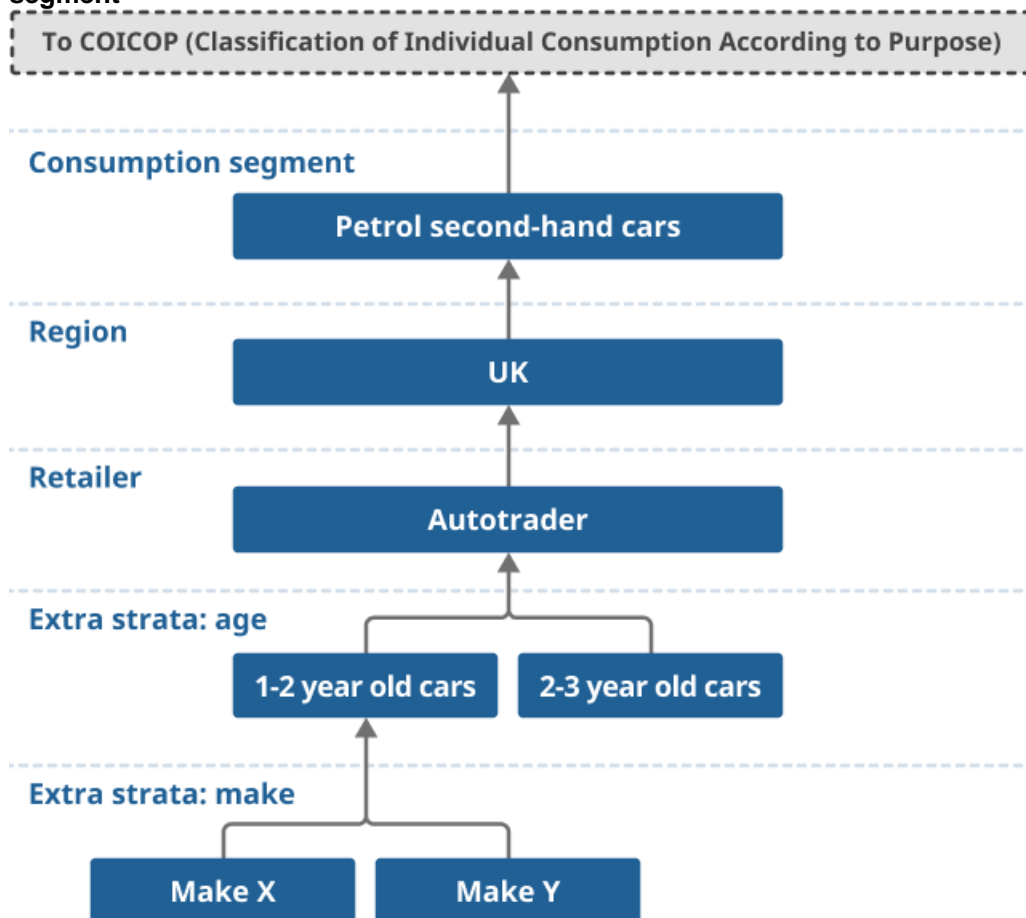
Many areas of the basket of goods and services for groceries, rice included, are regionally stratified and therefore rice is broken down into 12 regional aggregates. At the retailer-level, we have scanner data for individual grocery retailers, which each form their own retailer aggregate, along with strata for other multiple and independent retailers (whose product samples and weights will not contain the retailers represented by scanner data). For alternative data sources, we have no further extra strata, but for traditional data sources, we break the retailer level down into two items: basmati rice and microwave rice.

Note then that the scanner data elementary aggregates are broken down by region and retailer but are not broken down any further. We use every rice product available to us when calculating the elementary aggregate, which may include risotto, basmati, microwave and other forms of rice. By contrast, the traditional data elementary aggregates only consist of two forms of rice (basmati and microwave), and we use a sample of products to represent these items. This example shows how we can combine a sampling-based approach for traditional aggregates, along with a census-based approach for scanner aggregates, into a single aggregation structure.

Case study B: Second-hand cars (petrol cars)

The aggregation structure for the second-hand cars consumption segment can be found in Figure 2. Note that while groceries used data from both alternative and traditional sources, for second-hand cars we will use only alternative data sources, supplied by Autotrader.

