



Rail Fares Index: Quality and Methodology Report

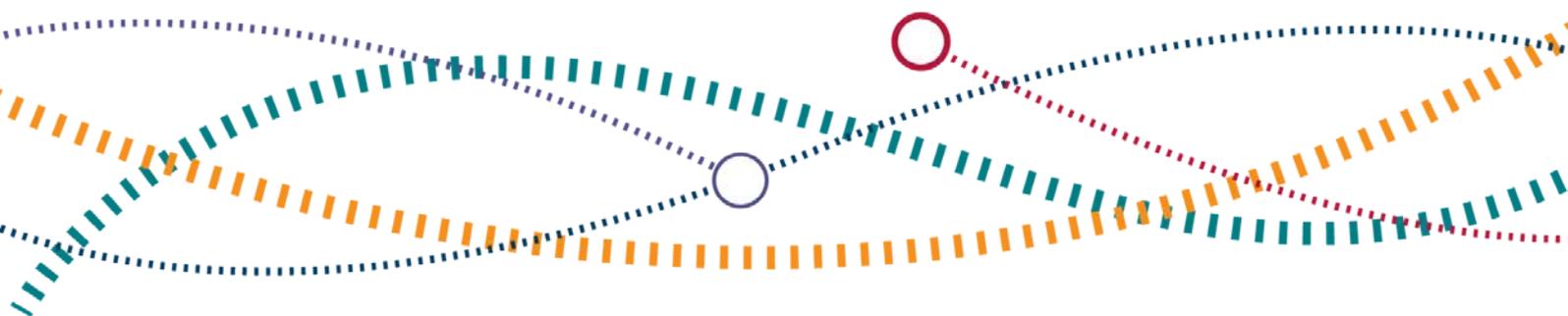
17 June 2021

Contact information:

Responsible statistician: P Moran

Public enquiries: rail.stats@orr.gov.uk

Media Enquiries: Tel: 020 7282 2094



Contents

Introduction	2
<hr/>	
Data sources, methodology and definitions	2
<hr/>	
Historical background	8
<hr/>	
Relevance	8
<hr/>	
Accuracy and reliability	11
<hr/>	
Timeliness and punctuality	13
<hr/>	
Accessibility and clarity	14
<hr/>	
Coherence and comparability	15

Introduction

This is a report on the quality of the annual Rail Fares Index statistical release and associated data tables. It helps users to understand the quality of our statistics, and also ensures we are compliant with the three quality principles in [the Code of Practice for Official Statistics](#) - Q1: Suitable data sources, Q2: Sound methods, and Q3: Assured quality. This report also provides information on the methodology and data sources used to produce the statistics.

This report covers the following areas:

- Data sources, methodology and definitions – detail on the various data sources, methodology used to compile the statistics, and definitions;
- Historic background – a background to rail fares index statistics and details of any changes throughout the time series;
- Relevance to users – the users of the statistics, and our engagement;
- Accuracy and reliability – the accuracy, data coverage and quality assurance of the statistics;
- Timeliness and punctuality – our timescales for the production and publication of the statistics;
- Accessibility and clarity – the format of our statistics and where they can be found;
- Coherence and comparability – similar statistics published elsewhere and the degree in which the statistics can be compared over time.

The rail fares index is an annual release and provides a measure of the change in prices charged by train operators to rail passengers. New fares are normally introduced in January of each year, although in 2021 the implementation of the new fares was delayed until 1 March 2021.

