Office for National Statistics

Article

Using transaction-level rail fares data to transform consumer price statistics, UK

Rail fare transaction data will improve measurement of consumer prices from 2023. This article details methods and provides research indices using these data.

Contact:
Helen Sands and Joe Barker
cpi@ons.gov.uk
+44 1633 456900

Release date:
28 June 2022

Next release:
To be announced

Table of contents

1. Overview

2. Background to using rail fare prices in consumer price statistics

3. Aims of new data and methods

4. New transaction-level rail fare data

5. Proposed methodology

6. Results: price indices using new data and methods

7. Future developments

8. Related links

1 . Overview

- Alternative data sources, and methods to use these data sources, are being introduced from 2023, as detailed in our Transformation of consumer price statistics: April 2022 article.
- In 2021, we obtained access to daily transaction-level sales data for rail fares in Great Britain (GB) dating back to January 2019, sourced from the rail industry's Latest Earnings Networked Nationally Over Night (LENNON) ticket revenue system; it does not retail.
- This article details our proposed methodology for these rail fares data to be implemented in the headline Consumer Prices Index including owner occupiers' housing costs (CPIH) and Consumer Prices Index (CPI); details on our existing methods can be found in our Consumer Prices Indices Technical Manual, 2019 .
- The research indices presented are broadly in line with the trends seen in our published data, highlighting the quality of our historic measurement of rail fare inflation; however, with these new data we can produce more granular statistics which offer additional insights into the components driving rail fare inflation in the UK, including regional breakdowns.
- This work also ensures that any future changes in pricing policies for rail fares are more appropriately captured, and that the methods and systems that have been developed can be used to onboard further data sources in future, such as electronic point of sale scanner data, as part of our programme of continuous improvement.
- If we had used these data and methods between February 2019 and February 2022, there would have been negligible impact on the headline rate for CPIH and CPI, but our understanding of what was driving price changes in this category would have been substantially improved.
- We plan to introduce these changes in our CPIH and CPI calculations in February 2023, published in March 2023.

2 . Background to using rail fare prices in consumer price statistics

Rail fares in the UK are complex, with around 40% of rail fares being "regulated". Regulated fares are standard class fares including most saver and standard returns, off-peak fares between major cities, and season tickets for most journeys. Unregulated fares include first class and advance purchase. Train operators are free to determine the price of these unregulated fares, although they can be capped in certain circumstances.

Price changes for regulated fares in Great Britain (GB) are all capped by the government based on the annual change in the Retail Prices Index (RPI) in July of each year. This annual uplift to fares (reported by the Rail Delivery Group each year) is currently used to calculate the consumer price index for GB rail fares and is then aggregated with a similar annual figure for Northern Ireland (NI). The weights for this aggregation of GB and NI are based on the total franchised passenger revenue published by the Office of Rail and Road (ORR) versus the total passenger receipts as published by the Department of Infrastructure in NI.

The transformation discussed in this article regards the price index for rail fares only; the calculation of weights for the rail fares category remains unchanged. These weights are subject to the annual updating of the inflation basket and corresponding weights, as detailed in our Consumer price inflation basket of goods and services: 2022 article. In 2022, rail fares have a weight of 2.8 parts per thousand (0.28%) in the Consumer Prices Index including owner occupiers' housing costs (CPIH), and 4.2 parts per thousand in the Consumer Prices Index (CPI) (0.42%).

3 . Aims of new data and methods

The proposed methodology will offer some key improvements, including:

- increased product coverage, improving the representativity of our rail fares index and allowing for the calculation of more granular indices
- the daily delivery of data will result in much more timely prices, allowing a better understanding of, and improved responsiveness to, any seasonal fluctuations in price
- producing rail fare indices on a regional basis, allowing analysis of geographical variations in price change
- ensuring that our indices will be more responsive to potential future changes in rail pricing policies

The methodology proposed in this article relates to our current research indices which may be subject to minor alterations. Our final methodology and impacts will be published in November 2022.

