

How multilateral index methods help us understand grocery scanner data

How we use the GEKS-Törnqvist index method with grocery scanner data. Includes comparison with other methods.

Contact:
Consumer Prices Methods
Transformation
cpi@ons.gov.uk

Release date:
28 January 2026

Next release:
To be announced

Table of contents

1. [Overview of new scanner data methodology](#)
2. [The fixed basket](#)
3. [GEKS-Törnqvist methodology](#)
4. [Benefits of the GEKS-Törnqvist](#)
5. [Comparing the GEKS-Törnqvist to other methods](#)
6. [Next steps](#)
7. [Related links](#)
8. [Cite this methodology](#)

1 . Overview of new scanner data methodology

We have started using scanner data within our consumer price statistics, and from 2026, we will greatly expand our coverage by introducing scanner data for groceries.

For scanner data, we are using an index method, called the GEKS-Törnqvist, for aggregating prices into price indices. We previously explained the methodology underpinning the GEKS-Törnqvist, along with the reasons why we have chosen to use the GEKS-Törnqvist in our [Introducing multilateral index methods into consumer price statistics methodology](#).

In this article, we give a summary description of this method and consider some of the theoretical considerations of using the GEKS-Törnqvist in our price statistics.

2 . The fixed basket

National Statistical Offices often use the description of a "fixed basket" to explain the way in which inflation statistics are measured. The idea is to measure the price change of a basket based on fixed quantities. A simplified example might be to measure the evolving price of a basket featuring three apples, two bananas and an orange. Such a basket may cost £2 in January, £2.20 in February and £2.10 in March. We can then measure price change by dividing the total price of the measurement period basket (in February, March) by the total price of the base period basket (January).

The quantities we use in the basket are intended to represent consumer spending. In the example, we used three apples, two bananas and one orange, but these quantities may have been different, depending on which period we measure the quantities in.

Different index methods use quantities measured within different periods:

- Young and Lowe – quantities from before the base period
- Laspeyres – quantities from the base period
- Paasche – quantities from the measurement period
- Fisher – quantities reflecting an averaging of base and measurement period

Typically, the earlier the period we obtain our quantities from, the higher our fixed basket index will be (as described in section 8.125 of the [IMF Technical Manual](#)). In general, the Lowe, Young and Laspeyres give higher indices than the Fisher, and in general, the Fisher gives higher indices than the Paasche.

As the goal is to measure the price change between the base and measurement periods, it may be considered preferable for the basket quantities to be representative of both periods, leading to a preference for a Fisher basket index. The difference between the Fisher index and the other methods can then be described as "representativity bias", as explained in [European Central Bank Working paper 130 \(PDF, 856KB\)](#). This bias tends to be positive for earlier-weighted indices (Lowe, Young, Laspeyres) and negative for later-weighted indices (Paasche).

However, it would not be practical for us to use a Fisher, since the sources we use to obtain our weights provide data on a lag. As a result, we use the Lowe index for higher-level aggregation. Previous empirical analyses suggest the resulting representativity bias from using early-weighted indices to be small, as shown in [Empirical findings on upper-level aggregation issues in the Harmonised Index of Consumer Prices \(PDF, 448KB\)](#).

Historically, weights tend to only be available for higher-level aggregation. The data we use to calculate weights are generally granular enough for distinguishing between different product groups (for example, comparing apples as a whole and bananas as a whole) but not for comparing specific products (for example, comparing two banana products). Therefore, we tend to use unweighted index methods for elementary aggregation.

However, the scanner data that are now available to us not only gives us weights at the product level but also gives us the contemporary weights we need to explore methods beyond the Lowe. We have chosen to use the GEKS-Törnqvist for this.

3 . GEKS-Törnqvist methodology

There are several parts to the GEKS-Törnqvist methodology:

- the Törnqvist index
- applying the GEKS methodology to the Törnqvist index
- using splicing and choosing a window length

Transitivity

It is important to understand the concept of a transitive index method, before discussing the various methods.