Data sources, methodology and definitions

The data contained within the release and the data tables are sourced from:

- **Latest Earnings Networked Nationally Over-Night (LENNON)** provides revenue data for each origin-destination flow in Great Britain. This dataset uses the post-allocation dataset within LENNON that distributes passenger revenue to train operators who operate on all, or part, of the origin-destination flow. The revenue data is used to determine the weights that will be applied to calculating the average price change and are based on total revenue in the twelve months prior to the annual rail fare change.
- **Rail Delivery Group (RDG)** provide a [fares data feed](#) containing the price for every ticket sold through the LENNON system. This is used to calculate the price change between the two reference periods.
- **Office for National Statistics (ONS)** 'All Items' RPI data is used to compare the average change in rail fares with the average change in other goods and services. Whilst the RPI is no longer the headline measure of inflation, it is used as a comparator due to its relationship with rail fares. The annual change in regulated fares continues to be based on the 12-month change in the July 'All Items' RPI that precedes the annual fare change. We also use the March 2021 'All Items' RPI in our data tables to compare how the overall change in rail fares compares with the price change in other goods and services over the same period.

Methodology

Weights

The expenditure weights (%) used to calculate the average change in rail fares are the proportions of market share in terms of revenue for the twelve months preceding the annual fare change. This usually covers revenue between January and December as the annual fare change has, historically, taken place in January. As this year's annual fare change was effective from 1 March 2021, revenue weights were aggregated based on revenue between March 2020 and February 2021 inclusive.

To calculate the index weights, an extract is taken from the rail industry's ticketing and revenue database (LENNON) for the twelve months preceding the annual fare change.

This provides the revenue taken for each available fare on every flow (origin-destination pair) and usually contains more than 10 million records.

The revenue weights from each of these available flows are aggregated to create weights for each sector, class of travel, ticket type and regulated/unregulated fare (these categorisations are defined in the Definitions section). They are then used to calculate the weighted average price change.

Due to the high volume of refunds processed since March 2020, we have adjusted our weights methodology to take this into account. Within LENNON, there are two broad entries for refunds: refunds by flow (ordinary) and refunds by flow (season). Refunds by flow (ordinary) covers any non-season tickets (Advance, Anytime, Off Peak, Super Off Peak, Other) and, by default, revenue from these refunds were mapped to the Anytime ticket type. The effect of this on the index was negligible as refunds typically account for around 2% of total fare revenue.

For this year's release, we have re-allocated refunds between each individual ticket type based on the distribution of ticket sales data. For example, Off Peak tickets account for 40.4% of all non-season ticket revenue (33.3% of total revenue) so 40.4% of the refunds by flow (ordinary) revenue has been assigned to Off Peak tickets.

Prices

To calculate the index itself, the price of each of these fares is taken from a combination of LENNON and an [RDG open data feed](#). These datasets provide us with a price following the fare change and the price of that same product from the previous fare change. Not all flows/fares have prices in both reference points. This is because either:

- The flow and/or fare were introduced in the intervening period between the two fare changes so there is no price from the previous fare change
- The flow and/or fare were discontinued in the intervening period between the two fare changes so there is no price for the most recent fare change

Currently, if either of these exceptions apply, then that record is omitted from the index calculation (i.e. only flows that have a price in both the most recent fare change and the previous fare change are included). Each record in the final dataset is then assigned to:

- a sector based on the train operator running the service;
- a class based on whether it is first class or standard class travel;
- a ticket type based on the specific product code and description; and

- a regulated or unregulated fare based on the regulated fares basket.

The final indices are produced by calculating the price change between the most recent fare change and the previous fare change, weighted by earnings.

Comparators

Alongside our rail fares index we publish two comparators:

Revenue per journey is used as a comparator as it can reflect changes in customer behaviour when passengers switch to cheaper tickets. The fares index does not capture this phenomenon, other than in the expenditure weighting for those products in the following year. Furthermore, revenue per journey reflects changes in usage as well as the cost of fares and can be affected by different growth rates on different parts of the network as some sectors have a higher rate per journey than others.

The revenue per journey figures are based on the passenger revenue and passenger journeys statistics published within our [passenger rail usage statistics](#). It should be noted that the sector mappings used within the rail fares index and passenger rail usage statistics differ. This is explained further in the Definitions section of this report.

Please note that revenue per journey figures at sector level were unavailable for publication with the 2021 release. Further work is required to improve the methodology for allocating non-LENNON usage data to sectors. Therefore, [Table 7182 - Average change in fares by ticket type](#) only includes revenue per journey data at Great Britain level. We will re-instate the sector level revenue per journey data in the 2022 release.