4 . New transaction-level rail fare data

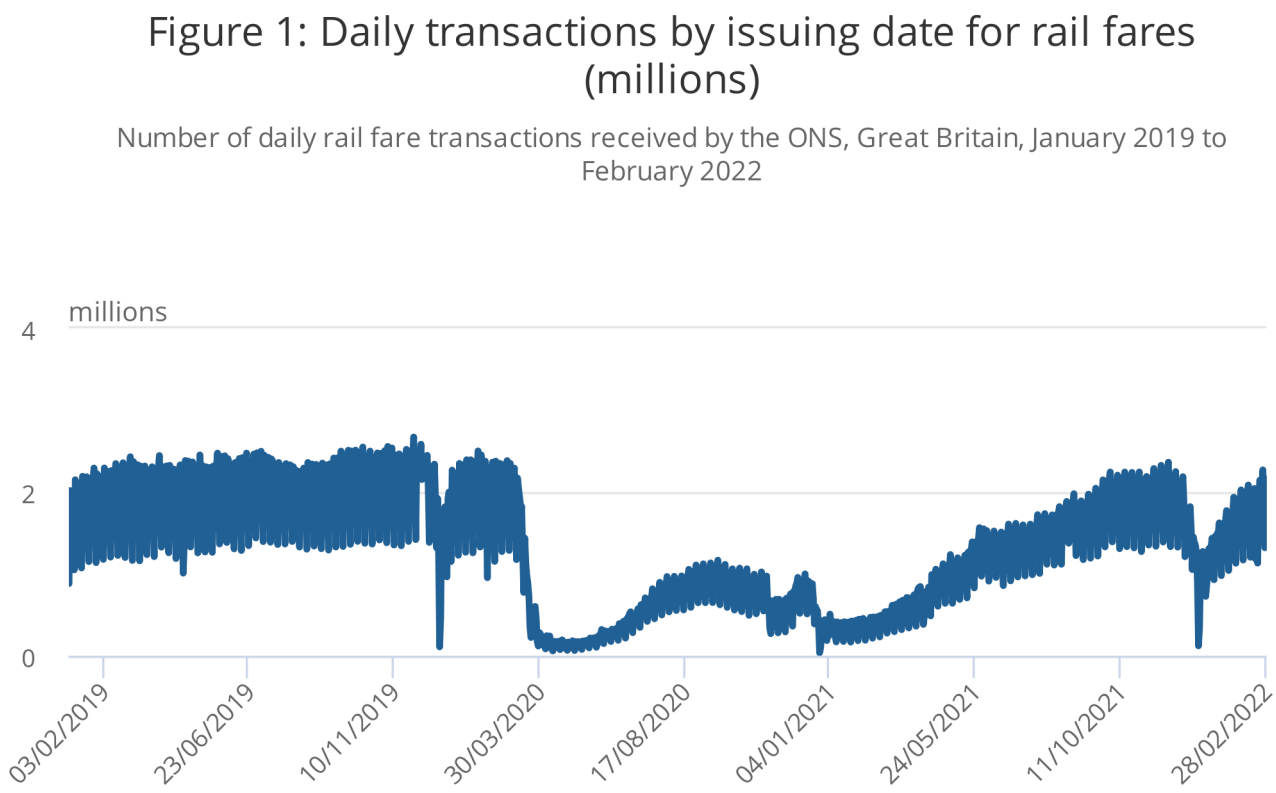
In April 2021, we obtained access to transaction-level sales data for rail fares in Great Britain (GB) dating back to January 2019, sourced from the rail industry's Latest Earnings Networked Nationally Over Night (LENNON) ticket revenue system. It does not retail.

These data are delivered daily and, in general, are extremely timely; we receive 85% of data within one day of the issuing date of the ticket, and 97% of data within seven days.

Since these data are transaction level, we are processing a vast amount of data. We receive approximately 2 million transactions per day, equating to approximately 60 million per month (in a typical year), as shown in Figure 1.

Figure 1: Daily transactions by issuing date for rail fares (millions)

Number of daily rail fare transactions received by the ONS, Great Britain, January 2019 to February 2022



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

The number of daily transactions was substantially reduced during months related to restricted movement during the coronavirus (COVID-19) pandemic, as well as on Christmas day in 2019, 2020 and 2021.

These data are highly informative, with almost 70 variables in total including origin and destination stations, ticket class, ticket name and sales values. These variables are primarily used in calculating an index value but also enable us to perform more detailed analyses on the underlying data. Note that the information in these variables relate to the journey; we do not ask for, nor receive, data that would personally identify passengers.

