Consumer/Retail Prices Indices

Microdata

10th April, 2017

Edition: *Two*

Reference period: February 1996 – February 2017

Office for National Statistics

Table of Contents

Basic Information

Title
Consumer, and Retail Prices Indices micro data
Previous titles
None
Topics Covered/ key words
Quote prices and associated meta data and weights underpinning the monthly production of the Consumer Prices Index (CPI), Consumer Prices Index including Owner Occupied Housing costs (CPIH) and Retail Prices Index (RPI).
Time Covered
January 1996 to August 2016
Data Source
Price quote data, item indices and associated meta data used in the construction of the UK Consumer Prices Index and its variant (Consumer Prices Index, including Owner Occupied Housing) and the Retail Prices Index.
Geographic Coverage
Economic territory of the UK (E.g. England, Wales, Scotland and Northern Ireland), excluding offshore islands (the Channel Islands and the Isle of Man).
Lowest level of Geography
Region
Breakdowns
Classification of Individual Consumption by Purpose (COICOP) and RPI classification system (comprising of broad groups and sections).
Sample Size
Approximately 200,000 prices. There are currently over 720 representative items in the CPI, CPIH and RPI basket of goods.
Frequency of Release
Monthly
Revision Policy
The Retail Prices Index is not revisable. The Consumer Prices Index is revisable, but causes for revision have been rare.
Next Publication
See section 1.7 below.
Data Owner and Supplier
Rowan Kelsoe, Prices Division, ONS Newport.

1 Introduction to CPI, CPIH and RPI

1.1 Aim of survey

Consumer price inflation measures the change in price of a fixed basket of goods and services consumed by households and is an important indicator of how the UK economy is performing. The CPI is currently the inflation measure used in the Government's target for inflation and, since 2011, it has been used for deflating consumer spending within the National Accounts.

1.2 Background of survey

Consumer Prices Index;

- Governed by European regulation, the CPI was first published in 1997 as the Harmonised Index of Consumer Prices (HICP)
- Between 2000 and 2002, the HICP's coverage of goods and services was extended to cover health, education, childcare and insurance.
- In 2003, the HICP was renamed the Consumer Prices Index (CPI) to reflect its new role as the main UK domestic measure of inflation.
- The CPI continues to be produced to the required international standards and in line with European (HICP) regulations.
- Since 2011 the CPI has replaced the RPI for deflating consumer spending within the National Accounts.
- In 2016, the CPI and CPIH were re-referenced from 2005 equals 100 to 2015 equals 100.

Consumer Prices Index (including Owner Occupied Housing Costs), CPIH;

- CPIH was introduced in 2013 and uses the same underlying principles in its construction as that of the CPI, but additionally includes owner occupiers' housing costs.
- Owner occupier's housing costs are measured using the rental equivalence approach. For more information see the CPIH Compendium.
- Data are available from 2005
- The CPIH is not currently an official statistic, but is undergoing a monitoring period for it to obtain National Statistics accreditation; we hope to achieve accreditation during the first quarter of 2017.

Other measures of consumer prices inflation include;

- The Retail Prices Index (RPI) is a long standing measure of UK inflation that has been used for a wide range of purposes, such as the indexation of pensions, rents and index-linked gilts. The RPI has recently been de-designated as a National Statistic as

it uses an unsuitable formula at the lowest level of aggregation. Currently, only routine changes are made to the RPI index. Data are available from 1987.

1.3 Main themes / key words

A copy of the Consumer Prices Indices Technical Manual (2014 Edition) and CPIH Compendium, which are definitive guides to the compilation of consumer price indices within the UK, can be found as companions to this data release. A glossary of terms, concepts and abbreviations can be found at Annex A below.

1.4 Relevance

Consumer price indices are important indicators of how the UK economy is performing, the main users and uses of consumer price statistics are as follows:

- **Monetary Policy Committee** (Bank of England) - the CPI is mainly used as the Government's target for inflation. Currently set at two percentage points, if inflation is more than one percentage point higher or lower than the target, then the Governor of the Bank of England is required to write an open letter explaining what measures the MPC are taking to bring inflation back to its target.
- **Public -** to show the impact of inflation on family budgets and how income and expenditure, when linked to inflation, can affect interest rates, taxes (tax allowances, personal tax, business tax and indirect tax), wage negotiations, savings (index linked Government bonds and National Savings), indexation of pensions and state benefits, maintenance contracts, business contracts, regulated charges (such as rail fares), and student loans.
- **Deflation** - to remove the effect of price change in the UK National Accounts. For example, from 2010, ONS switched from using the RPI to using the CPI to revalue gross domestic product (GDP) at constant prices.
- **Academia** the academic community uses Consumer Price statistics for research and analysis purposes.
- **The media** – there is significant media coverage of the current Consumer Prices releases which are often reported on in comparison to other statistics to help the public understand how their living standards are changing over time.

1.5 Longitudinal

Prices are collected for a set basket of goods for at least one year.

1.6 Geography

For the local price collection, the UK is divided into regions with a number of locations selected in each region. The lowest aggregate of prices (elementary aggregate) covers all prices collected for one item in one stratum.

1.7 Status of the data in the VML

Datasets currently available are lower level price quote data linked to associated meta data (item descriptions, shop and stratum weights) as well as item indices (and associated item weights) and higher level COICOP weights and COICOP and RPI mappings between January 1996 and August 2016.

Table summarising the contents of the Consumer Prices VML datasets;

No	File names	Summary of contents
1	*price_quote_yyyy_qn or price_quote_yyyymm*	Detailed monthly price quotes that have underpinned the production of published consumer prices indices. These values are used to create elementary aggregates.*
2	*locationTable.csv*	List of locations and the region in which they reside. This file should be used in conjunction with the price quote file at 1.
3	*stratum_weights_yyyy*	A separate file containing yearly stratum weights from 1996 to 2006. Use this file in conjunction with 1 above to aggregate stratum (elementary aggregate) level indices to item level indices.* Note: From 2007 these values are included within the individual price quote files (1 above).
4	*item_indices_yyyy_qn* or *item_indices_yyyy_mm*	Pre-calculated item level indices and corresponding item level weights for CPI, CPIH and RPI
5	*cpi_coicop_map* and *cpih_coicop_map*	Item to COICOP4 level mappings for CPI and CPIH. Use this file in conjunction with 1 to 3 above and 6 below to derive COICOP4 and parent level aggregations (COICOP 3:1).
6	*cpi_coicop_descriptions* and *cpih_coicop_descriptions*	COICOP 4:2 level descriptions. Use this file in conjunction with 4 above.
7	*cpi_coicop_weights_yyyy* and *cpih_coicop_weights_yyyy*	COICOP4:2 level weights for CPI and CPIH. Use this file in conjunction with 1 to 4 above to construct COICOP3:2 level parent aggregations.
8	*rpi_sections*	RPI Class and Group level descriptions. Note, item level indices are mapped to RPI classes and groups using the 4th and 3rd foremost digit of item ID. Use in conjunction with 1 to 3 above.
9	*Consumer Prices Technical Manual – 2014 – v5*	A definitive guide to the computation of UK Consumer Prices Indices (CPI, CPIH and RPI).
10	*CPIH Compendium*	A definitive guide to the computation of Owner Occupied Housing and its use within the CPIH.
*Note - the same values are used for CPI, CPIH and RPI construction		

The CPI is published each month on the Office for National Statistics website, www.ons.gov.uk, in a Statistical Bulletin with accompanying detailed Briefing Notes, on a Tuesday about a month after its index date. Low level micro data are published on a monthly basis directly after the publication of the current month's index. Publication dates are announced six months in advance and the latest available dates for 2016/2017 are provided in the table below.

Index Month	Publication Date
October 2016	15 November 2016
November 2016	13 December 2016
December 2016	17 January 2017
January 2017	14 February 2017*
February 2017	21 March 2017*
March 2017	11 April 2017*
April 2017	16 May 2017*
May 2017	13 June 2017*
June 2017	18 July 2017*
July 2017	15 August 2017*
August 2017	12 September 2017*
September 2017	17 October 2017*
October 2017	14 November 2017*
November 2017	12 December 2017*
December 2017	16 January 2018*

*Provisional dates

1.8 Other important points to note

Consumer Prices microdata is produced to include a regional variable which approximates to Government Office Regions. While adequate as inputs to the calculation of higher level consumer prices indices, source data underpinning the calculation of stratum level indices at this level or below are generally subject to greater uncertainties and/or are of relatively poorer quality due to smaller sample sizes.