12-month change in the 'All items' RPI is used as a comparator as it measures the change in price of a basket of goods and services. This gives us an indication of how the price of rail fares has changed in comparison to other goods and services. The RPI figure used relates to the month that the fare change came into effect. Therefore, the most recent fare change uses the March 2021 'All items' RPI.

We continue to use RPI as a comparator because of the relationship between RPI and rail fares. Current government policy caps the annual change in regulated rail fares based on the July 'All items' RPI. We took the decision to continue using RPI as a comparator to avoid using two measures of inflation (CPI and RPI) which could potentially confuse users.

The RPI is also used to calculate the change in fares in 'real terms'. Real terms change illustrates how the price of rail fares has changed when you take into account inflation. For example, if the real terms change in rail fares over a given time is 5%, this means that rail fares have increased by 5% more than the general inflation rate.

Definitions

Each mainline operator is assigned to one sector, unlike our [passenger rail usage statistics](#) where operators can cover more than one sector.

- **London and South East:** c2c, Chiltern, Govia Thameslink Railway, Greater Anglia, London Overground, Southeastern, South Western Railway, TfL Rail.
- **Long distance:** Avanti West Coast, Cross Country, East Midlands Trains, Hull Trains, Grand Central, Great Western Railway, London North Eastern Railway.
- **Regional:** Caledonian Sleeper, Merseyrail, Northern Trains, Scotrail, TfW Rail, Transpennine Express, West Midlands Trains.

The ticket types are broken down into the following categories:

- **Anytime:** fully flexible tickets that can be used on most trains and at most times. These are usually more expensive.
- **Advance:** single, one-way tickets for a specific train. These are usually cheaper than other ticket types.
- **Off Peak:** cheaper than anytime fares but cannot be used at busy times of the day.
- **Super Off Peak:** cheaper than off-peak fares but subject to similar restrictions (previously known as Super Saver tickets).
- **Seasons:** allows unlimited travel between two locations for a specified period (from a week up to a year).
- **Other:** includes promotional fares, rover tickets, group tickets and package tickets.

Fares are also broken down into regulated and unregulated fares:

- **Regulated fares:** standard class fares including saver returns, standard returns, off-peak fares between major cities and season tickets for most journeys. Price changes for regulated fares are capped by government based on the annual change in the RPI in July of each year.
- **Unregulated fares:** includes first class, advance purchase and saver tickets. Train operators are free to determine these fares, though these can also be capped in certain circumstances.

Impact of the coronavirus (COVID-19) pandemic

In response to the pandemic, the government announced in December 2020 that the [usual annual January change to rail fares would be delayed until March 2021](#), freezing prices at 2020 levels during January and February 2021.

At this point in time, it is not possible to quantify the impact of the price freeze on the average change in fares across the whole of 2021. Due to the ongoing uncertainty over loosening of restrictions, it is difficult to predict to what extent passenger rail travel will recover during the remaining months of 2021.

In addition, ticket purchasing patterns have changed because of the pandemic and it remains to be seen whether this is a permanent shift or not. Over recent years, revenue market share for ticket types has remained relatively stable. However, this year has seen a reduction in the market share for longer-term season tickets with passengers opting for on the day travel or shorter-term season tickets.

Historical background

The rail fares indices provide a measure of the change in prices charged by train operators for the following categories:

- Average change in price of rail fares by ticket type; and
- Average change in price of rail fares by regulated and unregulated tickets

The average change in rail fares by ticket type dates back to 2004 and is disaggregated by sector (London and South East, Long Distance and Regional (includes Scotland)) and by ticket type (Advance, Anytime, Off Peak, Season, Super Off Peak and Other).

The average change in price of rail fares by regulated and unregulated tickets dates back to 1995 and is disaggregated by sector, class, and regulated/unregulated fare.

Both statistics are calculated from transaction data held in the LENNON system and the RDG open data feed on fares.