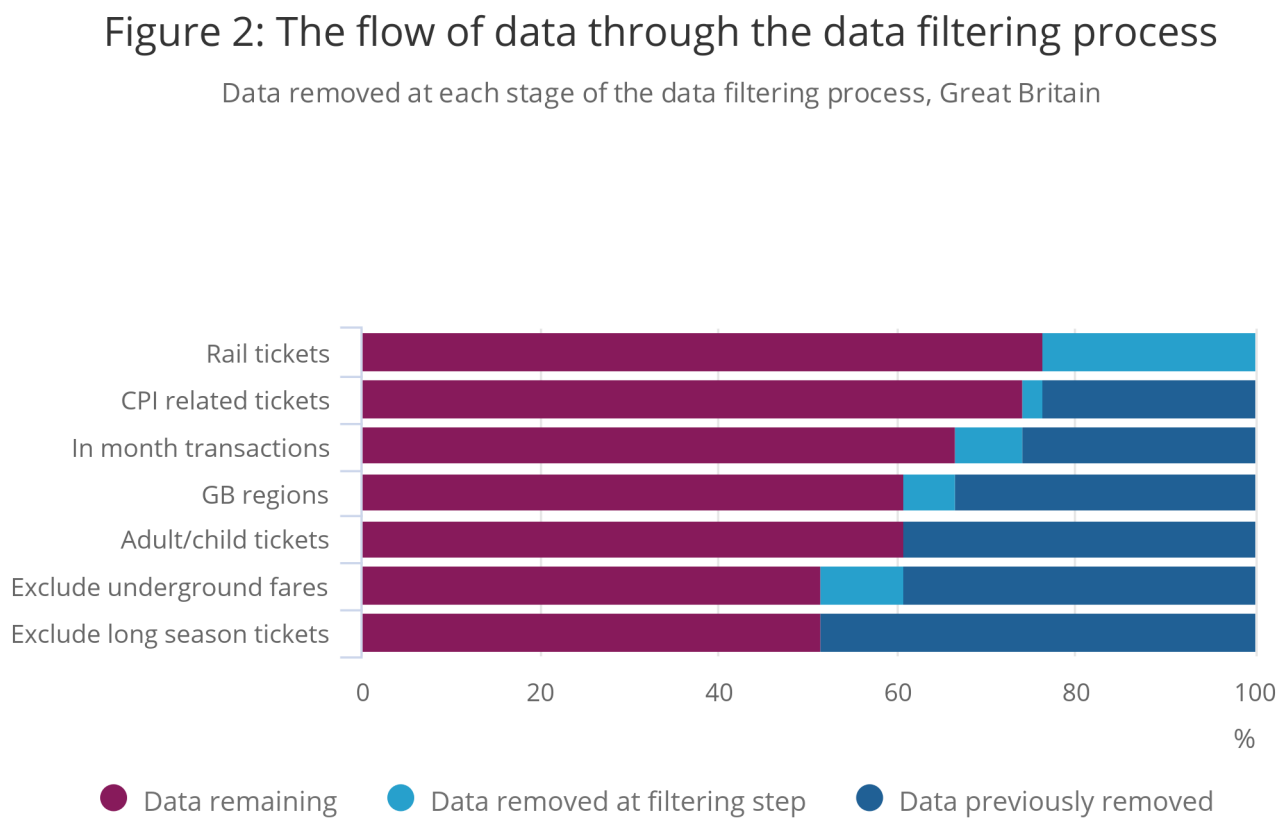
5 . Proposed methodology

Data classification and filtering

As well as data for GB rail fares, these data also include additional transactions that are not relevant to the typical consumer rail fares. These data are filtered so that we can produce the relevant indices. Each bar in Figure 2 relates to the data remaining, data removed at the specified filtering step, and the data that have already been removed in previous filtering steps, expressed as a percentage of the total. For example, when filtering for the Consumer Prices Index (CPI) related tickets, 24% of the data had already been removed as they were not rail tickets (such as car parking and seat reservations), and a further 2% are now removed because they are business transactions, this leaves us with 74% of the data remaining. We also exclude underground fares from these data since these fares belong to a separate area in the Consumer Prices Index including owner occupiers' housing costs (CPIH) and CPI basket. This filters out a further 9% of the dataset.

Figure 2: The flow of data through the data filtering process

Data removed at each stage of the data filtering process, Great Britain



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

Notes:

1. "Adult/child tickets" filters out tickets that have both adults and children on the same ticket, since we are unable to separate the price.

After filtering we are left with 52% of the data, accounting for 65% of expenditure. This means that we will use approximately 30 million transactions per month in calculating consumer price indices for rail fares in GB, in a typical month. Further details regarding the data cleaning carried out prior to producing the analysis in this article were discussed with the Technical Advisory Panel on Consumer Prices in January 2022.

Defining a product

Unlike conventional scanner data, rail fares do not have a barcode or similar unique product identifier that we can use to track price change over time. Instead, we create a product identifier from several variables within the data. This results in a challenging trade-off. Using a single identifier, such as "origin station" could introduce compositional effects into our price indices depending how many tickets of different types are bought. For example, if more peak tickets are bought in a month compared with the previous month this could lead to a disingenuous price increase. However, if our definition is too narrow, we may not be able to find a comparable product in a previous month to compare the price with.

For more information on conventional scanner data, see our Research into the use of scanner data for constructing UK consumer price statistics article.

The product definition that we propose to use is a combination of the following variables:

- origin station (for example, Cardiff Central)
- destination station (for example, London Paddington)
- route (for example, via London)
- product name (for example, standard day return)
- discount type (for example, "senior railcard")
- fare product group (for example, peak or off-peak)

The average transaction price is calculated each month for each product definition and is then tracked over time to produce a price index.

Discounts

Discounts are relatively common within rail fares because of purchasable railcards, which offer groups of the population a discount on rail travel, such as the "16-25 railcard", "senior railcard" and "disabled persons railcard". If these discounts were included in the index, it would be difficult to determine when a genuine price change has occurred as opposed to a change in the composition of travellers. For example, we would expect that more young people would travel in school or university holidays with cheaper tickets, either because of child fares or "16-25 railcards". This would cause the average ticket price to drop over that period despite no price change within the individual ticket prices. Subsequently, we have avoided this discount effect by including the discount type within the product ID. For example, whether the ticket is a standard adult fare, a child fare or whether a railcard has been used.

Refunds

As well as discounts on rail travel, the data also include refunded tickets. These appear as a new transaction with negative sales values. Refunds are far less common than discounts, though they were more prevalent during the coronavirus (COVID-19) pandemic, especially during earlier periods of restricted movement. In these cases, we use a set of variables to link the refunded transaction to the original transaction.

For the purpose of these research indices, we have excluded the refund transactions at the data filtering stage while we perform further work to improve our methodology in this area. However, given that refunds make up less than 2% of transactions in the full dataset, we expect this work to have minimal impact on the resulting index.