2 Sample Design

The scope and coverage of CPI (CPIH) and RPI are those goods and services which are included in the household final monetary consumption expenditure (HFMCE) component of the National Accounts, which is consistent with the HICP version of the international classification framework (COICOP).

Approximately 200,000 prices are collected in line with the COICOP/HICP classification system, from a selection of items which are representative of UK consumer expenditure; prices are only collected for those items selected. For example, for the item home-killed lamb, prices are collected for 'loin chops with bone' and 'shoulder with bone'. Other joints, and loin chops and shoulders without bones, are not priced; it is assumed that their price movements are close to those of the joints of lamb that are priced. There are currently over 720 representative items in the CPI/CPIH prices basket of goods.

2.1 Sample frames and sample selection

Full details of the CPI sampling framework and methodology can be found in **Chapter 4** of the CPI Technical Manual and within the CPIH Compendium.

2.2 Stratification and sample design

At the elementary aggregates stage, CPI, CPIH and RPI are stratified by shop type, by region or by region and shop type. The stratum types and cells are as follows;

Stratum Type	**Stratum Cell**
Not stratified	0
By shop	1 = Multiple (10 or more outlets) 2= Independent (fewer than 10 outlets)
By region	1= Catalogue collections 2= London 3=SE 4=SW 5=East Anglia 6=East Midlands 7=West Midlands 8=York's & Humber 9=NW 10 = North 11 = Wales 12 = Scotland 13 = NI
By region and shop type	If the shop type is a multiple, then the stratum cell is equal to the regional stratum code. If the shop type is an independent, then the stratum cell is equal to the regional stratum code +13. For example, if an item is stratified by region and shop and a shop at which that item is placed has fewer than 10 outlets (shop_type = 2) and is in London (region 2) then its stratum cell is equal to 2+13 =

	15, whereas, if the shop is a multiple then its stratum cell is equal to its region code of 2.

3 Estimation

3.1 Using weights

Elementary Aggregates

The individual raw price quote observations supplied to the VML are un-weighted, but shop weights (replication factors) are provided to obtain elementary aggregate (stratum level indices) results where needed.

The same shop weights are used in the compilation of CPI, CPIH and for RPI. However, different formula methods are used for computation of RPI stratum level indices.

For local and regional central shop collections, a Jevons (geometric mean) formulation is always used for both CPI and CPIH stratum index computation, whereas central spreadsheet items uses a mixture of the Dutot (Ratio of Averages) and Jevons formulation. Either a Dutot or Carli (Average Ratio) formulation is used for the computation of the RPI.

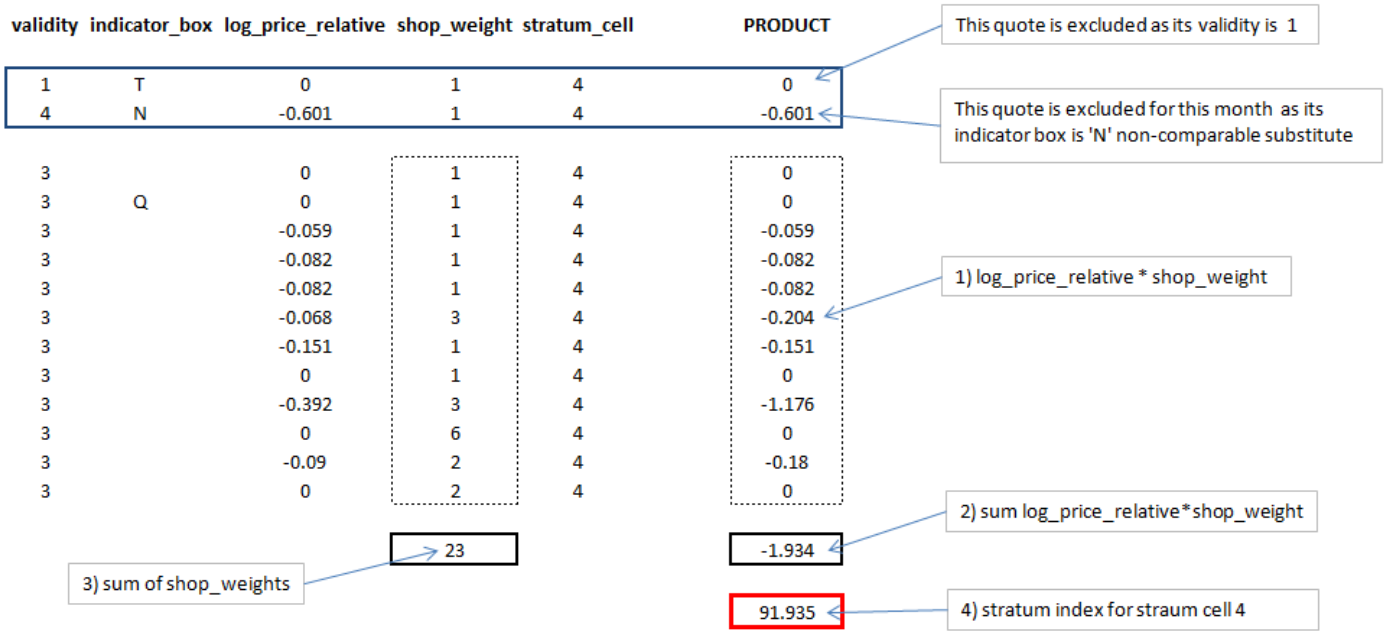
The formula method used for the RPI can be determined using the index_algortihm flag located in the *Item_Indices* file where an index algorithm flag of 1 equals Dutot and 2 equals Carli formulation.

A detailed guide to constructing a CPI, CPIH and RPI index can be found in **chapter 2, 3** and **10** respectively of the CPI Technical Manual and a guide to the formulation of weights for all index types can be found in **chapter 7.**

For example, the VML data set (*price_quote_201608)* can be used to calculate an elementary aggregate (CPI) for Home Killed Lean Mince (item 210403), stratum 4, for August 2016 as follows:

1) any non-contributing quotes are removed;
   - remove any invalid quotes (where 'validity' status is not in 3, 4, 53 or 54)
   - remove any quotes that have an indicator_box equal to 'N'. (N codes are non-comparable substitutions and do not contribute to the index during their first month of introduction).
2) the remaining observations are sorted and grouped by their stratum_cell value

3) the log_price_relative (log of price divided by base_price) is multiplied by the shop_weight [1] and then summed [2]
4) the sum of shop_weight is taken [3]
5) the stratum index is calculated as the exponent of the sum of all the products [2] divided by the total shop_weight [3] multiplied by 100. (Note, this is, in essence, one of many ways to calculate a geometric mean).



