Office for
National Statistics

Article

Transformation of consumer price statistics: April 2022

Our plans to transform UK consumer price statistics by including new improved data sources and developing our methods and systems for production from 2023.

Table of contents

1 . Main changes

- We are undertaking a programme of transformation across our consumer price statistics, including identifying new data sources, improving methods, and developing systems to improve both the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI).

- The majority of these changes will not feed through to the Retail Prices Index (RPI), which is a legacy measure of inflation; any proposed change to the RPI will go through the usual governance process – for more information, see Appendix 2 of our Consumer Prices Indices Technical Manual, 2019.

- The Office for National Statistics (ONS) currently collect most local price data from physical stores and shops; in the future, our consumer price statistics will be a mix of traditionally collected data, existing administrative data, and a variety of new data sources outlined in this article.

- These developments are part of a continuous programme of improvement; in 2023, we are prioritising the inclusion of new data and methods in rail fares and second-hand cars, because of new data availability.

- In 2024, our focus will be on groceries and private rents, building up the scale of transformation over time; our ambition is to bring in new data sources for further areas of the inflation basket, and continue to improve these statistics over the coming years.

2 . Overview of the transformation of consumer price statistics

Accurate measures of inflation play a vital role in business, government and everyday life. From rail fares to taxes and pensions, financial transactions and many other areas of our lives are regularly adjusted to reflect the change in prices over time. It is crucial to measure these changes in price as accurately as possible.

We are currently undergoing transformation across many areas of our consumer price statistics, including identifying new data sources, improving methods and developing statistical processing systems.

As part of this transformation, we are updating the way we collect and process price information to reflect our changing economy and produce more robust, timely and granular inflation statistics for businesses, individuals and government.

3 . Alternative data sources

We are introducing several new data sources to help us transform our collection. New data availability has meant we are prioritising data for rail fares and second-hand cars to be incorporated into headline statistics from early 2023.

Web-provided data on second-hand cars

The Office for National Statistics (ONS) have obtained access to data for second-hand cars from Auto Trader, a digital automotive marketplace. This is currently the largest and most visited vehicle advertising website in the UK. The data, dating back to January 2018, are highly informative. They include variables such as:

- date

- advertised price

- condition

- mileage

- year of registration

- make

- model

- engine size

- fuel type

As these data are for advertised vehicle listings, they do not include explicit sales revenue information.

Currently, our prices data for second-hand cars are collected from a sample of around 35 popular car models at three different ages. This is taken from a manual of industry guide prices. Working with Auto Trader will significantly boost the price quotes we use in this area to around 400,000 per month. It will provide a richness of information that will allow for further insights into the drivers of inflation.

Second-hand cars are an area of the basket where we have witnessed significant shifts in consumer trends, particularly over the coronavirus (COVID-19) pandemic. This highlights the growing importance of using alternative data sources to enhance our coverage and understanding of this market.

Transaction data on rail fares

We have gained access to transaction-level data for rail sales in Great Britain, sourced from the rail industry's Latest Earnings Networked Nationally Overnight (LENNON) ticketing and revenue system. This is provided to us by Rail Delivery Group, dating back to January 2019. These are transaction-level data, so explicit information is available on both cost and quantity of each ticket purchased. There is also a wealth of information on price-defining attributes, such as origin and destination station, ticket type and class.

These data cover a near-census of transactions for rail fares in Great Britain. We will be able to use approximately 30 million data points per month, to give us a much better picture of how rail fares are changing. This will include seasonality of pricing as well as accounting for different ticket types, travel classes and railcard usage. We will continue to use current methods of data collection for Northern Ireland, while we explore further data acquisition.

We will be publishing methodological research articles in June 2022 (as outlined in Section 8). These will explain how new methods will be used to process these rail fares and second-hand cars data, and to calculate the indices.

We are also continuing to acquire data for further priority item categories, and the following data sources will feed into our transformation.

Retail scanner data

Scanner data are collected electronically by retailers at the point of sale by "scanning" the barcodes for individual products. They provide us with significantly more information on the quantity and type of products sold, allowing us to reflect changing consumer spending patterns more accurately.

We are in the process of obtaining billions of prices every month from the UK's leading retailers. This will improve our understanding of how prices of products, as well as the number of each product sold, are changing in the UK economy.