The index shows the change in prices between the current fare (at the most recent annual price change) and the previous fare. This usually means that prices are compared from one January to the next, as January is normally the month when the fare change is implemented. In 2021 this was delayed until March, so comparisons are based on the March 2021 fare compared with the January 2020 fare. This should not affect the index calculation as prices only tend to change when the annual fare change comes into effect (i.e. the price would remain the same between January 2020 and February 2021).

Sampling

In 2014 and 2015, we calculated the fares index based on a sample of fares rather than a population of matched prices (i.e. exact same fare present in both the current and previous years). The non-sampling methodology generated an annual dataset of between 1.6m and 3.5m records. The production of the index from this annual dataset was resource intensive and time consuming. ONS recommended that we switch to probability sampling.

This was implemented in 2014, using a sample size of almost 100,000 (3% of dataset)

In 2016, we reverted to using the population dataset. This change was made because RDG permitted access to their fares data feed, which provided a more complete and robust reference file of fares than was used previously. This has reduced the number of

erroneous records in the final dataset and we felt that the subsequent reduction in the QA burden warranted re-introduction of the population dataset for calculating the index.

Advance fares

The indices for advance fares are calculated using a different methodology to other types of fares. There is a greater product turnover in advance fares than for other fare types which means that calculating a consistent matched price index can be difficult. ONS recommended that rather than calculate the change in advance fares through the matched price method, we should calculate an overall average price covering all advance ticket sales for each origin destination pair. This has been implemented from 2014 onwards.

Price change thresholds

Between 2004 and 2013, price changes between -40% and +60% were included within the final index calculations. Due to the size of the dataset, it was not possible to check the validity of all price changes and the flows with the largest weightings were checked and validated. ONS recommended that switching to a sample would enable us to quality assure a greater number of fares so all fares outside the price change limits of -20% and +20% have been checked for their validity from 2014 onwards.

Since reverting back to using a population dataset, we continue to quality assure price changes outside of the -20% to +20% bounds. This is done by extracting a dataset containing all price changes outside of these bounds. These are sorted by revenue and high-revenue flows (e.g. > £50,000) that are likely to have a noticeable/material impact on the index calculation are checked, verified and corrected so they can be included in the final index calculation. In addition, those that remain outside of these limits are excluded from the final index calculation as these are assumed to be errors.

Excluded revenue

Prior to 2014, only the revenue from those flows that had a matched price in the current and previous year were included as weighting in the index calculation. From 2014, expenditure on those flows that have been excluded at the various stages of compiling the index (e.g. no price record in one of the two reference years, price change exceeding the allowable thresholds) have been reintroduced to compile the aggregated sector/national/ticket category indices. This means that revenue included as weighting has increased from between 85% and 90% prior to 2014, to in excess of 95%.

Relevance

The degree to which the statistical product meets the user needs in both coverage and content.

Statistics on the changes in rail fares provides passengers and policy makers with important information on the cost of travel by rail in the wider context of cost of living.

Statistics on the changes in rail fares for train operators and specific origin destination flows are not published as they are deemed to be commercially sensitive.

Our last [user survey](#) took place from mid-January to mid-April 2020. The aim of the survey was to gather feedback on the data portal; this includes statistical releases, data tables and other supplementary material. There were 42 responses to the survey. We created an [implementation plan](#) following the 2020 user survey.

More detailed information on users of our statistics and meeting the needs of users is available on our [user engagement webpage](#).

Accuracy and reliability

The proximity between an estimate and the unknown true value.

Revenue weights data are all derived from the LENNON system. This is primarily an accounting tool and therefore faces limitations when being used for statistical reporting. With all large data sources there may be input errors though these are more likely to occur in journey data, rather than revenue data. Due to the size and complexity of the dataset we are unable to validate each and every entry.

The pandemic generated further uncertainty around the use of season tickets and refunds. An explanation of how we have adjusted the methodology to take this into account is explained in the Data sources, methodology and definitions section.