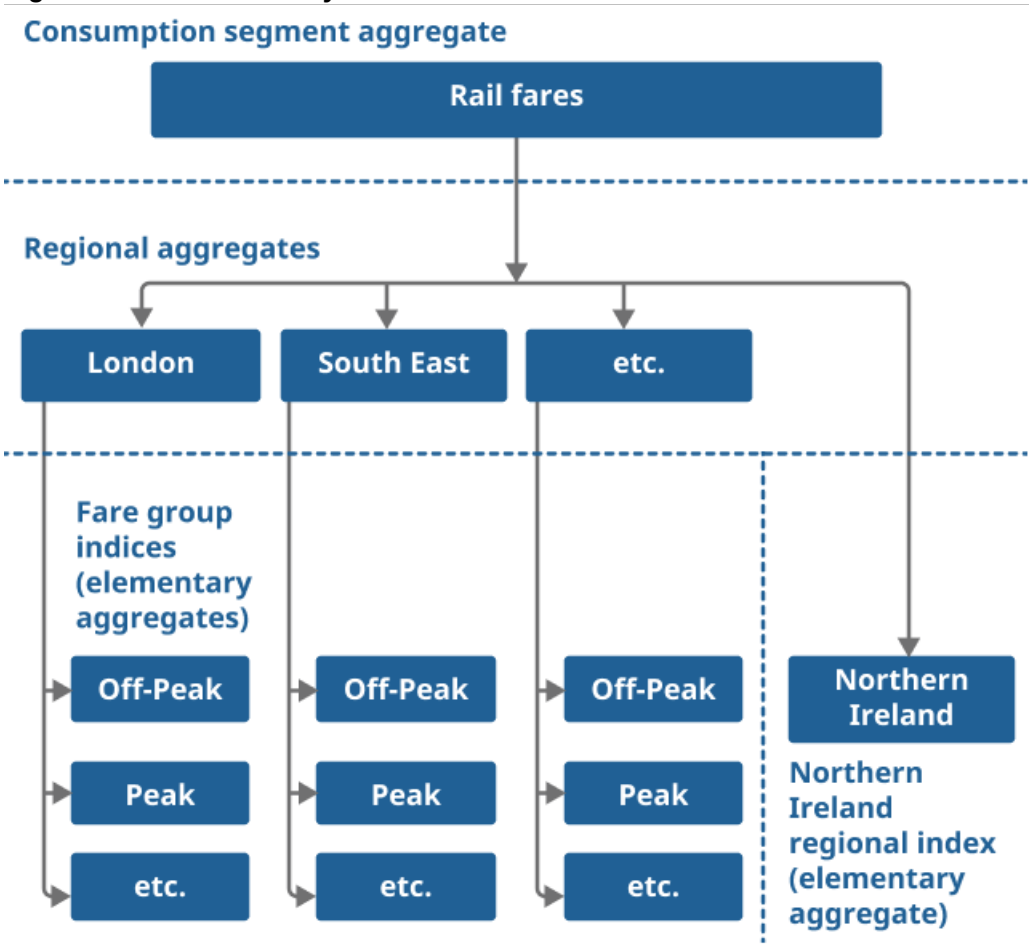
Season tickets

For the research indices presented in this article, we have excluded season tickets that are valid for longer than one month, while we further investigate how to treat their consumption. Especially for season tickets that are valid for longer periods, although the ticket is purchased on a single day, it can still be used for its entire validity period. This raises the question around the date of consumption of the ticket, and whether the price of the ticket should be distributed across the valid period of the ticket. We will be looking to make further methodological improvements surrounding season tickets prior to our next publication of these figures in November 2022.

Index methods

We stratify our indices to a fare product group (for example, advance, peak, off-peak) within each region (Figure 3). We also considered stratifying indices by ticket class (standard or first class) to provide an additional layer of information when interpreting the indices. However, we found the coverage of first class fares to be low so this information is now incorporated into the product ID.

Figure 3: Future hierarchy for UK rail fares index



Source: Office for National Statistics

Our previous work, and corresponding international guidance, has pointed towards multilateral methods being most appropriate for producing elementary aggregate price indices using large, dynamic datasets. For more information, see our New index number methods in consumer price statistics article.

A GEKS-Törnqvist index using a mean splice on the published series with a 25-month window, is used for calculation of these low-level stratum (elementary aggregate) indices. The process for choosing this method is discussed further in our Research and developments in the transformation of UK consumer price statistics: June 2022 article.

Elementary aggregate indices for GB rail fares in this analysis are aggregated based on the previous year (y-1) expenditure shares. For the first year of this analysis, where no historical data are available, they are based on the first year (y). Regional indices are aggregated with the existing index for Northern Ireland (for which Latest Earnings Networked Nationally Over Night do not provide data), using existing weights, to produce a UK rail fares index.

