

Article

Overview of how we use scanner data in consumer price inflation statistics: January 2026

Overview of how we use scanner data to produce consumer price inflation statistics, explaining the scanner data "basket", how we calculate indices, and how we treat discounts and refunds.

Contact:
Consumer Price Inflation team
cpi@ons.gov.uk
+44 1633 456900

Release date:
28 January 2026

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [What we are measuring](#)
3. [How we calculate price indices with scanner data](#)
4. [Accounting for changes in product quality](#)
5. [The scanner data basket](#)
6. [Accounting for discounts](#)
7. [Accounting for refunds](#)
8. [The impacts of scanner data](#)
9. [Cite this article](#)

1 . Overview

We are introducing groceries scanner data into our consumer price inflation statistics in March 2026.

Scanner data are data that are collected by retailers at the point of sale in-store or online. There are many benefits to using scanner data, compared with our current sources, including:

- collecting prices for all products sold by a retailer, rather than relying on a sample
- tracking them throughout the month, instead of on a single day
- providing more detailed and timely information on the quantity of each product purchased

These data help build a more comprehensive understanding of how prices and consumer spending patterns are evolving. Further details on the benefits of using groceries scanner data are available in [Section 3: Why we want to use scanner data](#) of our Transformation of UK consumer price statistics, groceries scanner data analysis: April 2025 article.

As we introduce new, bigger, "dynamic" datasets, we also need new approaches to the processing and calculation of price indices.

We will initially introduce scanner data for around 50% of the grocery market. We currently collect 25,000 prices per month directly from shops by price collectors. We will now instead use approximately 300 million price points derived from sales of over a billion units of products per month, collected directly from supermarket scanners in-store and online. For the remaining 50% of the groceries market, we will continue to manually collect prices in-store and online.

This article provides an overview of some of the main elements of how we are introducing scanner data into our consumer price inflation statistics. To help users understand the impact of these new data on the Consumer Prices Index (CPI), CPI including owner occupiers' Housing costs (CPIH), and the Retail Prices Index (RPI), we have also published our [Impact analysis on transformation of UK consumer price statistics: January 2026 article](#).

2 . What we are measuring

Consumer price inflation statistics are commonly described as measuring the change in price of a "fixed basket" of goods and services. Historically, we have identified a sample of basket "items" that are representative of what consumers buy and measured the change in price of items in this basket over time. The basket is "fixed" in terms of the items it contains, the quantity of each item in the basket and the quality of those items. This ensures that we only capture changes in price.

However, the idea of a fixed basket is illustrative, and is not a well-defined economic concept. It relates to two concepts: a "cost of goods index" (COGI) and a "cost of living index" (COLI). A COGI is often considered to align with the idea of a fixed basket index, while a COLI accounts for the fact that consumers may change what they buy to less inflationary goods and services. But both concepts measure changes in price.

In practice, a COGI and a COLI are not distinct. If consumers do not substitute what they buy over time, then a COGI is equal to a COLI. Equally, fixed basket COGI principles justify the choice of an index formula that is consistent with a COLI framework.

For example, when measuring price changes with a COGI approach, we could reprice the basket purchased in the base period at current period prices and measure the change in value of that basket. Alternatively, we could reprice the basket purchased now at base period prices as our comparator. Both approaches are equally valid.

We can consider this problem through the lens of "representativity bias". The choice between these two baskets is arbitrary, so a measure that averages across them may be preferred. This average "symmetric" basket would be more representative of consumer spending in both the base and current periods. The extent to which an index based on asymmetric weights leads to different results may be considered representativity bias.

Our preferred method for calculating scanner data price indices is the "GEKS-Törnqvist", which estimates the change in price of a basket that is free of representativity bias.

3 . How we calculate price indices with scanner data

Scanner datasets are dynamic. This is unlike the ongoing collection of prices in the field ("local" collection). For the local collection, price collectors visit outlets in locations across the country, and maintain a stable sample by collecting prices for the same products every month. Scanner data, however, reflect the real world. Product availability changes every month when new products enter the market or old products are discontinued. We therefore use the GEKS-Törnqvist "multilateral" index method to deal with this dynamic churn in product availability.

A technical description of this methodology is available in our [Introducing multilateral index methods into consumer price statistics article](#) and in an update to our [Consumer Price Indices Technical Manual](#), planned for March 2026. We have also published our [How multilateral index methods help us understand grocery scanner data](#) article, which includes a more accessible description of the method.