Data coverage

The indices include franchised and open access operators operating on the mainline rail network with the exception of Heathrow Express as their revenue data are not included in LENNON. Eurostar are also not included.

It does not include fares on London Underground, other light rail and metro systems, charter services or heritage operators.

Quality assurance

Due to the large and complex nature of the datasets used to derive the rail fares index, we use Python to process the data. Within the Python code, we run a series of checks on the data, the outputs of which are then reviewed for accuracy and reliability. These include:

- File containing products that have not been assigned to a ticket category
- File containing flows that have no price information in one or both reference periods
- File containing flows where the price has either fallen by more than 20% or increased by more than 20%

Whilst it is not possible to review every record due to the size of the datasets, these internal quality assurance processes provide us with an indication of whether they are any flaws in the process itself, and provide us with the opportunity to correct any potential discrepancies.

In addition to our own checks, any issues arising from our assurance processes may be escalated with the Rail Delivery Group or train operators directly. Explanations from the data suppliers regarding data anomalies are included within our commentary to explain the data and trends.

These data are then prepared for publication. The process includes quality assuring the tables and charts produced and providing supporting commentary regarding the key trends, methodology and quality measures. These reports are subject to peer review.

The final stage of the quality assurance process is a sign off by the statistics Head of Profession confirming the data meets the quality standards and are fit for publication.

Revisions policy

Our statement on [orderly release and revisions policy](#) outlines the revision policy we follow. Details of any revisions are available in the [revisions log](#). Further information on revisions and data series breaks can also be found in the data tables.

Timeliness and punctuality

Timeliness refers to the time gap between publication and the reference period.
Punctuality refers to the gap between planned and actual publication dates.

We aim to publish these statistics as soon as possible after the revenue data and new fares data is available. Rail fares index data is typically available on the data portal approximately three months after the fare change has come into effect.

The [publication schedule](#) available on the data portal outlines the publication dates for National Statistics quarterly and annual statistical releases and other official statistics up to 12 months in advance.

We are committed to releasing statistics in an open and transparent manner that promotes confidence.

Accessibility and clarity

Accessibility is the ease with which users are able to access the data, also reflecting the format in which the data are available and the availability of supporting information. Clarity refers to the quality and sufficiency of the metadata, illustrations and accompanying advice.

All rail statistics data tables can be accessed free of charge on the [Data Portal](#). Commentary about the statistics and trends are provided in the statistical releases.

The rail fares index tables currently published on the data portal are:

Rail fares

- Average change in fares by regulated and unregulated tickets (annual) - Table 7180
- Average change in fares by ticket type (annual) – Table 7182

Rail fares, and passenger revenue data, are viewed as commercially sensitive. Therefore, lower levels of disaggregation are not possible without permission from RDG and train operators.

Coherence and comparability

Coherence is the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar. Comparability is the degree to which data can be compared over time and domain.

Other related data

Passenger revenue statistics are published on the [passenger rail usage](#) page of the data portal.

Rail industry income and expenditure statistics are published on the [rail industry finance](#) page of the data portal.

[Data feeds including fares](#) can be accessed from RDG.

ONS publish [inflation time series on rail fares](#) though the scope and methodology differ from those used within our rail fares index release.

The Department for Transport (DfT) also publishes [rail statistics](#).

For more information on impacts of the pandemic, please see:

- [Transport use during the COVID-19 pandemic \(DfT\)](#)
- [All Change? Travel tracker \(DfT\)](#)
- [Coronavirus and the social impacts on Great Britain \(ONS\)](#)
- [Weekly travel during COVID-19 survey \(Transport Focus\)](#)
- [Public transport journeys by type of transport \(Transport for London\)](#)

For more information on comparability of data over time see Historical background section above which describes changes to the data over time.



© Crown copyright 2021

This publication is licensed under the terms of the [Open Government Licence v3.0](#) except where otherwise stated.

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available on the [data portal](#)

Any media enquiries regarding this publication should be sent to us at [orr.gov.uk/contact-us](#).