6 . Results: price indices using new data and methods

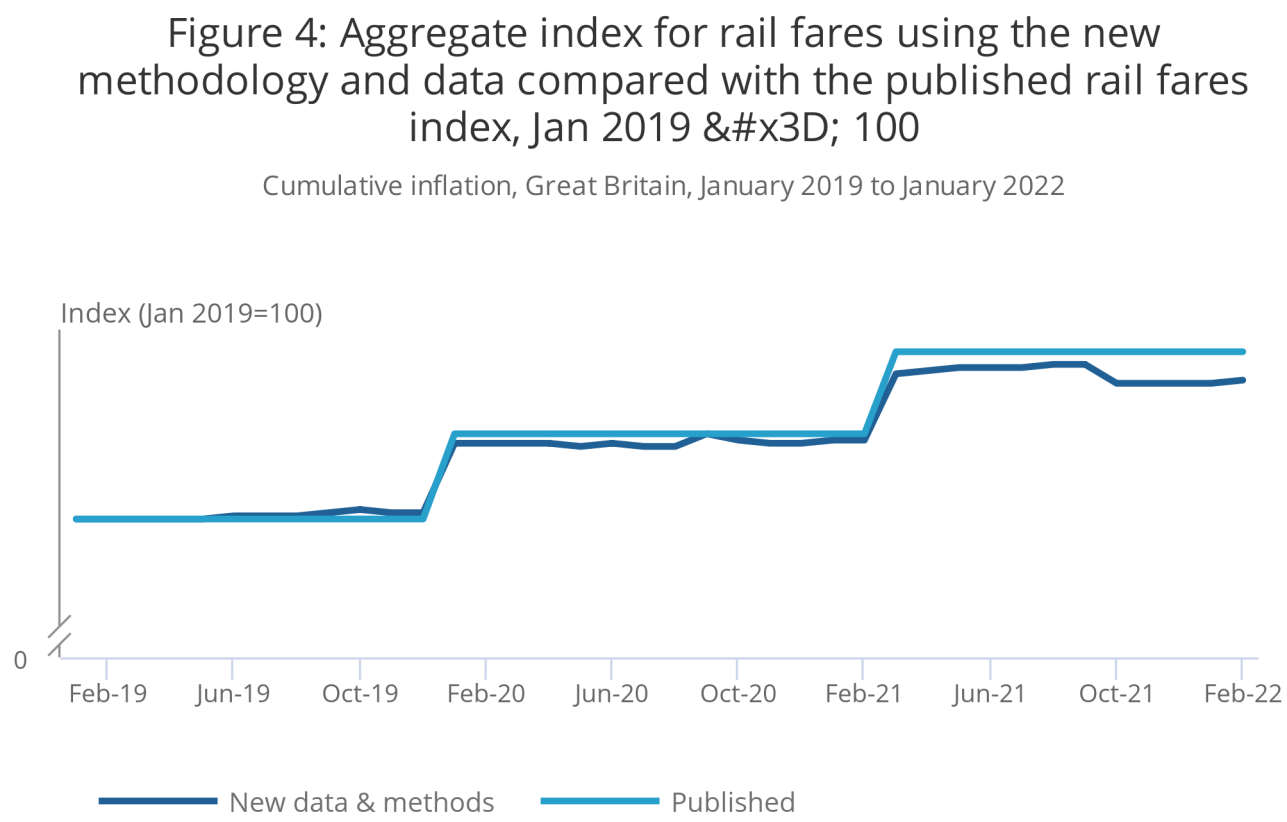
Consumption segment index

Figure 4 shows that the aggregate index based on the new methodology experiences an annual uplift broadly in line with our published index, though our aggregate index shows a slightly lower rate of inflation over the period.

Currently we report cumulative inflation of 5.3% between January 2019 and February 2022, whereas our index produced using new data and methods shows cumulative inflation of 4.4% over this same period. There is also more temporal variation in the index produced using new data and methods. To further explore these variations, we can look at the regional indices and indices by fare product group, as discussed in the following sections.

Figure 4: Aggregate index for rail fares using the new methodology and data compared with the published rail fares index, Jan 2019 = 100

