

ADVISORY PANEL ON CONSUMER PRICES – STAKEHOLDER

Prioritising suitable categories for the inclusion of alternative data sources

Status: FINAL

Expected publication: Alongside minutes

Purpose

1. This paper presents the recommended prioritisation of specific product categories that we will target for implementation of alternative data sources in our aggregate consumer price statistics in 2023.

Actions

2. Members of the Panel are invited to:
 - a) advise on the respective priorities of the product categories considered in this paper
 - b) advise on whether any other product categories should be considered during the first phase of the project, subject to data availability

Background

3. There are several areas of the existing consumer basket where it may be feasible to obtain and incorporate alternative data sources to replace or supplement our existing data collection methods. However, there is limited time before we plan to incorporate alternative data sources into consumer price statistics in 2023. We will therefore prioritise these categories according to several considerations, such as: data availability and quality (including suitable historical time series); the methodological complexity of the category chosen; the ease in which it can be processed using our new IT [processing pipeline](#); and the resources available.
4. In the first phase of the project, up until 2023, we intend to focus on categories where we already have alternative data (or are confident in receiving data shortly). These include:
 - technological goods (for example, laptops)
 - chart collected items (CDs, DVDs, Blu-Rays and books)
 - package holidays
 - airfares
 - clothing (garments)
 - rail fares
 - used cars
 - groceries (depending on the retailers that we can secure scanner data from)
5. The existing pipeline has been applied to technological goods using web-scraped data and should require little change to be used for scanner data items such as groceries. However, due to the complex nature of some categories, we envisage that further work may be required to expand the existing pipeline functionality to include new, or more complex, methods.
6. While we currently hope to implement alternative data sources for all these categories in 2023, there may be future challenges that mean that we will need to prioritise categories accordingly. The priorities, and the complexity of each task, will determine the level of resource we assign to each category. **Members of the Panel are invited to advise on the respective priorities of the product categories outlined in Table 1.**

Table 1: Prioritisation of product categories

Product category	CPIH weight	Description of research	Current priority
Groceries	11.4%	Once we have scanner data from a few retailers, we should be able to produce indices relatively easily using the existing pipeline functionality. The groceries market is dominated by a small number of retailers, so by focusing on scanner data for this category we should be able to improve estimates by using detailed expenditure weights.	High
Clothing (garments)	4.1%	Constructing price indices for clothing items can be problematic because of the high product turnover associated with fashion items. It is likely that work on defining product groups will help determine what methods should be applied to clothing items, but there may also be further research required before a final method can be recommended for clothing items to incorporate alternative data sources.	High
Technological goods	<1%	The pipeline has already been applied to technological goods using web-scraped data, and therefore should be a relatively straight-forward category to implement alternative data sources for by 2023. Technological goods are a good test case for the pipeline given their high level of product churn. The current method for these goods is highly resource intensive, so using alternative data sources in this area may also present opportunities to improve efficiency.	High
Used cars	1.4%	Measurement of changes in used car prices is challenging. Quality adjustments are required to ensure that age and mileage are controlled for throughout the year. The range of approaches employed across EU countries, along with divergent price trends, has prompted Eurostat to review car price index methodology, with a focus on used cars. In early 2018, the Advisory Panels on Consumer Prices recommended that we should review the current methodology for used cars.	Medium
Package holidays	3.4%	In 2015, Eurostat published the UK Compliance Report which stated that our method for package holiday calculation is non-compliant. Price relatives for package holidays are calculated by comparing the price in the current month with the price the same month a year ago, rather than comparing the price with a base period. Work is underway to develop an improved methodology using alternative data sources and also a parallel manual pilot collection (which we intend to drop in favour of alternative data sources).	Medium
Air fares	< 1%	Airfares are one of the more volatile items in the CPIH basket, largely as a result of when the collection day falls (for example, Easter). While their weight is small, they often have an undue influence over the headline rate. Alternative data sources can provide a daily collection of airfares with an increased number of routes. Methodology changes may be required to make more efficient use of this information.	Medium
Rail fares	< 1%	Due to existing data limitations, the current rail fares index is calculated using an imputation based on the cap on regulated rail fares set by the chancellor. The rail fares cap is increased each January and is set with reference to the RPI from the previous July. Alternative data sources can provide a wider dataset that incorporates both regulated and unregulated fares, allowing us to better capture movements in rail fares for consumers over time.	Low
Chart collected items	< 1%	This may require a change to the existing methodology, which follows chart positions over time rather than individual products. Some of these products may become obsolete in the coming years, and the weight for these items is low, hence the low priority currently assigned to these items. However, chart collections can lead to particularly volatile indices that also often have undue influence over the headline rate.	Low

Helen Sands
Prices, Office for National Statistics
September, 2019