Overview of the multilateral index method

The GEKS-Törnqvist multilateral approach works by calculating all possible combinations of "chain-linked" index series in a 25-month window of data, and then averaging them. A "chain link" is the mechanism we use for connecting indices with different baskets. In this context, it is used to refresh the sample of products to maximise the product matches available.

The process of linking indices is well established in consumer prices. It is the mechanism we use to introduce new items and remove old ones from the basket of goods and services every year, as described in our [Consumer price inflation basket of goods and service articles](#). It is also used to refresh samples and update the expenditure weights. The GEKS-Törnqvist approach uses the same technique to maximise sample matches.

For example, if we wanted to calculate price growth between January 2024 and January 2026, we could use February 2024 as a link month. Alternatively we could use March 2024, or April 2024 or any of the following months in the window.

While the method uses historical time points as the link month, this does not mean that the resulting index is lagged. The link month serves to maximise product matches. However, the price index shows the change in prices between the base month and the current month, as any other price index method would.

Unlike data collected in the field, we know how much consumers spend on different product varieties. This allows us to reflect the economic importance of different products through "expenditure weights". Because each linked index uses a different link month, they will also have different weights. This is important in a dynamic market where, for example, a new product may be introduced partway through the measurement period. Under this approach, the new product can be given a weight in the calculation.

However, this does not mean that the index is an expenditure index. Linking series with different weights is standard practice (as part of the annual basket update). Averaging over linked-price indices will result in a price index, since it shows the change in prices over time, not the change in expenditure. For example, if prices were to stay the same over the measurement period, but all expenditures doubled, then the index level would remain the same. Conversely, if expenditure was unchanged but all prices doubled, then the index level would double.

The GEKS-Törnqvist methodology has been thoroughly stress-tested as part of [the framework we used to shortlist index number methodologies](#). This included assessing how the method responds to high variance in prices, product obsolescence and economic shocks. The method was found to suitably represent each of the features tested.

A further important consideration is how to use the latest GEKS-Törnqvist estimates to extend the published series every month. This is not straightforward, since using a new 25-month window will also result in a new base. Our "splicing" approach is described in more detail in our [Introducing multilateral index methods into consumer price statistics methodology](#).

4 . Accounting for changes in product quality

It is important for any price index that the quality of products in the sample is held constant, so that changes in product quality do not affect the measurement of price change.

In the local collection, this is managed through a "matched sample". This is where price collectors aim to price the same products every month; where this is not possible, they follow clear procedures to maintain the comparability of the sample, as described in our [Consumer Price Indices Technical Manual, 2019](#). We treat scanner data in effectively the same way. Each unlinked component of the GEKS-Törnqvist is based on a matched sample of transactions. Additionally, the GEKS-Törnqvist mitigates for biases associated with the introduction of new goods into the market through its multilateral approach.

"Shrinkflation" effects (for example, where the size or volume of a product changes, but the price does not) are captured by standardising product sizes. We do this by tracking the price per unit of weight. This is consistent with the treatment of size changes in locally collected data, and ensures that price comparisons are on a like-for-like basis.

As with the local collection, changes in the underlying quality of a grocery product (for example, changes in the ingredients used) are not explicitly captured through this approach. The practical implementation of this would be costly and complex, and is beyond the scope of any price index with real-world applications in decision making. We would not expect all such quality changes to be significant for consumer satisfaction, particularly in the short-term.

There are methods to explicitly value quality change, for example, the use of a "hedonic" model. However, these methods are more appropriate when product lines change quickly and when there are well-defined product characteristics. For example, hedonics are appropriate for the treatment of technological items, where the pace of innovation is swift and improvements, such as increased processing speed, are quantifiable. However, the resource cost of this approach is high, so its use cannot be justified in other areas of the basket.

5 . The scanner data basket

A consumer price inflation basket that incorporates scanner data using the GEKS-Törnqvist method is fixed in terms of:

- the items it contains
- the basket weights
- the quality of basket items

But for some components, we will now use a dynamic sample of prices and transaction-level weights to get a better, representative measure of price change for each component.

For locally collected components, we define a sample of items in the basket of goods and services and follow the price development of these items over time. However, for scanner data, we do not need to rely on a sample, as we have near complete coverage of grocery expenditure categories for each scanner retailer. We have therefore introduced "consumption segments" for scanner data, which are more broadly defined than item definitions, and which make exhaustive use of the scanner data.

