

ADVISORY PANEL ON CONSUMER PRICES – TECHNICAL

Extending the classification structure

Status: final

Expected publication: alongside minutes

Purpose

1. As part of the Consumer Prices Data Collection strategy on alternative data sources, work has commenced on researching the impact of these data on existing methodology and aggregation structures.
2. This paper sets out one of these projects, in particular whether it is possible to expand the existing classification structure to make better use of the data that will be collected. It outlines some of the initial thinking with regards this project, and details some further work that is required before making a final decision on what changes should be made to the classification structure.

Recommendations

3. Members of the Panel are invited to:
 - a) comment on the initial discussion in Annex A
 - b) advise on whether there are any further considerations that need to be looked at as part of this work

Tanya Flower
Prices Division, ONS
May 2018

List of Annexes

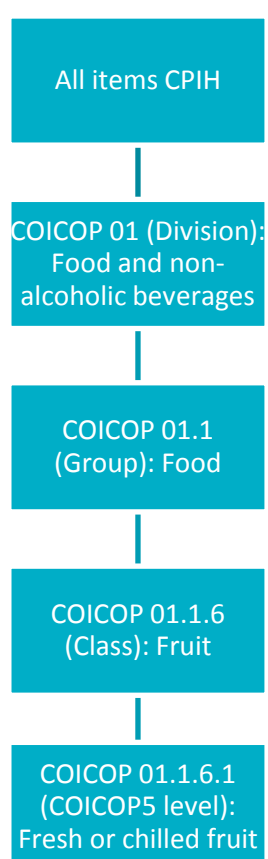
Annex A	Extending the consumer prices classification structure to include alternative data sources
----------------	--

Annex A – Extending the consumer prices classification structure to include alternative data sources

The coverage and classification of the consumer prices indices are based on the internationally agreed classification system for household consumption expenditure known as COICOP (classification of individual consumption by purpose). COICOP is a hierarchical classification system comprising of divisions, groups, classes and COICOP5 level categories, designed to capture all categories of consumer spending.

Figure 1 provides an example showing the classification structure for COICOP5 level 01.1.6.1 Fresh or chilled fruit.

Figure 1: Classification structure for COICOP5 level 01.1.6.1 Fresh or chilled fruit



Underneath the COICOP5 level, ONS adds another layer of categorisation called the item level. These items are selected to be representative of consumer spending patterns within the COICOP5 categories, for example, a representative item for Figure 1 would be “Bananas, per kg”. Price collectors then collect individual products that fit within these tightly defined item descriptions (therefore ensuring these items contain products that are relatively homogenous). However, as these are representative items, categories at this level do not capture an exhaustive list of products that consumers can buy within this COICOP5 category.

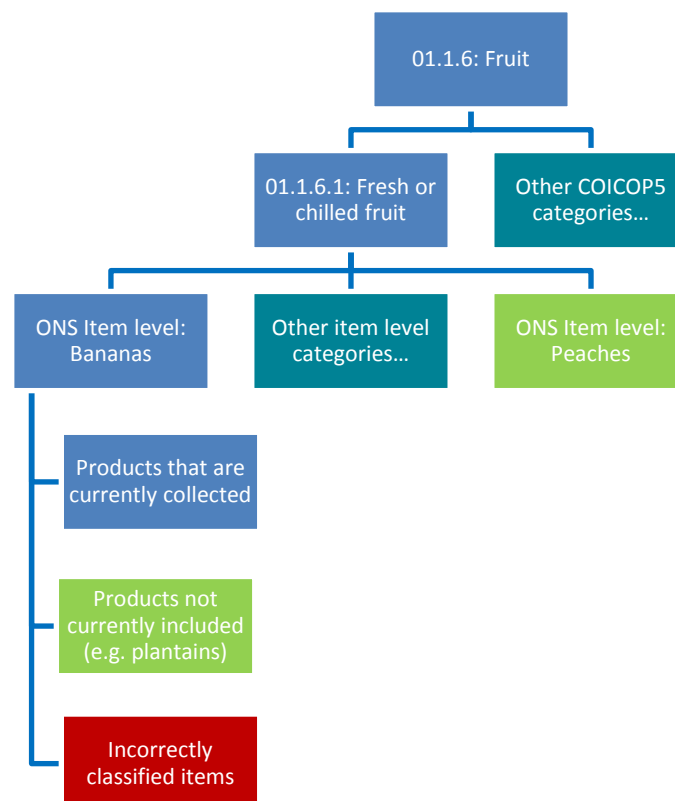
When testing the feasibility of incorporating alternative data sources within consumer price statistics, the first step should be to use the existing classification structure as a target. This means that we would map products collected using these data sources to the current item level categories and exclude all products that would not be captured by the existing classification.

However, this does not fully utilise all possible data that could be collected using these data sources. There are two possible extensions:

1. Expand the list of items that are captured underneath the COICOP5 level
2. Create a new layer of hierarchy at the sub-item level for certain items

Figure 2 demonstrates how this could work with the Fresh or chilled fruit category. At the item level, ONS currently collects prices of bananas but some other types of fruit are not included, for example peaches. Within the item category for bananas, the item description informs the price collector to exclude products such as plantains from the collection. Our alternative data sources may now include data on plantains, but their pricing behavior may be different enough to warrant another “sub-item” level of classification to ensure all products contained within each category remain relatively homogenous.