There are usually two ways of calculating an index method. The first is called a direct index, where we compare a measurement month with a base month, regardless of how far these two months are apart.

The second is called a chained index, where we compare a measurement month with a base month by multiplying short-term comparisons.

For example, calculating the Jevons from January to March would be considered a direct index between January and March, but multiplying the Jevons from January to February by the Jevons from February to March would be considered a chained index from January to March.

For some index methods, the direct and chained approaches give the same result. We call these "transitive index methods". For others, this is not the case and the difference between the direct and chained approaches is then described as "chain drift". Chain drift is problematic since the difference occurs as a consequence of the chaining process, rather than because of genuine price change. It can lead to situations where, for example, the prices are the same in both the base and measurement months, but the index does not return to unity as a result of prices changing in one of the chained months then resetting by the measurement period.

Törnqvist

Traditionally we have used the Jevons index for most elementary aggregation within the CPI and CPIH. The Törnqvist can be considered a weighted version of the Jevons.

Both the Jevons and Törnqvist involve calculating a (geometric) average of price relatives. The difference is that when averaging, the Jevons assigns each product an equal weight, whereas the Törnqvist assigns each product a weight by using the average of its base and measurement month expenditure shares. In doing so, each product is treated in line with its economic importance when comparing these two months.

Empirically, the Törnqvist and Fisher indices tend to approximate one another very closely, and so the Törnqvist index can be understood to approximate a fixed basket index based on symmetric use of base and measurement month weights.

However, there are two reasons we do not use a Törnqvist with scanner data. The first is that the Törnqvist is not a transitive index and therefore may be subject to chain drift. The second is that the Törnqvist cannot account for the influence that emerging and leaving products have on price change and is therefore unable to maximise use of the scanner data. Both challenges are solved with the GEKS-Törnqvist.

GEKS-Törnqvist

Traditionally, we update our basket at the start of each new year, using the new annual basket to update both the types of goods and services we collect, along with the product samples we collect for each of these types. This avoids our basket becoming unrepresentative. However, we may still experience products leaving the basket within a single year.

In traditional practices, we use a variety of techniques that allow us to replace leaving products and therefore retain our sample counts. These methods usually require manual scrutiny, so do not scale well to the large product counts we have with scanner data. These methods also do not allow us to systematically introduce emerging products into our sample at the point of their introduction to the market.

For example, a product that is available in February and March, but not in January could not be accounted for with the direct Törnqvist from January to March because of the lack of a price in January.

However, if we calculate a chained Törnqvist from January to March by using February as a link month – that is, a January-February Törnqvist multiplied by a February-March Törnqvist – then the influence of the emerging product can be captured in the latter Törnqvist.

Performing such month-on-month chaining can result in substantial chain drift and therefore tends not to be used by most statistical organisations. However, the GEKS methodology can be used to uphold the principle of using chained comparisons to capture the influence of new products, while still producing a transitive index.

The following is a simplified example with a five-month window covering January to May. The GEKS-Törnqvist from January to May is calculated as the geometric average of the following five pairs of Törnqvists:

- Törnqvist (January, January) multiplied by Törnqvist (January, May)
- Törnqvist (January, February) multiplied by Törnqvist (February, May)
- Törnqvist (January, March) multiplied by Törnqvist (March, May)
- Törnqvist (January, April) multiplied by Törnqvist (April, May)
- Törnqvist (January, May) multiplied by Törnqvist (May, May)

Each pair of Törnqvists represents a chained Törnqvist from January to May, each using a different link month. The GEKS-Törnqvist can therefore be thought of as a geometric average of many different (equally reasonable) chained Törnqvists.

A product that emerges in April and is present in May, but is not available for purchase in the other months, can be represented in the Törnqvist (April, May) comparison, and therefore influences the overall index. If we had calculated a simple direct Törnqvist index between January and May (that is, Törnqvist (January, May)), then it would not have been possible for this product to influence the index.