Consumption segments are necessarily less homogeneous than items. For example, the consumption segment "rice in all forms (excluding rice flour)" includes three different items:

- basmati rice (500 kilograms (kg) to 1 kg)
- microwaveable rice (220 grams (g) to 280 g)
- rice cakes (100 g to 180 g)

However, the matched model approach and use of transaction-level weights control for differences in quality between the components of a consumption segment. This ensures that only price changes are captured. The use of retailer weights, which reflect their market share in each consumption segment, further ensure that the consumption segment is representative of the market.

For more information on consumption segments, please refer to [Section 4: Case studies of aggregation structures](#) of our [Introducing alternative data into consumer price statistics: aggregation and weights](#) article.

6 . Accounting for discounts

We must also consider the appropriate treatment of discounts. Consumer price indices should reflect the actual prices paid by consumers, so discounted prices are in scope. However, there are challenges in capturing prices paid as a result of discounting.

In the local collection, we capture discounts where we know the discounted price paid by consumers. This excludes discount types such as loyalty card discounts and multi-buy offers, because we do not know the proportion of consumers who paid full price, compared with those who paid the discounted price. This approach assumes that the price evolution of the discounts that we do not capture is consistent with the broader price behaviour of similar products.

With the introduction of scanner data, we now know the average prices paid for every product line, including the prices paid through loyalty card and multi-buy transactions for the retailers involved. While it is not possible to reflect loyalty card and multi-buy offer prices in the local collection, incorporating more discount types through scanner data helps to reduce potential biases in the data collection.

Discounts introduced mid-year will be captured through the decrease in the average price paid by consumers for all discount types, including loyalty card promotions, multi-buy offers and price reductions. Any change in the quantities purchased because of discounting would be reflected in the "representative baskets" used in the GEKS-Törnqvist calculation. However, because GEKS-Törnqvist is a price index and not an expenditure index, we are measuring the change in price rather than the change in baskets.

While the level of prices may be lower, it is important to note that the inclusion of more discount types does not imply that the level of the index will be lower than it was before. This would only be the case if prices for the newly introduced discount types systematically increase more slowly than for other prices. If this were true, then annual rates in the first year would also be subject to a base effect because the lower index in the current year would be compared with the higher index from the previous year, resulting in a lower annual rate. However, if there are no systematic differences in price growth, there will be no shift in the level of the index (or corresponding base effect).

For more information on the treatment of discounts, please refer to our [Introducing grocery scanner data into consumer price statistics methodology](#).

7 . Accounting for refunds

Products that are returned to the retailer for a refund are not consumed and are therefore out of scope of a consumer prices index. The treatment of refunds in the raw scanner data is not consistent between suppliers. Therefore, it is not always possible to remove refunded expenditure. However, within groceries, refunds are a very small proportion of expenditure, and we therefore expect the impact to be negligible.

For locally collected data, we do not have transaction-level weights, and lower-level expenditure weights are only available at a lag. We would therefore expect scanner-based indices to more accurately reflect what households have consumed.

For more information on the treatment of refunds, please refer to our [Introducing grocery scanner data into consumer price statistics methodology](#).

8 . The impacts of scanner data

In this article, we have described how we deal with common issues in producing consumer price inflation statistics when using scanner data. We give an indication of the expected impact from introducing scanner data into our consumer price inflation statistics in our [Impact analysis on transformation of UK consumer price statistics: January 2026 article](#). These impacts are because of the following factors, acting in combination:

- the use of consumption segments, which means that prices for a broader range of goods are captured (as described in [Section 5: The scanner data basket](#))
- the inclusion of multi-buy and loyalty card discount types, which have not previously been captured (as described in [Section 6: Accounting for discounts](#))
- the use of the GEKS-Törnqvist to construct indices, rather than unweighted methods (as described in [Section 3: How we calculate price indices with scanner data](#))
- scanner-data prices reflect the average price paid over three weeks of the month, whereas locally collected prices reflect the price at a particular point in the month (for more information, please refer to our [Introducing grocery scanner data into consumer prices statistics methodology](#))
- scanner data include transactions from all of that retailer's stores, whereas locally collected data only include a sample of stores

It is not possible to isolate the impact of each of these points separately.

9 . Cite this article

Office for National Statistics (ONS), released 28 January 2026, ONS website, article, [Overview of how we use scanner data in consumer price inflation statistics: January 2026](#)