Cumulative inflation, Great Britain, January 2019 to January 2022



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

Regional aggregate indices

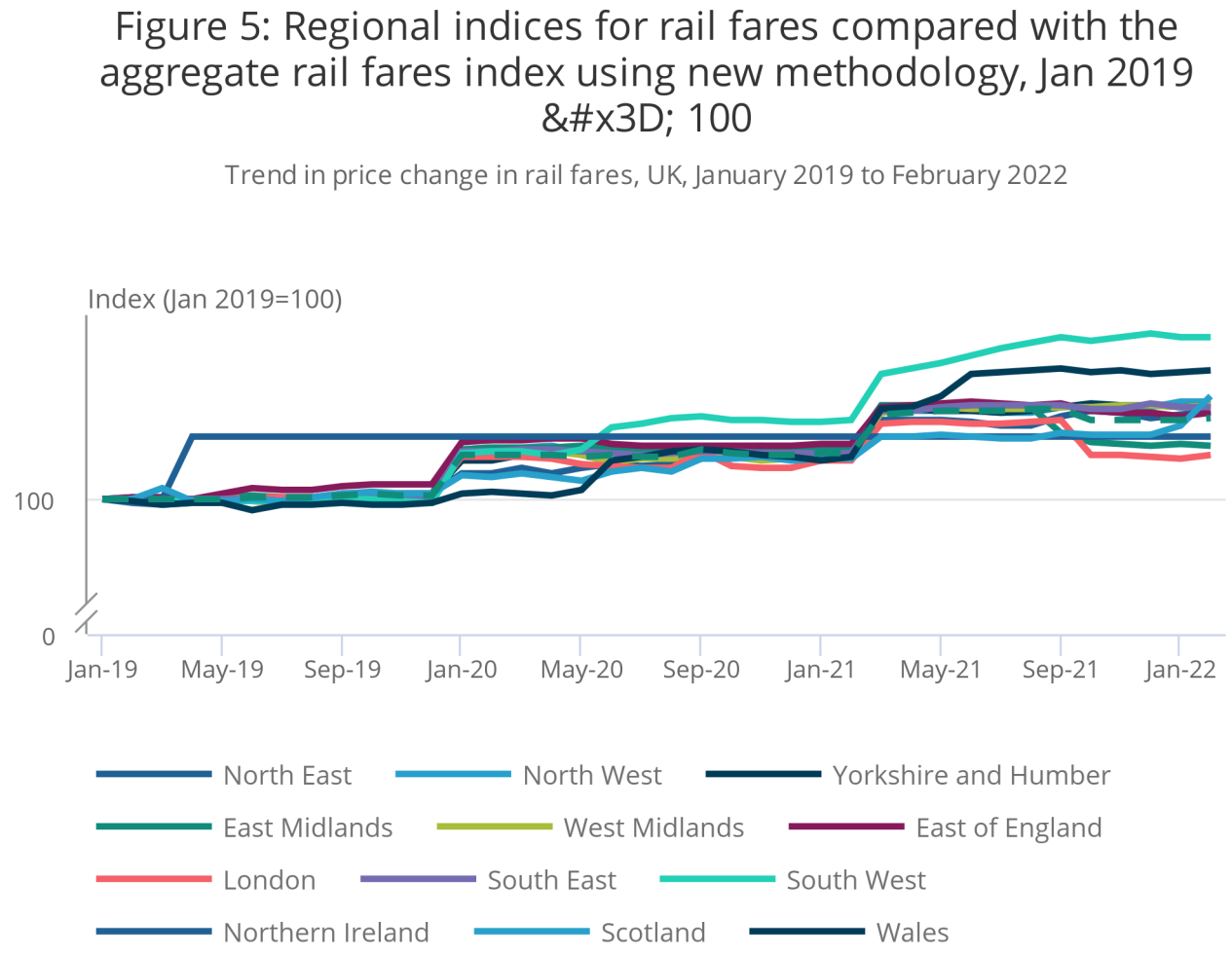
One objective of using alternative data sources in the UK is that they will allow us to produce sub-national price indices more readily. There is a conceptual challenge in defining a region for a rail fares transaction and there are several options. We have chosen to base the region for a rail fares transaction on the region that the origin station resides in (based on postcode). This is because intuitively we would expect this to be the region the ticket was bought in, and therefore typically reflective of households in that region.

There are potential issues when consumers perform ticket splitting. This is where consumers buy multiple tickets to cover a single journey, that would then appear as multiple transactions in our data. Given the information we have available, we believe this is the best approximation for the region of consumption.

Figure 5 shows regional indices for rail fares in UK compared with the aggregate index, where the Great Britain regions are determined based on the origin station of our new data. We use our existing index for Northern Ireland.

Figure 5: Regional indices for rail fares compared with the aggregate rail fares index using new methodology, Jan 2019 = 100

Trend in price change in rail fares, UK, January 2019 to February 2022



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UKOffice for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

While there is some regional variation in rail fares, broadly speaking the regional indices follow a similar trend. There are some exceptions to this, for example we see that the South West is showing the highest level of inflation of all the regions. In this region, rail inflation is 4.5% higher in September 2020 than January 2019, and 9% higher by December 2021. By further inspecting the elementary aggregate indices for the South West, the main driver of this inflation is mostly because of a gradual increase in the price of advance fares (that are unregulated) in this region.