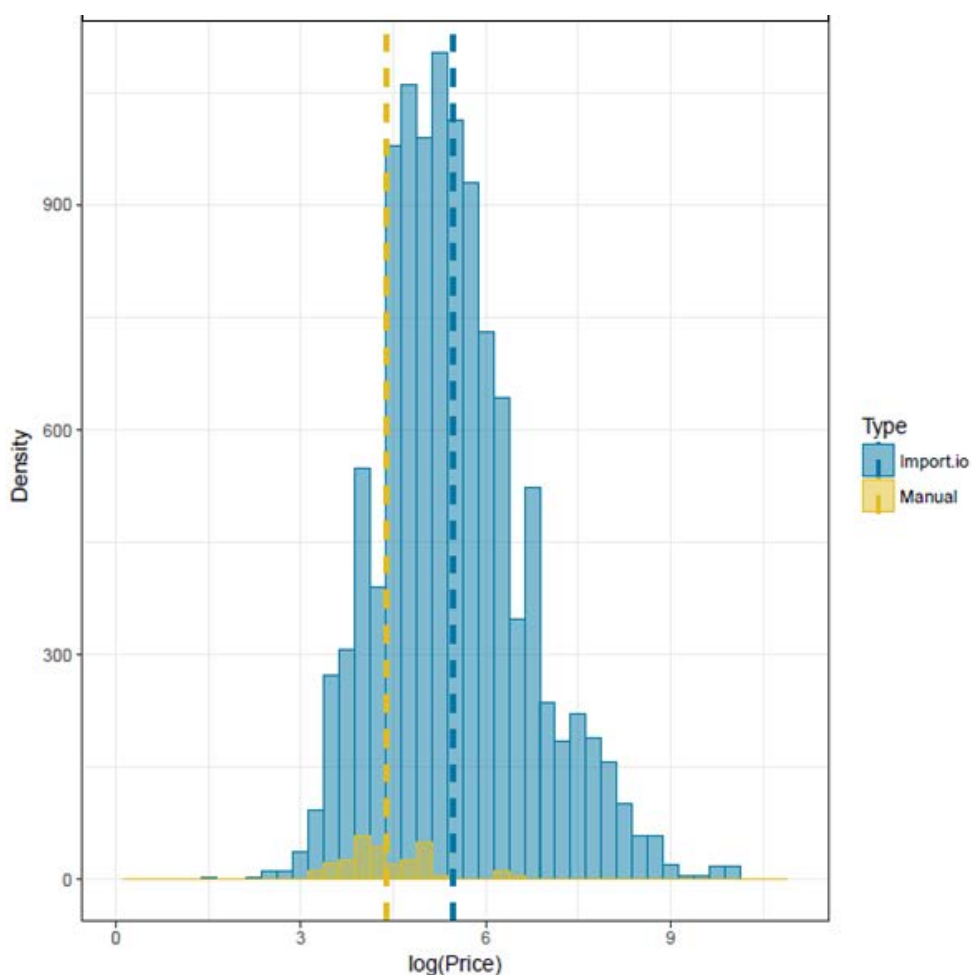
Figure 2: Extending the classification structure for COICOP5 level 01.1.6.1 Fresh or chilled fruit



Another example of where these sub-item categories may be required is for technological goods. Figure 3 shows a frequency plot of the log(price) for printers, where prices have been collected

manually and from web scraping (using the web scraping software, import.io). As well as showing the relative increase in sample size that can be gained, Figure 3 also demonstrates the greater dispersion of prices in the web scraped data. These products could be feasibly separated into sub-item groups (for example, basic printers vs photo printers) if the heterogeneity in prices is identified as an issue in our index calculations.

Figure 3: Distribution of web scraped printer prices



There is therefore scope to extend the classification in both directions, although items that have been incorrectly classified to a particular item category should still be excluded from any aggregation.

However, there are several issues that need to be considered before this extension to the classification structure can be implemented:

1. *Expenditure weights for new categories.* To be able to aggregate up to headline CPIH, these additional categories will need to have associated expenditure weights. This will not necessarily be an issue if we get scanner data but could be more of a problem for web scraped data. We are currently conducting a project that looks at proxying expenditure weights for web scraped data and this could be incorporated into the project scope.
2. *Different levels of hierarchy for different items.* What are the system/methodological implications for having “sub-item” categories for some areas of the basket? For example,

there are some items (as demonstrated) where the current item definition could include heterogeneous products, but there are other areas of the basket (for example, most grocery items) where it is unlikely that products are going to be hugely different at the item level. Will it be possible to adjust the item definitions to include these additional products, rather than creating a sub-item category?

3. *Local collection from independent retailers.* We will still need to run a local collection for some areas of the basket (for example, it is unlikely we will be able to source alternative data to cover independent retailers). Will it be possible to continue to use the existing item list for this collection, but aggregate it together with an extended item list for the alternative data sources? International guidance recommends the use of a different elementary aggregate structure to overcome this problem, for example all bakery items from a supermarket could potentially enter the index, whereas for the traditional bakeries only the items selected for price collection enter.
4. *Does the choice of index methodology determine the criteria used to identify additional groupings?* Certain formulae may behave more sensibly when there is less heterogeneity in the prices/price relatives for certain products.