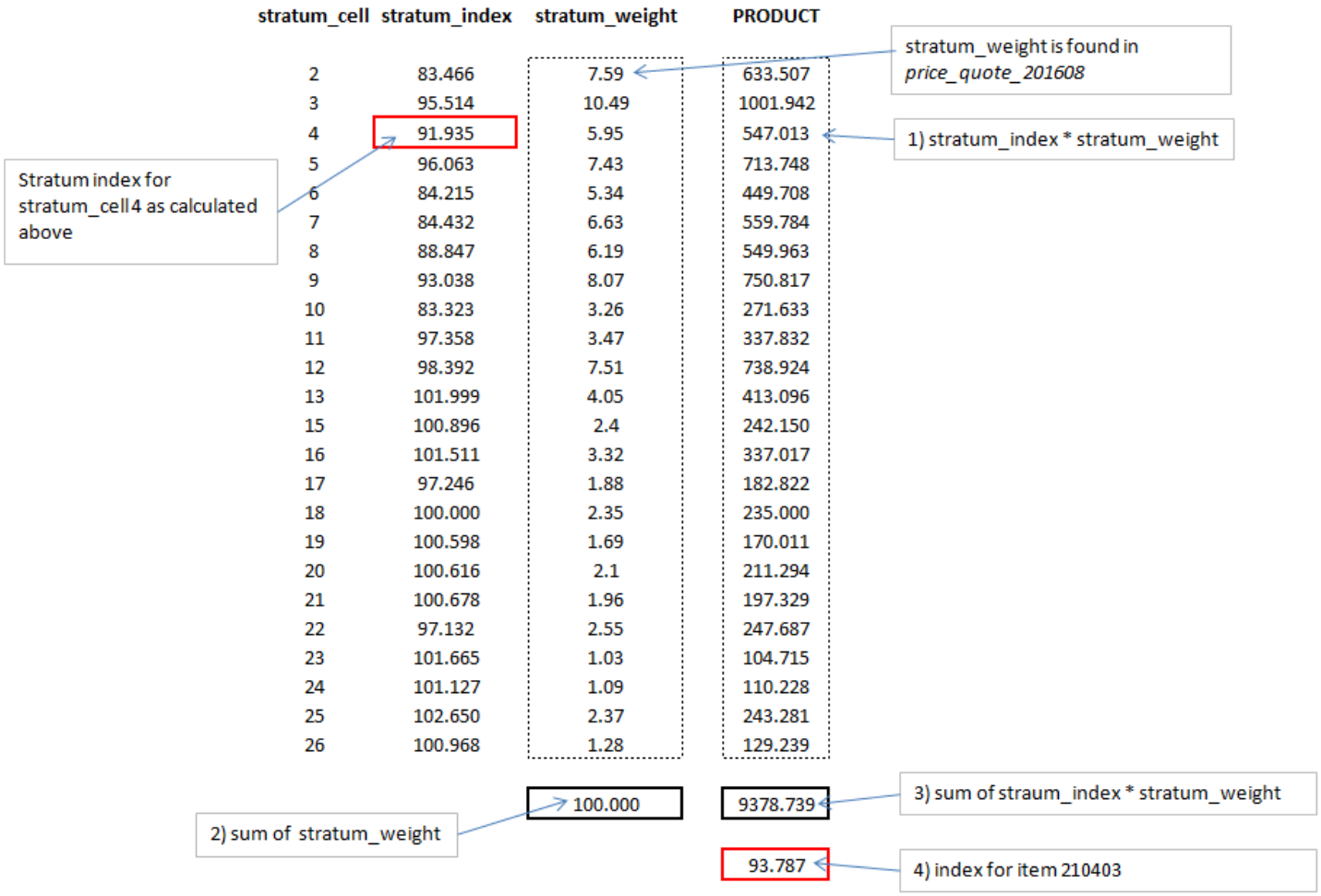
Steps 3 to 5 are repeated for the remaining stratum cells for the item.

Item indices

Item indices are weighted for CPI, CPIH and RPI following the elementary aggregates stage.

For example, an item index (CPI or CPIH) for Home Killed Lean Mince (item 210403) for August 2016 can be calculated using elementary aggregates (stratum indices) calculated above.

1) the sum of stratum_weight [2] is taken
2) the stratum_index is multiplied by the stratum_weight [1] and then summed [3]
3) the item index [4] is calculated as the sum of products [3] divided by the total stratum_weight [2]



stratum_cell	stratum_index	stratum_weight	PRODUCT
2	83.466	7.59	633.507
3	95.514	10.49	1001.942
4	91.935	5.95	547.013
5	96.063	7.43	713.748
6	84.215	5.34	449.708
7	84.432	6.63	559.784
8	88.847	6.19	549.963
9	93.038	8.07	750.817
10	83.323	3.26	271.633
11	97.358	3.47	337.832
12	98.392	7.51	738.924
13	101.999	4.05	413.096
15	100.896	2.4	242.150
16	101.511	3.32	337.017
17	97.246	1.88	182.822
18	100.000	2.35	235.000
19	100.598	1.69	170.011
20	100.616	2.1	211.294
21	100.678	1.96	197.329
22	97.132	2.55	247.687
23	101.665	1.03	104.715
24	101.127	1.09	110.228
25	102.650	2.37	243.281
26	100.968	1.28	129.239
		100.000	9378.739
			93.787

Step 1 to 3 is repeated for all other items.

<u>Above item aggregation;</u>

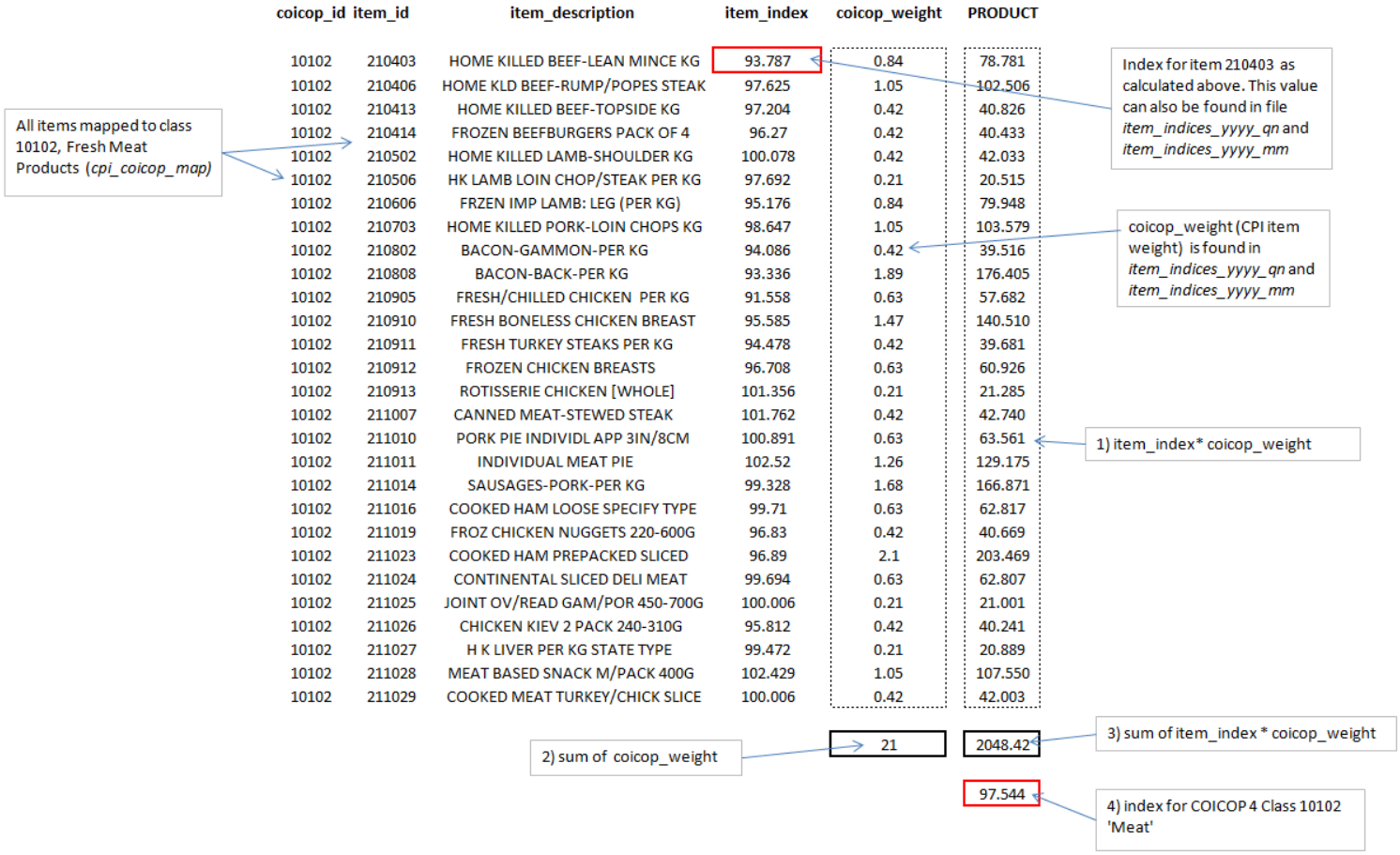
An additional set of item and COICOP class level weights, along with required mappings, are provided to enable item aggregation to the standard international COICOP classification system used by the CPI and CPIH. A pared-down version of the COICOP structure can be found at Annex B below.