We have engaged directly with the UK's largest retailers to gain access to scanner data. Importantly, these data will not disclose individual purchases. Retailers will provide product sales totals for each day, week, or month (dependent on retailer) along with additional information about the products being sold. Many of these retailers, including the Co-op, are already sending us regular feeds of data at daily, weekly or monthly frequencies. Together, they account for nearly 50% of the UK grocery market. We are seeking to acquire data from additional retailers to further increase our market share coverage. More information is available in the article about our retail scanner data on The Times website.

Administrative data

Administrative data are information collected by government departments and other organisations (public, private or third sector), primarily for operational reasons rather than statistical ones. They can provide greater coverage than survey data alone.

We will continue to use administrative data sources that are already currently used for some areas of the consumer basket, such as for owner occupiers' housing costs and private rents. We are now receiving rental information at a microdata level from the Valuation Office Agency, Welsh Government, Scottish Government and the Northern Ireland Housing Executive. These data are being used to develop a new series of private rental prices statistics that will feed into the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI). This is detailed in our Private rental prices development plan, UK: updated February 2022.

For more information on the administrative data sources, see our Quality assurance of administrative data used in consumer price inflation statistics methodology, published 19 July 2017.

Web-scraped data

Web-scraped price data are collected from retailers' websites and can provide a wealth of additional product information about online prices, such as product descriptions. For example, as well as obtaining the price of a laptop, we can collect information such as random-access memory and processor speed. This helps us understand how the quality of products changes over time, an important factor when calculating inflation.

We receive web-scraped data covering areas such as clothing, electronic goods, and package holidays. As there are no historical series available for these data (unlike scanner data), we have been building up a sufficient time series and continuing research before they can be used in official measures.

4 . Why we want to use alternative data sources

Alternative data sources provide many benefits compared with our current sources, including improved product coverage, high frequency of collection, as well as potential cost savings. Scanner and transaction data provide additional information, such as expenditure per product. Web-scraped data contains a rich source of product information that is useful for things like accurate classification and determining quality.

Some types of data also have the potential to provide greater regional coverage of prices and expenditure so that we could produce additional metrics, for example, regional inflation measures. Through the richness of information that is increasingly available to us, we hope to not only improve our headline measures of inflation, but also to better inform the narrative around the drivers of inflation.

5 . Current methods of data collection

We will continue to use traditionally collected price data and our existing administrative data sources in the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI). They will be used when they cannot be replaced by scanner or web-scraped data, such as for small independent shops who do not have a website.

We envisage that in the future our consumer price statistics will be a mix of traditionally collected data, existing administrative data, and the new data sources outlined in Section 3.

6 . Changes in how we process data

These new data sources will result in hundreds of millions more price quotes being processed each month. To maximise their benefit, we will also be making changes to our methodologies and systems.

For any changes in methodologies and systems we have undertaken, and will continue to undertake, extensive stakeholder engagement. We will also continue to publish our Research and developments in the transformation of UK consumer price statistics articles series, released biannually, and impact analyses.

7 . Implementation plan

We plan to include data from alternative sources in our headline consumer price statistics from Quarter 1 (Jan to Mar) 2023, dependent on continuous engagement with users, as outlined in Section 8.

The first divisions we aim to include alternative data sources for in the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI) in 2023, are rail fares (classification of individual consumption by purpose COICOP 07.3) and second-hand cars (COICOP 07.1).

These will be followed by food and non-alcoholic beverages (COICOP 01), alcohol and tobacco (COICOP 02) and private rents in Quarter 1 2024 [note 1]. These divisions will be partially based on scanner, transaction and administrative data, in conjunction with traditionally collected data. More information on the redevelopment plans for private rental prices statistics can be found in our Private rental prices development plan, UK: updated February 2022.

Throughout 2020 and 2021, we have worked on developing the systems for alternative data sources to enable research and impact analysis, alongside research into the methods needed to produce high-quality indices using these data. The results of this work have been published as part of our Research and developments in the transformation of UK consumer price statistics articles series, released biannually, and The redevelopment of private rental prices statistics, intended methodology: March 2022.

The next phase of the project involves the publication of aggregate research indices, using the new rail fares and second-hand cars data, as well as a range of impact analyses. The research will use the data so makes sense to keep in same sentence Following feedback from our stakeholders, we have decided on a publication schedule to allow for user scrutiny and feedback.