Splicing and window length

In the GEKS-Törnqvist example given previously, we showed the calculation of indices within a five-month window running from January to May. To extend this index series into June, as the June data become available, we calculate a second five-month window, this time running from February to June, and use splicing to combine these two windows into a single-index series running from January to June.

While the GEKS-Törnqvist is free of chain drift, introducing splicing causes us to introduce some chain drift into the method. However, it is empirically shown in [Multilateral index number methods for Consumer Price Statistics](#) that using a longer window allows us to mitigate chain drift. In the example we gave, we showed how a five-month window GEKS-Törnqvist is calculated. In line with most other countries, in a production setting, we will be using a 25-month window. That means to calculate the GEKS-Törnqvist from the base period to some measurement period, we will calculate a geometric average of 25 pairs of chained Törnqvists, each pair using a different link month.

In an example where we calculate two windows:

- Window 1 covers January 2024 to January 2026 (with January 2024 as a base month)
- Window 2 covers February 2024 to February 2026 (with February 2024 as a base month)

We now wish to combine these windows into a single-index series running from January 2024 to February 2026, with a single base month. To do this, we splice the second window onto the first. The second series contains many months that overlap with the first series (February 2024 to January 2026) along with one non-overlapping index (February 2026). To splice, we calculate ratios to quantify the multiplicative difference between the overlapping index series, which we can use as a multiplier to adjust the non-overlapping index. This adjustment allows us to link the new index onto the existing series with a shared reference month.

For a more mathematical explanation of these steps, see Table 9 in Section 6 of our [Introducing multilateral index methods into consumer price statistics](#).

4 . Benefits of the GEKS-Törnqvist

For elementary aggregation within our alternative data sources, we will use the GEKS-Törnqvist with a 25-month window, using a "mean splice", as described above (on the published series) to splice our windows together. There are several advantages to using the GEKS-Törnqvist:

- it makes use of contemporary weights offered to us within scanner data
- it allows emerging and leaving products to influence the overall index, circumventing the need to apply a product replacement strategy often used in traditional measurement
- while not a full international consensus, it is the most widely adopted method for the earliest adopters of scanner data (including Australia, Canada, Belgium, Luxembourg, Norway, New Zealand)
- it is among the easier-to-understand multilateral index methods

The GEKS-Törnqvist generally gives similar results to the Fisher index, which can be understood as a fixed basket with symmetric weights (and therefore minimises overall representativity bias).

5 . Comparing the GEKS-Törnqvist to other methods

In this section, we will use scanner data to compare the GEKS-Törnqvist against a variety of traditional "bilateral" index methods to determine how close it comes to the other methods. The index methods we will compare against includes the:

- Laspeyres
- Paasche
- Fisher
- Törnqvist
- Jevons

For these comparison indices, we will use an annually chain-linked basket with January as the link month. Unlike traditional practices, we will not use any replacement strategies for handling products that drop, because such methods cannot be scaled to big datasets.

For the weighted methods (Laspeyres, Paasche, Fisher, Törnqvist), we can use the quantities present in the scanner data as weights. However, the unweighted method (Jevons) implicitly uses equal weights for all products, meaning the index will react just as strongly to a product with £10,000 annual sales as it would to a product with £1 million annual sales. Traditionally we circumvent this issue by focussing our price collection on the products that we believe to have the highest annual sales. Therefore, when using the Jevons, we will filter to the top-selling products accounting for the top 95% of expenditure (which covers approximately 50% of the products). This will result in statistics that are more similar to our traditional statistics.

For this analysis, we will use scanner data from multiple retailers covering 4,935 elementary aggregates from January 2019 to November 2024 (therefore covering annual rates for 59 months between January 2020 and November 2024). This leads to having 275,563 annual rates per index method covering this time period (noting that not every elementary aggregate has annual rates available for every month within this 59-month period).