The RPI uses an aggregation structure that is unique to the UK which has evolved over time. The RPI is aggregated to sections using the leading digits of item ID. For example, Home Killed Lean Mince Meat (item ID 210403) would be allocated to section 2104 (see *rpi_sections* file for other examples). Once that fact has been established, then aggregation from item index level follows similar principles to that of the CPI and CPIH.

For example, the Home Killed Lean Mince Meat (210403) item calculated above can be combined (weighted) with other fresh meat products under COICOP 4 category 01.1.2, Fresh Meat;

1) the sum of coicop_weight [2] is taken
2) the item_index is multiplied by the coicop_weight [1] and then summed [3]the item index [4] is calculated as the sum of products [3] divided by the total stratum_weight [2]

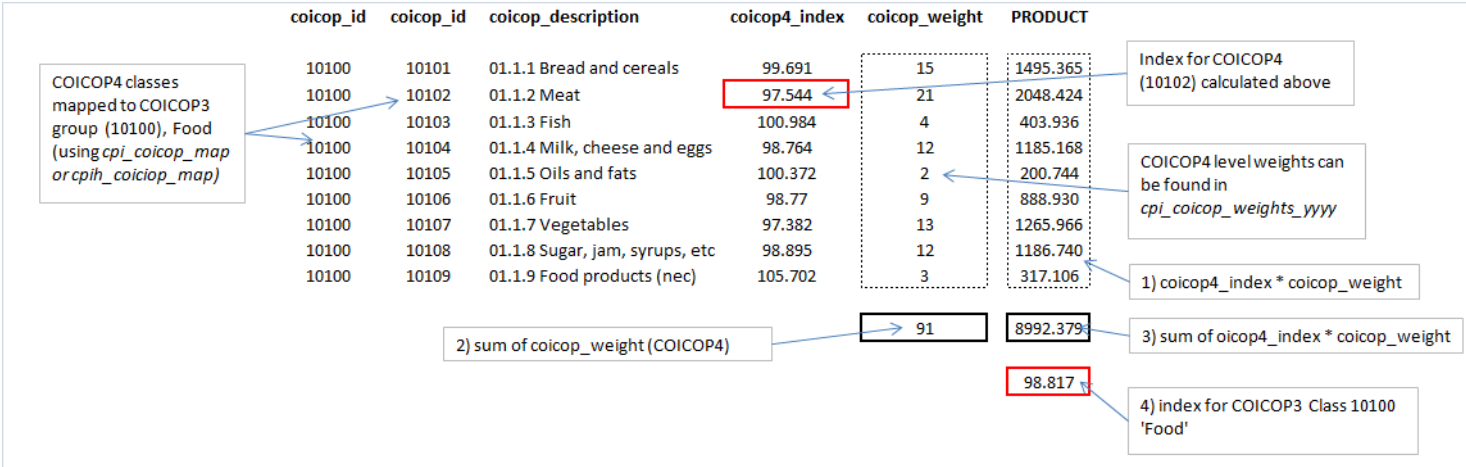
3) the item index [4] is calculated as the sum of all the products [3] divided by the total coicop_weight [2]



coicop_id	item_id	item_description	item_index	coicop_weight	PRODUCT
10102	210403	HOME KILLED BEEF-LEAN MINCE KG	93.787	0.84	78.781
10102	210406	HOME KLD BEEF-RUMP/POPES STEAK	97.625	1.05	102.506
10102	210413	HOME KILLED BEEF-TOPSIDE KG	97.204	0.42	40.826
10102	210414	FROZEN BEEFBURGERS PACK OF 4	96.27	0.42	40.433
10102	210502	HOME KILLED LAMB-SHOULDER KG	100.078	0.42	42.033
10102	210506	HK LAMB LOIN CHOP/STEAK PER KG	97.692	0.21	20.515
10102	210606	FRZEN IMP LAMB: LEG (PER KG)	95.176	0.84	79.948
10102	210703	HOME KILLED PORK-LOIN CHOPS KG	98.647	1.05	103.579
10102	210802	BACON-GAMMON-PER KG	94.086	0.42	39.516
10102	210808	BACON-BACK-PER KG	93.336	1.89	176.405
10102	210905	FRESH/CHILLED CHICKEN PER KG	91.558	0.63	57.682
10102	210910	FRESH BONELESS CHICKEN BREAST	95.585	1.47	140.510
10102	210911	FRESH TURKEY STEAKS PER KG	94.478	0.42	39.681
10102	210912	FROZEN CHICKEN BREASTS	96.708	0.63	60.926
10102	210913	ROTISSERIE CHICKEN [WHOLE]	101.356	0.21	21.285
10102	211007	CANNED MEAT-STEWED STEAK	101.762	0.42	42.740
10102	211010	PORK PIE INDIVIDL APP 3IN/8CM	100.891	0.63	63.561
10102	211011	INDIVIDUAL MEAT PIE	102.52	1.26	129.175
10102	211014	SAUSAGES-PORK-PER KG	99.328	1.68	166.871
10102	211016	COOKED HAM LOOSE SPECIFY TYPE	99.71	0.63	62.817
10102	211019	FROZ CHICKEN NUGGETS 220-600G	96.83	0.42	40.669
10102	211023	COOKED HAM PREPACKED SLICED	96.89	2.1	203.469
10102	211024	CONTINENTAL SLICED DELI MEAT	99.694	0.63	62.807
10102	211025	JOINT OV/READ GAM/POR 450-700G	100.006	0.21	21.001
10102	211026	CHICKEN KIEV 2 PACK 240-310G	95.812	0.42	40.241
10102	211027	H K LIVER PER KG STATE TYPE	99.472	0.21	20.889
10102	211028	MEAT BASED SNACK M/PACK 400G	102.429	1.05	107.550
10102	211029	COOKED MEAT TURKEY/CHICK SLICE	100.006	0.42	42.003
				21	2048.42
					97.544

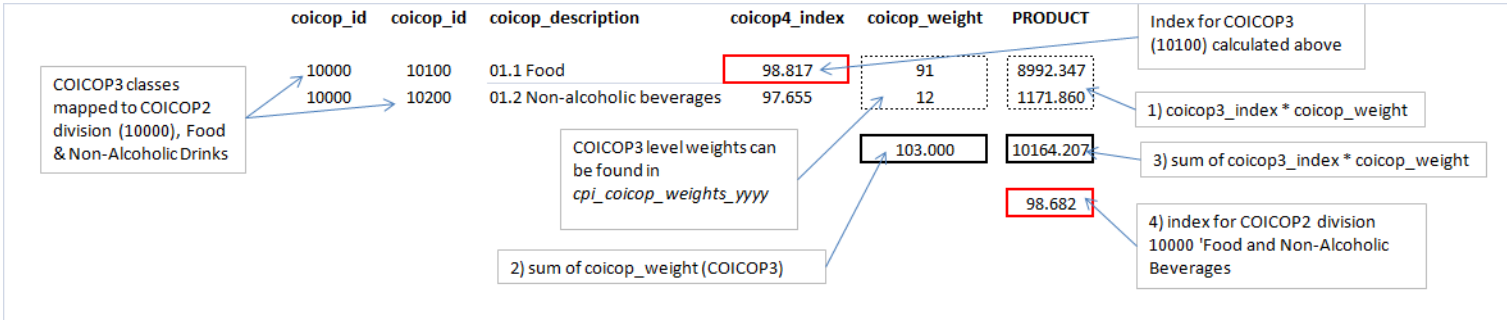
Step 1 to 3 is carried out for all other COICOP4 classes.

4) Class 10102 is then combined with other classes (COICOP4) to create its parent level aggregation (COICOP3), 10100 'Food'.