Our publication schedule will be as follows:

- in June 2022, we will produce a set of research publications as part of our methodology series, covering our methods and price indices using the new rail fares and second-hand cars data; this will cover data up to February 2022

- in November 2022, we will produce an impact analysis and experimental statistics release, covering data up to July 2022, published following feedback from the [Advisory Panels for Consumer Prices](#) in October

- in Quarter 1 2023, these new estimates will be incorporated in [our Consumer price inflation, UK Statistical bulletins](#).

We expect this publication model to be followed in future years. We will, however, continue to gather feedback from users on this approach and our communications from users and ensure we are best meeting user and stakeholder needs.

Notes for: Implementation plan

1. Private rental prices' statistics are used to inform the owner occupiers' housing (OOH) costs element of the Consumer Prices Index including OOH (CPIH), the Office for National Statistics' (ONS) lead measure of consumer prices inflation, as well as "actual rentals for housing" aspect of Consumer Prices Index (CPI) and CPIH.

8 . Future developments

The changes outlined are the initial stages of a continuous programme of improvement for consumer price statistics. Our ambition is to bring in new data sources for further categories and continue to improve consumer price statistics over the coming years.

Throughout each phase, we will also be liaising regularly with our [Advisory Panels for Consumer Price Statistics](#), our users, and the Office for Statistics Regulation. This is to ensure that our plans for consumer price inflation measurement are appropriate for improving the quality of our statistics and meeting our ongoing user requirements.

Timeline

The following timelines are subject to continued systems development, research and impact analysis to ensure the quality of our statistics, which is our priority. Decisions will be made through continuous engagement with our stakeholders. The timelines are as follows:

- 2020: continued research into the methods required to process alternative data sources

- 2020 to 2021: development of systems for processing alternative and traditional data sources

- 2021: application of methods and impact analyses for priority items

- 2022: recommendations of methods for each item and data source and stakeholder engagement

- June 2022: publication of methods papers and research estimates of the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Price Index (CPI), incorporating alternative data sources for rail fares and second-hand cars

- November 2022: publication of final experimental estimates and impact analysis of using alternative data sources for rail fares and second-hand cars in CPIH and CPI

- 2022: continued research and application of methods, systems development, and impact analyses for additional priority items

- Quarter 1 (Jan to Mar) 2023: alternative data sources for rail fares and second-hand cars used in headline aggregate measures of consumer price statistics

- June 2023: publication of methods papers and research estimates of CPIH and CPI, incorporating alternative data sources for food and non-alcoholic beverages, alcohol and tobacco, and private rents

- November 2023: publication of final experimental estimates and impact analysis of using alternative data sources for food and non-alcoholic beverages, alcohol and tobacco, and private rents in CPIH and CPI

- Quarter 1 2024: alternative data sources for food and non-alcoholic beverages, alcohol and tobacco, and private rents used in headline aggregate measures of consumer price statistics

- 2025 and beyond: rolling programme of improvements for the use of alternative data sources, including the use of additional data in a category, the roll-out of alternative data sources for new item categories, and the methodological and systems improvements required for the use of alternative data sources

9 . Related links

Transformation of consumer price statistics: November 2021
Article | Released 9 November 2021
Our plans to transform UK consumer price statistics by including new improved data sources and developing our methods and systems for production from 2023.

Consumer prices development plan: updated March 2021
Article | Released 16 March 2021
The proposed development plan for consumer price statistics considers a range of priorities, including the rationale for these priorities and their potential impact. Resource requirements and a broad timeline are included to provide users of prices statistics with a greater understanding of the planned development work.

Research and developments in the transformation of UK consumer price statistics: November 2021
Article | Released 9 November 2021
The third in a series of biannual articles to update users on our research to modernise the measurement of consumer price inflation in the UK.

The redevelopment of private rental prices statistics, intended methodology: March 2022
Article | Released 25 March 2022
Overview of the methodology that we intend to use for our redeveloped private rental prices statistics.

Private rental prices development plan, UK: updated February 2022
Article | Released 8 February 2022
Overview of our plans for the statistical development of rental prices statistics, including a timeline for development.

Consumer price inflation, UK: March 2022
Bulletin | Released 13 April 2022
Price indices, percentage changes, and weights for the different measures of consumer price inflation.