Figure 2. Illustrative example of the aggregation structure for the “petrol second-hand cars” consumption segment



Source: Office for National Statistics

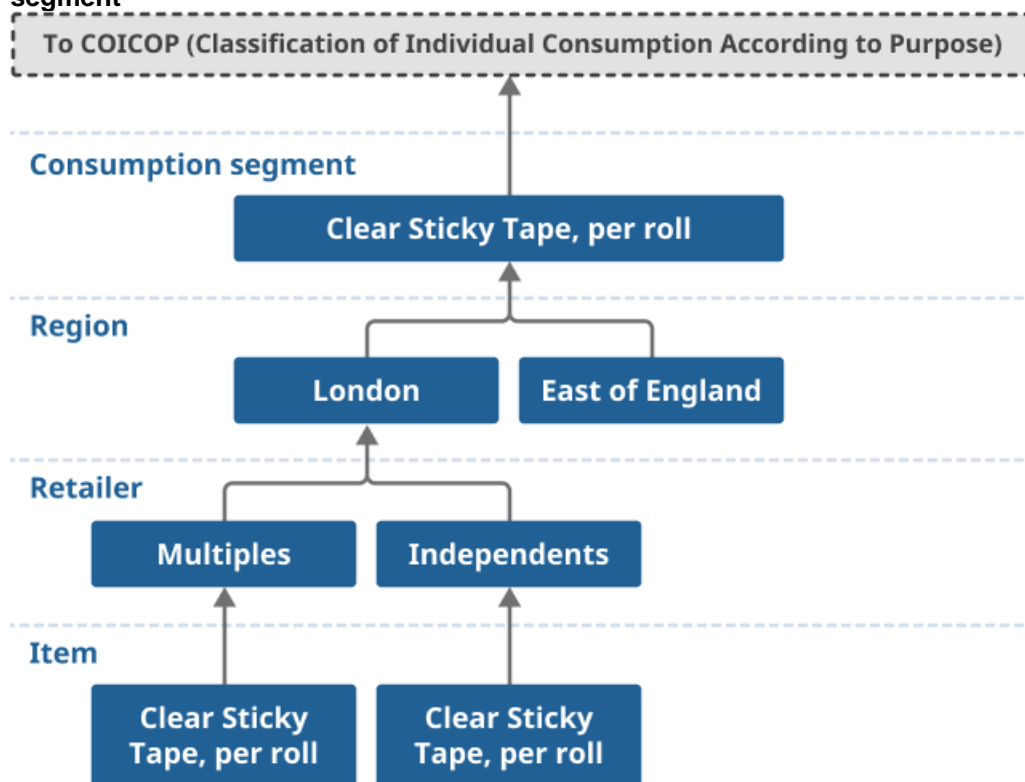
Figure 2 shows the petrol second-hand car consumption segment. These second-hand car listings are available online and can be bought from most regions in the UK. Therefore, it makes sense for second-hand cars not to be stratified by region. Because of this, we use a single UK region at the region level to account for the fact we will not perform a regional split for second-hand cars.

We have one data supplier covering many retailers, so also use a single stratum at the retailer level to encompass the data supplier. Finally, for second-hand cars we use two levels of extra strata that are bespoke to second-hand cars alternative data sources: seven age-bandings encompassing the age of the vehicle, along with approximately 25 makes of car.

Case study C: Clear sticky tape

The aggregation structure of clear sticky tape at the consumption segment can be found in Figure 3.

Figure 3: Illustrative example of the aggregation structure for the “clear sticky tape” consumption segment



Source: Office for National Statistics

Clear sticky tape does not currently use alternative data and is represented only by traditional data. Clear sticky tape is stratified by both region and retailer. Since there is no alternative data, the item matches the consumption segment. An example item elementary aggregate may be "clear sticky tapes sold in London within multiple stores".

5 . Future developments

In our [previous aggregation and weights article](#), we discussed potentially introducing "big-small" retailer stratification as an alternative to "multiple-independent" stratification, which would mean defining and grouping retailers based on their market share rather than the number of physical stores that the retailers own. There are still some challenges within the data for us to overcome, and so we are assessing this option for potential introduction in 2026.

6 . Related links

[Consumer price inflation basket of goods and services: 2025](#)

Article | Released 18 March 2025

The "shopping baskets" of items used in compiling measures of consumer price inflation, updated annually to ensure the measures are representative of consumer spending patterns.

[Impact analysis on transformation of UK consumer price statistics: January 2025](#)

Article | Released 23 January 2025

Indicative impacts of the planned improvements to our consumer price statistics from January 2019 to June 2024.

[Consumer Prices Indices Technical Manual, 2019](#)

Methodology | Released 26 March 2025

Explanation of how measures of consumer price inflation and associated indices are compiled.

7 . Cite this article

Office for National Statistics (ONS), released 29 April 2025, ONS website, article, [Introducing alternative data into consumer price statistics: aggregation and weights](#)