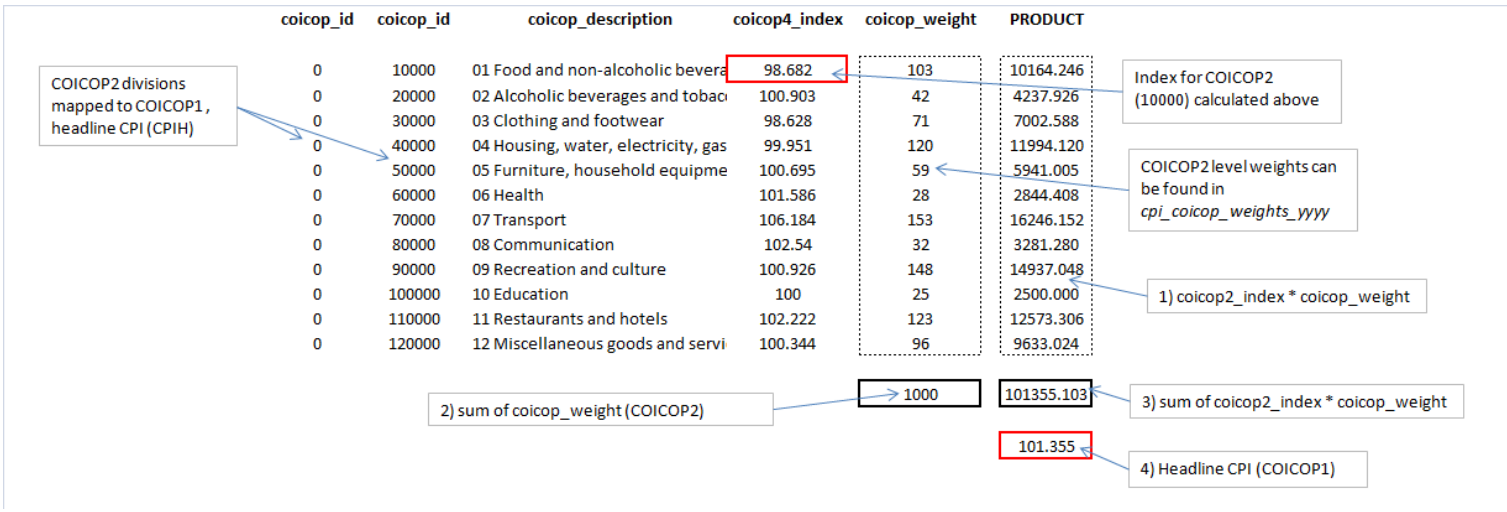


coicop_id	coicop_id	coicop_description	coicop4_index	coicop_weight	PRODUCT
10100	10101	01.1.1 Bread and cereals	99.691	15	1495.365
10100	10102	01.1.2 Meat	97.544	21	2048.424
10100	10103	01.1.3 Fish	100.984	4	403.936
10100	10104	01.1.4 Milk, cheese and eggs	98.764	12	1185.168
10100	10105	01.1.5 Oils and fats	100.372	2	200.744
10100	10106	01.1.6 Fruit	98.77	9	888.930
10100	10107	01.1.7 Vegetables	97.382	13	1265.966
10100	10108	01.1.8 Sugar, jam, syrups, etc	98.895	12	1186.740
10100	10109	01.1.9 Food products (nec)	105.702	3	317.106
				91	8992.379
					98.817

5) Group 10100 is then combined with other groups (COICOP3) to create its parent level aggregation (COICOP2), 10000 'Food and Non-Alcoholic Beverages'.



6) Division 10000 is then combined with other divisions (COICOP2) to create headline CPI (COICOP1).



3.2 Imputation

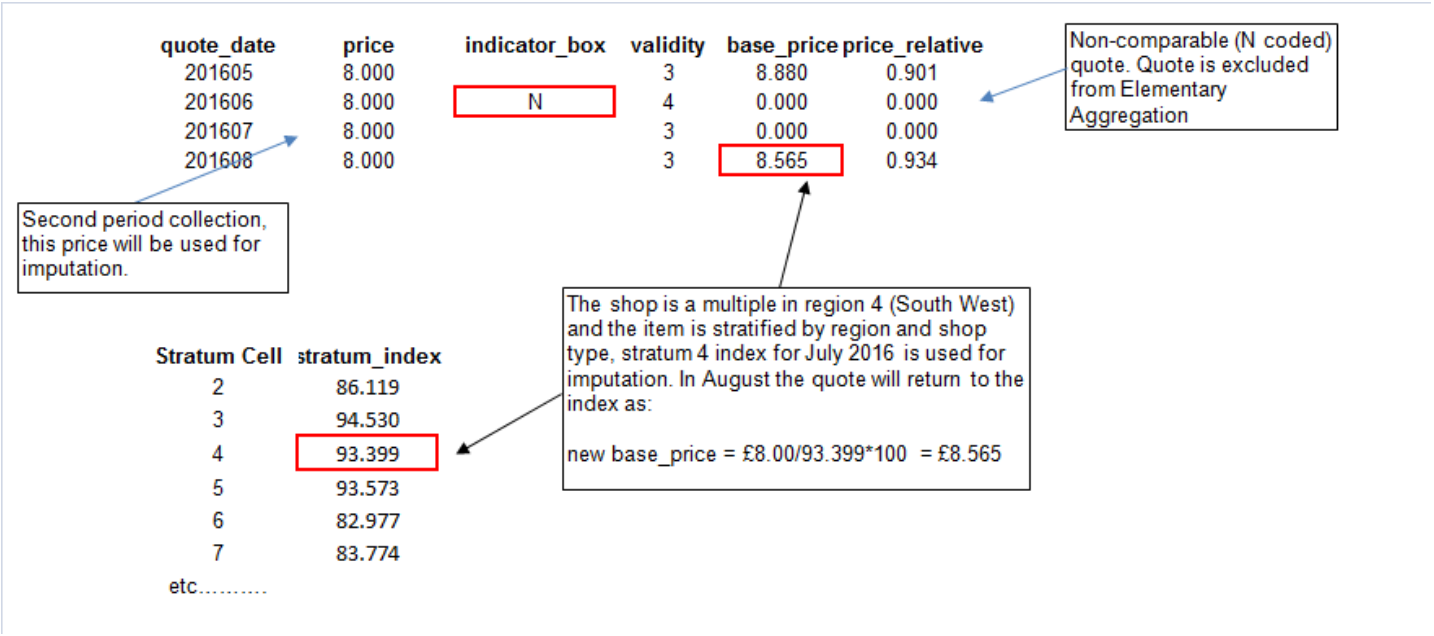
CPI, CPIH and RPI uses class mean imputation using RPI stratum level indices.

Imputation is invoked for individual price quote observations (base_price) under the following circumstances:

- If the previous months product (brand and variety) is unavailable in the current month and its replacement is of a different quality or specification. These types of observation can be identified within the price quote file (*price_quotes*) as having an 'indicator_box' equal to 'N' and are excluded from index compilation for two months.
- If a sampled retailer closes or refuses to participate further in the inquiry, then a new replacement outlet is selected. Base prices for the new outlet are imputed in the following month.
- If an item is moved from one retailer in the sample to another. Base prices for the new outlet are imputed in the following month.

Base price imputations can usually be detected within the *price_quotes* data file where base_price values are shown to three decimal places.

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 change in the elementary aggregate for that item. For example, if the price is £8.00 and the elementary aggregate index for that item (calculated excluding the product in question) in that stratum is 93.399, the new base price is: £8.00/93.399*100 = £8.565



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. Approximately one percent of the sample is imputed each month.

3.3 Outliers

An outlier detection process known as the Tukey algorithm is used to identify price movements which differ significantly from the norm. Once an outlier in the current month's collection has been identified, the failed quote observation is subjected to further scrutiny with a number of follow up actions taking place:

- The failed price quote is accepted if there is sufficient accompanying evidence (notes) to support the individual price that has been recorded
- The failed price quote is queried with the data provider and the price is either subsequently amended and accepted or confirmed with supporting evidence and accepted.
- The quote is manually rejected from the index for the current month if none of the above criteria can be met. Quotes that have been rejected following validation can be identified in the *price*_quote file as having a 'validity' status equal to 2. These

quote observations are excluded from CPI, CPIH and RPI index calculations, but are included in the microdata set in the interest of full disclosure.

The operation of the Tukey algorithm is explained in more detail in **Chapter 6** (sub-section 6.3.3.4) of the CPI Technical Manual with outliers being treated as described in sub-section 6.3.2.