We can then calculate distributions of differences of each index method to the Fisher. To give an example, in March 2022, one of the elementary aggregates may have an annual rate of 2.2 for the Laspeyres and 1.8 for the Fisher. Then we can see the difference between the two to be 0.4. Using all 275,563 annual rate differences, we can plot a distribution of differences between the methods, with the y-axis indicating the count of aggregates with the stated level of difference in the x-axis.

Most of our comparisons involve comparing some index methods against the Fisher. The Fisher index is not a target index, but it is used as a reference point.

As a benchmark, we first explore a histogram of differences between the Törnqvist and Fisher. As can be seen in Figure 1, the Törnqvist produces annual rates that are extremely close to the Fisher, with the vast majority of Törnqvist annual rates being within 0.5 index points of the Fisher. This is in line with theory, as the Törnqvist is expected to approximate the Fisher.

Figure 1: Difference in annual rates between the Törnqvist and Fisher

Törnqvist – Fisher elementary aggregates, annual rates, January 2020 to November 2024

In Figure 2, we now compare the other methods against the Fisher. As expected, the histograms demonstrate that the Laspeyres and Paasche give, in general, higher and lower annual rates than the Fisher, respectively.

Figure 2: Difference in annual rates between various index methods and the Fisher

Index method – Fisher elementary aggregates, annual rates, January 2020 to November 2024

On the other hand, the differences between the GEKS-Törnqvist and the Fisher produce a symmetric distribution. Both the GEKS-Törnqvist and Fisher use symmetric weights, so it is unsurprising that the GEKS-Törnqvist is not producing systematically higher or lower indices compared with the Fisher. The GEKS-Törnqvist distribution of differences with the Fisher is wider than the Törnqvist distribution of differences. This may be partly explained by the GEKS-Törnqvist also capturing the influence of emerging and leaving products (in a way neither the Fisher nor the Törnqvist is able to do).

Interestingly, the distribution of differences between the Jevons and Fisher is also symmetrical, although it is much wider. As both the Jevons and GEKS-Törnqvist distributions of differences with the Fisher are symmetric, we might expect that switching from using the Jevons (with traditional data) to the GEKS-Törnqvist (with scanner data) might lead us to more closely approximate a Fisher, without the overall index being higher or lower because of our chosen index methodology. In Figure 3, we compare the annual rates of the GEKS-Törnqvist and the Jevons, and as expected, the distribution is also symmetric.

Figure 3: Difference in annual rates between the GEKS-Törnqvist and Jevons

GEKS-Törnqvist – Jevons elementary aggregates, annual rates, January 2020 to November 2024

6 . Next steps

We have already started using the GEKS-Törnqvist in our price statistics when using alternative data for rail fares and second-hand cars. However, these are relatively small areas of the basket. In March 2026, we will be using the GEKS-Törnqvist as our index method for elementary aggregation when we introduce grocery scanner data into our basket. We have published an [Impact analysis on transformation of UK consumer price statistics](#).

7 . Related links

[Introducing multilateral index methods into consumer price statistics](#)

Methodology | Updated 28 November 2022

How we will use the GEKS-Törnqvist to introduce alternative data into our consumer price statistics, including how the method works and its advantages.

[Overview of how we use scanner data in consumer price inflation statistics: January 2026](#)

Article | Released 28 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.

[Introducing grocery scanner data into consumer price statistics](#)

Methodology | Updated 29 April 2025

From 2026, we intend to introduce grocery scanner data into our consumer price statistics. This methodology article gives an overview of the data and methods we will use.

[Impact analysis on transformation of UK consumer price statistics: January 2026](#)

Article | Released 28 January 2026

Indicative impacts of the planned improvements to our consumer price statistics from January 2019 to June 2025.

8 . Cite this methodology

Office for National Statistics (ONS), revised 28 January 2026, ONS website, methodology article, [How multilateral index methods help us understand grocery scanner data](#)