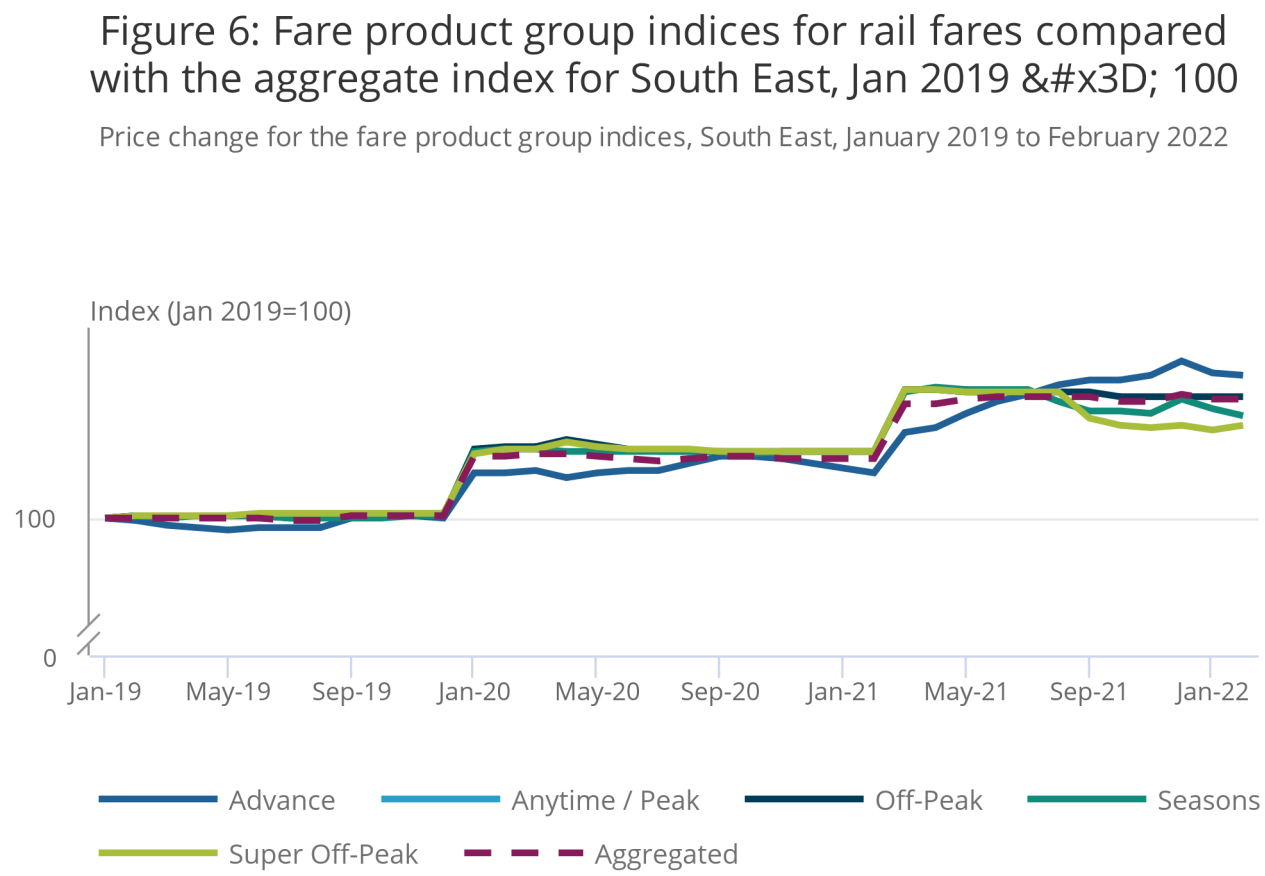
It is clear from these results that some further investigation is needed into the London index, especially since it holds a higher weight in the aggregate index compared with other regions. In particular, we see a small spike in the London index in September 2020 (an increase of 0.9 percentage points) and a decrease of 1.8 percentage points in October 2021 that carries forward. From further analysis, the latter is potentially because of "pay as you go" tickets available on some London journeys (tube excluded), where we are seeing a large drop in sales starting in October 2021.

Elementary aggregate indices

Another aim of alternative data sources is to produce more granular indices that can be used to better understand drivers of inflation. Figure 6 shows indices at the elementary aggregate level for rail fares within a single region, that is, indices for each fare product group within the South East region of the UK as well as the aggregated index for this region.

Figure 6: Fare product group indices for rail fares compared with the aggregate index for South East, Jan 2019 = 100

Price change for the fare product group indices, South East, January 2019 to February 2022



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

In this case, the fare product groups all seem to experience very similar price change. However, we can see those advance tickets, in particular, deviate the most from the aggregated index, experiencing slightly lower levels of inflation in the uplift of 2020 before consistently increasing across 2021. Advance fares are unregulated whereas most other fares are regulated, this might explain why they diverge more from the typical annual uplift in this case.

These lower-level indices give us more insight into the underlying fares that are driving inflation. This helps to explain the impact on the aggregated indices, something that is harder to achieve with our traditional method.