4 Questionnaire and Data Collection

4.1 Data collection

Approximately 200,000 prices are collected in line with the COICOP/HICP classification system, from a selection of items which are representative of UK consumer expenditure. Collections are carried out monthly, quarterly, half yearly or annually depending on the commodity being collected:

Local collections

A third party company, Kantar TNS, are currently contracted to collect approximately 100,000 prices for around 580 items from a variety of retail stores in around 150 locations across the UK each month.

The TNS collectors physically attend sampled outlets and collect the majority of the prices for CPI, CPIH and RPI on the second or third Tuesday of each month. Price, brand and variety descriptions for each price quote observation are manually keyed into an electronic hand-held device at the point of collection.

Each price quote entered is subjected to some initial validation (a month on month percentage price range check, a minimum and maximum level check and checks to ensure that the correct combination of 'indicator-box' and price has been recorded). The TNS collector provides additional commentary (messages) for use by internal ONS staff for any prices that have failed these basic checks.

Once the collection is complete, the data is transmitted to the ONS using specialised file transfer protocols prior to the data being loaded into the ONS (Prices) index processing system by ONS staff.

Regional Central Shop Collections (CES);

Locally collected price data are further augmented by an internal internet collection of approximately 20,000 prices across 429 items from around 46 retailers. Price, brand and variety descriptions are collected directly from retailer web sites and entered onto a central database. Locally collected data are merged with the CES collection prior to a more thorough process of validation being carried out. Validation is explained in more detail in **chapter 6** of the CPI Technical Manual.

Central Spreadsheet Items

There are approximately 80,000 additional prices collected via our central spreadsheet mechanism for around 150 items. The individual item indices are computed within the spreadsheets and then fed onto a central database for combining with other indices computed from the local collections and CES collections (above) before aggregating into the required RPI and COICOP (CPI/CPIH) headings.

These prices are not included within the VML micro data for practical reasons (due to their relative complexity), but pre-calculated item index level values for these items can be found within the *item_indices* file within the micro data release.

4.2 Response rate

Data collected by TNS are delivered in two tranches. The contractual target delivery rate is 79% (interim delivery) with a final response of 84%.

Internal central shop and spreadsheet collections tend to operate closer to 100% response in comparison to the TNS collections. This is because internal sample sizes are smaller and the types of products collected are more consistent over time. Furthermore, it would not be cost effective for ONS to demand a higher, or unachievable, response rate for the TNS collection.

5 Quality Assurance and Validation

5.1 Accuracy

See **Section 4.1, 7.6.2** and **12.1** of the CPI Technical Manual for further information on sampling and non-sampling errors. **Chapter 6** provides information on the forms of validation applied to CP/CPIH data.

5.2 Comparability

See section 3.2 above and **8.2** of the CPI Technical Manual, covering product substitution, quality adjustment and imputation procedures.

6 Datasets

6.1 Types of microdata produced

Consumer Prices Indices are published monthly. Low level micro data (price quotes, weights and mappings) are available within the VML directly after the current month's publication, as set out in section 1.7 above.

The CPI index is revisable, though the only times it was revised were when the index was re-referenced to 2005 equals 100 in January 2006 and to 2015 equals 100 in January 2016. It is unusual to revise the CPI when methodological improvements are introduced, though this is considered on a case by case basis. Consequently there is no regular scheduled revision to

the CPI. The revisions policy for CPIH is the same as that for CPI, but it does not necessarily follow that a revision to CPI would result in a revision to CPIH and vice versa.

In accordance with the Statistics and Registration Services Act (updated 2007) the ONS are not permitted to change the methodology or inputs that underpin the production of the RPI if it changes the index significantly enough to be detrimental to holders of bonds and index linked gilts. This excludes the regular updating of the RPI/CPI basket of goods that takes place each year.

Data are provided in more than one file which are described in section 7.1 below.

6.2 Changes to the dataset

Date	Reason
2006 Q4	From Quarter four 2006, an error for the variable start_date and end_date has been corrected to reflect the base period price start and end date. Prior to this the variable had erroneously shown the items start and end date. It has not been possible to correct this due to the age of the data.
2007 Q1	From Quarter one2008, stratum indicator (stratum_indicator) is no longer available within the microdata price quote file (price_quote_yyyy_qn.csv) and has been replaced by stratum weight (stratum_weight)
2012 Q2	From Quarter three 2013 item descriptions are included within the microdata price quote file (price_quote_yyyy_qn.csv).
2016 Q3	CPIH item index, item weights and higher level (COICOP) weights and COICOP mappings were added to the VML data set for both current and historical periods. To note - unlike CPI, which begins in 1996, the official CPIH series does not begin until 2005
February 2005 – January 2017	CPIH item and higher level weights and COICOP mappings were updated to reflect revised owner occupiers rental weights and the inclusion of council tax within the measure.
February 2017	CPI and CPIH COICOP mappings were updated to reflect the introduction of COICOP Five (ECOICOP) to reflect an additional level of aggregation required by Eurostat.

7 Variables

7.1 Types of variables

The following datasets are supplied (their variable descriptions are given in sections 7.2 to 7.5 below):

- Elementary aggregation;
    - Detailed monthly price quotes (price_quote_yyyy_qn.csv) where yyyy refers to the year in which the price collection took place and n are the quarter months
    - A location table (locationTable.csv) which is linked with the price quote table above to identify the location from which an individual price quote has been observed.
- Item level aggregation;
    - Annual Stratum level weights (stratum_weights_yyyy.csv) where yyyy is the year that the stratum level weights apply to.
- Above item level aggregation (Classification of Individual Consumption by Purpose).
    - Monthly item indices and their associated item weights; CPI, RPI and CPIH (Item_Indices_yyyy_qn.csv)
    - COICOP Mappings (cpi_coicop_map.csv and cpih_coicop_map.csv) for CPI and CPIH respectively
    - COICOP Descriptions (cpi_coicop_descriptions.csv and cpih_coicop_descriptions.csv) for CPI and CPIH respectively
    - Annual COICOP level 4 to 2 weights (cpi_coicop_weights_yyyy.csv and cpih_coicop_weights_yyyy.csv) for CPI and CPIH respectively
- Above item level aggregation (RPI)
    - RPI Mappings (rpi_sections.csv)

7.2 Detailed Monthly price quotes (price_quote_yyyy_qn.csv) and location (locationTable.csv) – Elementary Aggregation

Further information on the construction of Elementary Aggregates for CPI and its variants can be obtained from the CPI Technical Manual, 2014 Edition (**Chapters 2, 10** and **11**).

quote_date The year and month in which the individual price observation was collected

item_id Unique identification number of the item priced

Notes: Item IDs consist of 6 digit reference numbers which can be used to link to COICOP mapping and COICOP description files in order to allocate each item to its constituent COICOP as well as to stratum weights and item indices and weights tables.

item_desc High level description of the item being priced.

validity The validity of the individual price quote observation.

Label values/coding:

Coding	Description
0	Price is outside the min-max range
1	Zero price or failed credibility check
2	Rejected by user
3	Validated
4	Accepted by user
5	Price change failed % test
7	Unknown Indicator Code
8	Ind.= Q/C/N/W but no message exists
9	Price is 0 but Ind. is NOT T or M
10	Ind. is T or M but Price is NOT 0
11	Quote is valid but Ind. is Q/W
15	Scotland eye test charges 2006 - free
51*	Zero Price, temporarily out of stock or missing
52*	Rejected by user
53*	Valid, quote contributed depending on indicator box
54*	Valid, quote contributed depending on indicator box

Notes: Only price quote observations with validity statuses 3,4, 53 or 54 are used for Elementary Aggregation and final index production. Other statuses are provided for additional information.

shop_code Outlet code from which the individual price quote was obtained.

Notes: It is possible to have duplicate shop codes within a region, but these should not be mistaken for duplicate quote observations. Within the CPI/CPIH processing systems a quote record is keyed using its shop code and location value, this makes a unique key for the individual shop record. Since the location variable is not present (due to unreliability and disclosure) within the microdata set, this can give rise to the appearance of duplication.

price Observed price on date of collection.

Range: >=0

- if the variable = 0 then the product to be collected was either out of stock or not available within the store. The price has been excluded from the production of the index for the period.

Units: £GBP

Coverage: All locally collected outlets and internal Central Shop Collections

indicator_box Describes any features of the price observation recorded.

Label values/coding:

Coding	Description
S	Sale price or special offer
R	Recovery of price at end of sale or special offer
T	Temporarily out of stock
M	Missing – not sold at shop
C	Comparable – change in product being priced, new product is similar to the previous product
N	Non-comparable – change in product being priced, new product is not comparable to the previous product
Q	Additional comments have been supplied to assist with internal validation
W	Change in the weight or volume of the product
X	Comparable item on sale
Z	Non-comparable item on sale

Notes: 'N' coded price quotes do not contribute to the months index in which they appear. The quote re-enters index production after its base price has been imputed in subsequent months.

orig_indicator_box_ Describes any original features of the price observation recorded.

Label values/coding:

Coding	Description
S	Sale price or special offer
R	Recovery of price at end of sale or special offer
T	Temporarily out of stock
M	Missing – not sold at shop
C	Comparable – change in product being priced, new product is similar to the previous product
N	Non-comparable – change in product being priced, new product is not comparable to the previous product
Q	Additional comments have been supplied to assist with internal validation
W	Change in the weight or volume of the product
X	Comparable item on sale
Z	Non-comparable item on sale

Notes: Describes the features of the price observation as originally recorded and prior to any internal automated processes (indicator code reversals) or further validation and subsequent amendment.

price_relative The quote index for the individual price observation

Derivation: The individual price relative value is obtained by dividing the current price (price) by the base price (base_price) as follows;

$$PR = \frac{P_t}{P_o}$$

log_price_relative The natural logarithm of the individual price relative (price_relative).

Derivation: Derived from the individual price_relative variable.

stratum_weight The stratum weight applicable to the item for the period.

Notes: The same stratum weights are used in the CPI, CPIH, RPI and RPIJ. Stratum weights are updated in February of each year and lasts until the following January. Stratum weights are used to aggregate the individual stratum level indices to an item level index.

stratum_type Describes the items stratification method.

Label values/coding:

Coding	*Description*
0	Not stratified
1	Stratified by region
2	Stratified by region and shop type
3	Stratified by shop type

start_date Base period price start date

Range: yyyy02 to yyyy01 (following year t+1)

Notes: Base prices start in February, but are based on the collected price from the January collection. A base price start date that is not equal to February (02) indicates that a base price imputation has taken place. This is because either a non-comparable substitution has been made in the preceding months, an item has been placed at an outlet for the first time, or that a new item has been placed at an existing outlet for the first time.

end_date Base period price end date

Range: yyyy02 to 999999

Notes: 999999 denotes a live base period price (whether imputed or not).

region Approximation to Government Office Regions (GOR)

Label values/coding:

Coding	*Description*
1	Catalogue collections
2	London
3	SE
4	SW
5	East Anglia
6	East Midlands
7	West Midlands
8	Yorks & Humber
9	NW
10	North
11	Wales
12	Scotland
13	Northern Ireland

shop_type Shop type indicator

Label values/coding:

Coding	*Description*
1	Multiple (10 or more outlets)
2	Independent (fewer than 10 outlets)

shop_weight The relative weighting of the shop by item

Notes: The same shop weights are used for all index types (CPI, CPIH, RPI and RPIJ). Shop weights reflect the market share of chain shops. They are not strictly weights, but replication factors indicating the number of times each central shop price should appear in each stratum.

base_price Observed or imputed base period price

Range: *>=0*

Units: *£GBP*

Coverage: All locally collected outlets and Central Shop Collections

base_validity The validity of the individual price quote observation.

Label values/coding:

Coding	*Description*
0	Rejected (New quote)
1	Rejected (3 Month Syndrome)
2	Rejected by user
3	Validated by base calculation
4	Accepted by user

stratum_cell Stratum cell of the individual price quote

Label values/coding:

There are 3 types of stratum; by shop, by region and by shop and region.

The stratum types and cells are as follows,

stratum_type	*Coding*	*Description*
Not Stratified	0	Not Stratified
By shop Type	1	Multiple
	2	Independent
By region	1	Catalogue collections
	2	London
	3	SE
	4	SW
	5	East Anglia
	6	East Midlands
	7	West Midlands
	8	Yorks & Humber
	9	NW
	10	North
	11	Wales
	12	Scotland
	13	Northern Ireland

Notes: By region and shop type; if the shop type is a multiple, then the stratum cell is equal to the regional stratum code. If the shop type is an independent, then the stratum cell is equal to the multiple code +13.

location Unique identification number of the location at which an individual quote has been observed.

Location_description Location descriptions (note these values are on a separate table locationTable.csv)

7.3 Item level aggregation (*stratum_weights_yyyy.csv*)

The additional stratum weights file is included as an additional data set to supplement the lower level detailed monthly quote files (*price_quote_yyyy_qn.csv*) before 2007. Stratum weight values are included within the detailed monthly quote files themselves from 2007 onwards.

Item_id Unique identification number of the item

Notes: Item IDs consist of 6 digit reference numbers which can be concatenated with stratum_cell to map the stratum weights values from *stratum_weights_yyyy.*csv to the detailed monthly price quote file *price_quotes_2016_qn.csv* between 1996 and 2006.

stratum_cell Stratum cell number

Label values/coding:

There are 3 types of stratum; by shop, by region and by shop and region.

The stratum types and cells are as follows,

stratum_type	***Coding***	***Description***
Not Stratified	0	Not Stratified
By shop Type	1	Multiple
	2	Independent
By region	1	Catalogue collections
	2	London
	3	SE
	4	SW
	5	East Anglia
	6	East Midlands
	7	West Midlands
	8	Yorks & Humber
	9	NW
	10	North
	11	Wales
	12	Scotland

	13	Northern Ireland

Notes: By region and shop type; if the shop type is a multiple, then the stratum cell is equal to the regional stratum code. If the shop type is an independent, then the stratum cell is equal to the multiple code +13.

start_date Year and month that stratum weight begins

Range: yyyy02 (where yyyy is the year and 02 is February)

end_date Year and month that stratum weight ends

Range: yyyy01 (where yyyy is the year and 02 is February)

stratum_weight The stratum weight applicable to the item for the period.

Notes: The same stratum weights are used in the CPI, CPIH, RPI and RPIJ. Stratum weights are updated in February of each year and lasts until the following January. Stratum weights are used to aggregate the individual stratum level indices to an item level index. To correctly match the correct quote period and stratum weight the quote_date should fall between the stratum weight start_date and end_date.

7.4 Above item level aggregation (Classification of Individual Consumption by Purpose).

Further information on the construction of Elementary Aggregates for CPI and its variants can be obtained from the CPI Technical Manual, 2014 Edition (**Chapters 2, 10** and **11**).

a) Monthly item indices and their associated item weights; CPI, RPI and CPIH (*item_indices_yyyy_qn.csv*)

index_date The year and month to which the index refers

item_id Unique identification number of the item

Notes: The RPI uses its own classification system comprising of groups and sections and was specified and developed by earlier RPI Adivisory Committees. Aggregates can be derived by assigning items to their consistuent groups and sections by extracting the 3rd and 4th

foremost digits from the 6 level item identifier. ***Chapter 10.2*** of the CPI Technical Manual provides a broad relationship between RPI Groups and COICOP Divisions used for CPI/CPIH groupings.

item_desc High level description of the item

index_algorithm The formula used for RPI calculations (as set out in ***section 10.3*** of the CPI Technical Manual, 2014 edition).

Label values/coding:

Coding	*Description*
1	Dutot (ratio of average prices) – RA
2	Carli (average of price relatives) – AR

Notes: CPI and CPIH uses a Jevons or Dutot formulation, the use of a Carli index in the CPI/CPIH is prohibited by HICP regulation.

stratum_ind The stratification method assigned to the item

Label values/coding:

Coding	*Description*
0	Not calculated (central) – items processed separately
1	Stratified
2	Not stratified

intem_index RPI Item Index (using arithmetic means)

Notes: Use the index_algorithm variable to determine the formulae method used

all_gm_index CPI Item Index using Jevons formulation (Geometric Mean)

gm_ra_index CPI Item Index using Dutot formulation – not used in any Consumer Prices Indices production (for information only)

coicop_weight CPI Item level weight

item_weight RPI Item level weight

cpih_coicop_weight CPIH Item level weight (CPI including OOH)

b) COICOP Mappings (*cpi_coicop_map.csv* *and* *cpih_coicop_map.csv*) for Consumer Prices and Consumer Prices including Owner Occupied Housing respectively.