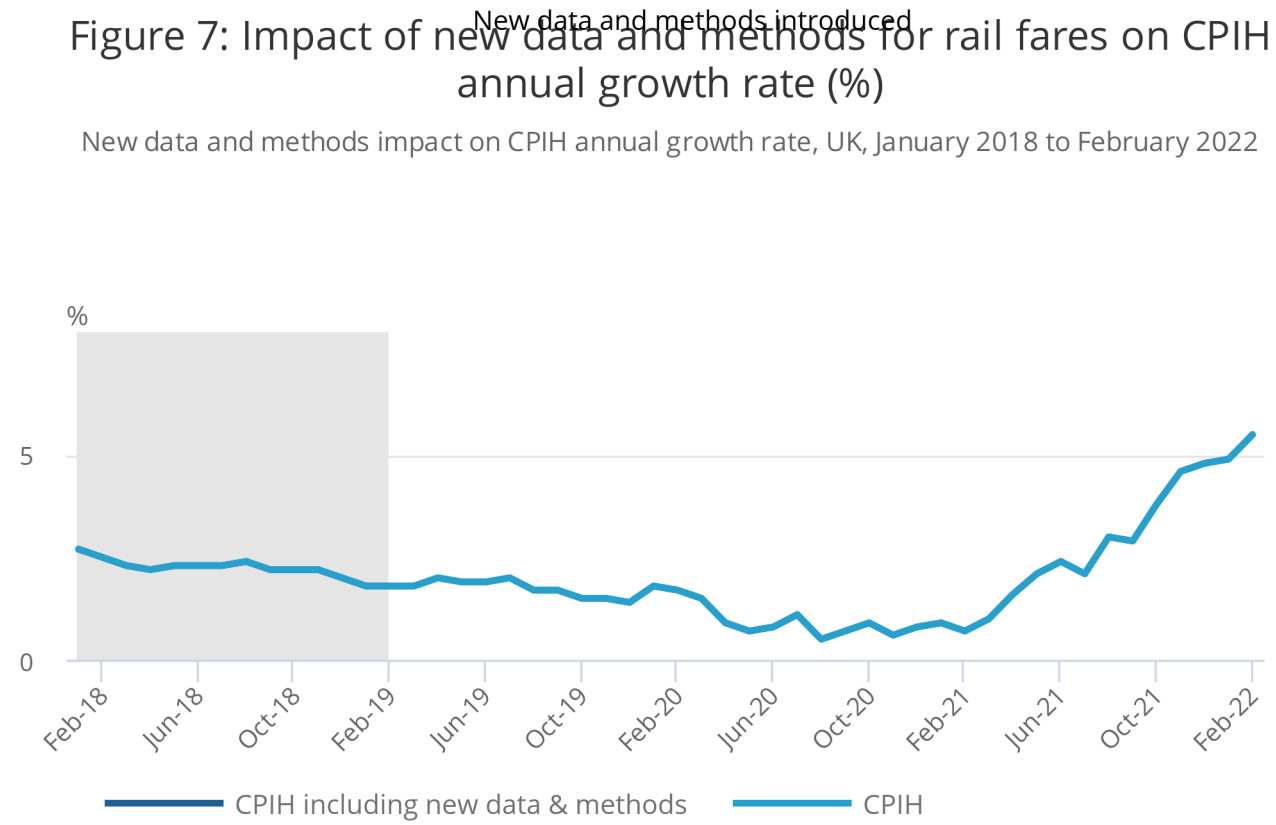
Impact of new data and methods for rail fares on headline consumer price statistics

To show the impact of these new data and methods, had we introduced them sooner, we produce a revised Consumer Prices Index including owner occupiers' housing costs (CPIH) index using the updated index values. This is indicative, and the CPIH and Consumer Prices Index (CPI) will not be revised as we introduce these new data and methods into our published figures from 2023.

The aggregate index for CPIH (Figure 7) has been produced between January 2018 and February 2022, including the new rail fares index from February 2019 onwards. This is so the growth rates in the year of introduction can be seen as well as annual growth in the years following introduction. The new index is aggregated together with published series using the existing annual weights and chain-linking methodology.

Figure 7: Impact of new data and methods for rail fares on CPIH annual growth rate (%)

New data and methods impact on CPIH annual growth rate, UK, January 2018 to February 2022



Source: Office for National Statistics – Using transaction-level rail fares data to transform consumer price statistics, UK

While our index for rail fares is more timely and more granular, enabling us to better understand drivers of inflation, the impact on CPIH as a result of this change is negligible with an average absolute difference of 0 percentage points across the entire period when rounded to one decimal place. There is also no impact at headline on CPI.

Note that since March 2020, there have been a number of unavailable items that have been imputed in some periods based on price movements of the headline index. For this impact analysis, we have not recalculated these imputations because of the complexity of their calculations, but we would expect the impact of recalculating imputations to be negligible. This is based on the minimal impact of these new data and methods on the headline indices, and because imputations were designed to have a negligible impact on the headline rate. For more information see our Coronavirus and the effects on UK prices article.

7 . Future developments

Following our publication of the final impacts in November 2022, a decision will be made on whether we move these new data and methods into use in live production of the Consumer Prices Index including owner occupiers' housing costs (CPIH) and the Consumer Prices Index (CPI).If we are satisfied our data, methods and systems are ready for live monthly production of these indices, the first time they would be introduced is in the figures for February 2023, published in March 2023. The CPIH and CPI will not be revised.

Our broader plans to transform UK consumer price statistics by including new improved data sources and developing our methods and systems for production from 2023, are discussed in Transformation of consumer price statistics: April 2022 article.

8 . Related links

Research and developments in the transformation of UK consumer price statistics: June 2022
Article | Released 28 June 2022
Research to modernise the measurement of consumer price inflation in the UK: fourth in a series of biannual articles to update users.

Using Auto Trader car listings data to transform consumer price statistics, UK: June 2022
Article | Released 28 June 2022
Car listings data will improve measurement of consumer prices from 2023. This article details new methods and provides research indices using these data.

Transformation of consumer price statistics: April 2022
Article | Released 27 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.

Consumer price inflation, UK: May 2022
Bulletin | Released 22 June 2022
Price indices, percentage changes, and weights for the different measures of consumer price inflation.

Consumer Prices Indices Technical Manual, 2019
Methodology | Released 18 September 2019
This technical manual is a reference tool for anyone wanting to understand how measures of consumer price inflation and associated indices are compiled.