Further information on the structure of the CPI and its classification system, COICOP, can be found in ***Chapter 2.2*** of the CPI Technical Manual, 2014 Edition.

Item_id Unique identification number of the item

coicop_id COICOP 4 level classification number that each item is assigned to

Range: *10101 to 120700*

coicop4_id COICOP 4 level classification number that each item is assigned to (applies from 2015 onwards)

Range: *10101 to 120700*

coicop5_id COICOP 5 level classification number that each item is assigned to (applies from 2015 onwards)

Range: *1010101 to 12070004*

start_date Year and month that the item was assigned to its COICOP 4 or COICOP 5 category

Notes: The start_date represents, on the whole, the date that the item entered the RPI/CPI basket of goods or, less frequently, the classification of an item from one COICOP 4 category to another after 1996.

end_date Year and month that the item was removed from its COICOP 4 category or left the RPI/CPI basket of goods

Notes: The use of 999999 denotes that the item is still active within its category

c) COICOP Descriptions (*cpi_coicop_descriptions.csv* *and* *cpih_coicop_descriptions.csv*) for Consumer Prices and Consumer Prices Including Owner Occupied Housing respectively.

Further information on the structure of the CPI and its classification system, COICOP, can be found in ***Chapter 2.2*** of the CPI Technical Manual, 2014 Edition.

coicop_id COICOP 4:2 level classification numbers

Range: *10101 to 120700*

coicop_type Classification of the COICOP number

Label values/coding:

Coding	*Description*
A	COICOP 1 level
T	COICOP 2 level
G	COICOP 3 level
S	COICOP 4 level

start_date Year and month that the COICOP category became active

end_date Year and month that the COICOP category became inactive

Notes: End dates = 999999 denote that the category is still active.

coicop_description Text description of COICOP categories

d) Annual COICOP level 4 to 2 weights (*cpi_coicop_weights_yyyy.csv* and *cpih_coicop_weights_yyyy.csv*) for Consumer Prices Index and Consumer Prices Index including Owner Occupied Housing respectively.

Further information on the use and construction of COICOP level weights can be found in ***Chapters 2.5*** *and* ***7.6.1*** of the CPI Technical Manual, 2014 Edition.

aggregate_type Classification of the COICOP number

Label values/coding:

Coding	*Description*
A	COICOP 1 level
T	COICOP 2 level
G	COICOP 3 level
S	COICOP 4 level

aggregate_id COICOP 4:2 level classification numbers

household_type Household type

Label values/coding:

Coding	*Description*
1	Standard household
2	One pensioner household (not used in CPI)
3	Two pensioner household (not used in CPI)

start_date Year and month that COICOP 4:2 level weight begins

Range: yyyy01

end_date Year and month that COICOP 4:2 level weight ends

Range: yyyy12

coicop_weight CPI COICOP 4:2 level weight

7.5 Above item level aggregation (RPI).

aggregate_id RPI section or group level classification number

All items index	An index which is constructed using price indices which represent every type of expenditure within the scope of the Consumer Prices Index (CPI) , CPI including owner occupiers' housing costs (CPIH), Retail Prices Index (RPI), the RPI measure using a Geometric (Jevons) formula (RPIJ). It is an average measure of the change in the prices of goods and services bought for the purpose of consumption in the United Kingdom.

description Text description of RPI section and group categories

Annex A - Glossary of terms, concepts and abbreviations

Coverage	Those transactions which it is possible to identify and measure in practice. This is determined by the expenditure categories for which weights are compiled.
Classification of Individual Consumption by Purpose (COICOP)	**Division (COICOP2);** In the CPI and CPIH, all categories of expenditure on which significant amounts of money are spent are arranged into twelve divisions, such as clothing and footwear, transport and recreation and culture. **Group (COICOP3);** In the CPI and CPIH, all categories of expenditure on which significant amounts of money are spent are arranged into twelve divisions, which are subdivided into groups. Examples of groups are food, postal services and insurance. **Class (COICOP4);** In the CPI and CPIH, all categories of expenditure on which significant amounts of money are spent are arranged into twelve divisions, which are subdivided into groups and then into classes. Examples of classes are bread and cereals, water supply and transport insurance.
Elementary Aggregate	The lowest aggregate of prices covering all prices collected for one item in one stratum.
Index Day	The CPI, CPIH, RPI and RPIJ are intended to reflect prices on one particular Tuesday of the month (either the second or third Tuesday) which is known as Index Day. Index Day is therefore the day on which the majority of prices are collected.
Indicator codes	Codes used to identify any special features in the prices recorded. For example, collectors enter an S if the item is on sale or special offer, or an N if the current price is a non-comparable substitute of the previous month's price.
Inflation Rate	The percentage change on a year earlier of a price index. It is usually used to mean the all items inflation rate.
Items	An item is any type of consumer goods or service that can be purchased, for example women's jeans. A number of different brands of that item may be available, for example, women's Levi 501s.

Laspeyres	A base weighted index, i.e. one where the prices are combined using weights derived from data from the base period $I_{t,0} = 100 \times \frac{\sum_i P_{it} Q_{i0}}{\sum_i P_{i0} Q_{i0}}$ where: P_{it} = price for i^{th} item at time t P_{i0} = price for i^{th} item at base rate, time 0 Q_{i0} = quantity of i^{th} item purchased in the base period, time 0
Laspeyres-type	An index such as the CPI, CPIH, RPI or RPIJ which has the basic characteristics of a Laspeyres index. In other words it is a fixed base weight index, being the price of the basket at a given time as a percentage of its price on the base date. The CPI, CPIH, RPI and RPIJ are not true Laspeyres as the base period does not coincide with time 0 (see Laspeyres) but is the most recent available 12 months.
Local price collection	Individual price quotes collected by our external contractor, TNS.
Location	Locations are intended to be broadly representative of a central shopping area and the areas where the local shopping population tend to live.
Outlets (shop codes)	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.
Products/Varieties	These are the varieties in good or service available within an item specification. For example, there are a number of different firms producing automatic washing machines, each firm produces a number of models each with different specifications, but they are all automatic washing machines.
Regional central shops	Regional central shops are chains of shops without a national pricing policy but for which it can be assumed that prices collected in a branch in one region apply to all the branches in that region.
Representative item	Those items that are in the basket of goods and services.
Retailing inquiry	Produced by the ONS, the Annual Retailing Inquiry supplies data on sales by shop type broken down into commodity and service groups and then outlet type, i.e. whether they are independents or multiples. This information is used to construct shop weights, which

	are used in Elementary Aggregation of Consumer and Retail Prices indices.
RPI Classification system	**Section;** In the RPI and RPIJ, all categories of expenditure on which significant amounts of money are spent are arranged into 14 groups, subdivided into about 85 sections. Examples of sections are bread, cigarettes, and postage, footwear and rail fares. **Group;** In the 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.
Sampling frame	A complete list of the objects to be sampled, together with sufficient information on each object to stratify if required.
Scope	All those transactions which one would ideally want to measure.
Shop Weights	Shop type weights were updated annually until 1999 using data collected from the Annual Retailing Inquiry. Following the termination of this Inquiry, shop type stratum weights have been updated where possible using data from various sources, including the LCF. The same shop type stratum weights are used in the CPI, CPIH, RPI and RPIJ.
Strata	Strata are classifications that the raw data can be separated into. In the case of the CPI, CPIH, RPI and RPIJ the strata used are region and shop type within item. The data within each stratum are combined and the resulting indices for each of the strata are then combined together using stratum weights.
Tukey algorithm	The Tukey algorithm identifies and invalidates price movements in the current period which differ significantly from the norm.
Weight	A factor by which a component is multiplied to reflect the level of consumers' expenditure on that component.

Annex B – Diagram illustrating a pared-down version of the COICOP structure.