



Rail Fares Index



Quality & Methodology Report

Release date: May 2020

Contents



Introduction	3
Methodology	4
Historical background	6
Relevance	9
Accuracy and reliability	10
Timeliness and punctuality	11
Accessibility and clarity	12
Coherence and comparability	13

Introduction

This is a report on the quality and methodology of the Rail Fares Index statistical release and data tables. It helps users to understand the quality of our statistics, and also ensures ORR is compliant with the Quality pillar, and supporting principles, of the [Code of Practice for Official Statistics](#).

The report covers the following areas:

- **Methodology** – detail on the various data sources and methodology used to compile the statistics;
- **Historic background** – a background to each statistic and details of changes throughout the time series;
- **Relevance of the data** – the users of the statistics, and our engagement;
- **Accuracy and reliability** – the accuracy of each statistic;
- **Timeliness and punctuality** – our timescales for the production, quality assurance and publication of each statistic;
- **Accessibility and clarity** – the format of our statistics and where they can be found;
- **Coherence and comparability** – comparisons to similar statistics published elsewhere.

The Rail Fares Index is an annual release and provides a measure of the change in prices charged by train operating companies to rail passengers, which are normally introduced in January of each year.

ORR publishes quality and methodology reports for other ORR statistical releases which can be found on the [ORR data portal](#).

Methodology

The data contained within this release are sourced from Latest Earnings Nationally Networked Over Night (LENNON) system, Rail Delivery Group (RDG) (formerly ATOC) and the Office for National Statistics (ONS), collated and processed using Python and stored in a secure data warehouse maintained by ORR. The statistics are subject to an extensive quality assurance process, including a suite of validation checks to ensure the information meets the required specification and is in line with previous trends.

These statistics 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 statistics meets the quality standards and are fit for onward use.

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 calendar year preceding the publication of the statistics (e.g. the January 2015 index weightings are based on expenditure during 2014).

To calculate the index weights, an extract is taken from the rail industry's ticketing & revenue database (LENNON) for the calendar year, detailing the revenue taken for each available fare on every flow (origin-destination pair). Each year, this dataset contains in excess of 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 section Definitions). These are then used to calculate the weighted average price change.

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 for January in the current year (Jan_x) and a price from January in the previous year (Jan_{x-1}). Not all flows/fares have prices for Jan_x and Jan_{x-1} because:

- 1) The flow and/or fare were introduced in the calendar year so there is no price for Jan_{x-1}
- 2) The flow and/or fare were discontinued in the calendar year so there is no price for Jan_x

Currently, if either 1) or 2) apply, then that particular record is omitted from the index calculation (i.e. only those flows that have a price in both Jan_x and Jan_{x-1} are included). Each record within the final dataset is assigned to the following;

- a sector based on the train operating company (TOC) 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 indices are produced by calculating the price change between Jan_x and Jan_{x-1} weighted by earnings.

Definitions

The sectors are defined as follows:

- 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.

These differ from sectors in the [passenger rail usage statistics](#) as these disaggregate some operators across multiple sectors (e.g. Great Western Railway operate services in all three sectors in the passenger rail usage statistics).

The classes are broken down into:

- First Class; and
- Standard Class

The ticket types are broken down into the following categories:

- Anytime: can be purchased right up to departure but can be used on any train without restrictions on time or day;

- Advance: restricted fares that can be purchased up to twelve weeks before travel but are train-specific;
- Off Peak: restricted fares for travel during periods when train services are generally less busy. They can be bought right up till departure but carry restrictions on days/times of travel;
- Super Off Peak: cheaper than off-peak fares but subject to similar restrictions (previously known as Super Saver tickets);
- Seasons: All season tickets lasting one week or longer; and
- Other: Promotional fares and those not assigned to the above categories.

The regulated/unregulated fares are broken down into:

- Regulated: Saver returns, standard returns, season tickets and standard singles/returns within London Travelcard zones; and
- Unregulated: First class travel, advanced tickets, tickets which include a non-rail element, and some saver and weekly season tickets.

Historical background

The rail fares indices provide a measure of the change in prices charged by train operating companies (TOCs) to rail passengers 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 price of rail fares by ticket type shows statistics back to January 2004 and is disaggregated by sector (London & South East, Long Distance and Regional (includes Scotland)) and 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 shows statistics back to January 1995 and is disaggregated by sector, class and regulated/unregulated fare.

Both statistics are calculated from transaction data held in the Latest Earnings Nationally Networked Over Night (LENNON) system, the rail industry's ticketing and revenue database, and the RDG open data feed on fares.

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

The index shows the change in prices from one January to the next as, generally, new fare prices are introduced at the beginning of January. The average change in fares is presented alongside the January all items Retail Price Index (RPI). The all items RPI measures the changes in the cost of a representative 'basket' of goods and services bought by consumers within the UK. Presenting these measures alongside one another illustrates how the average change in rail fares compares with the average change in the cost of goods and services. More information on the RPI and other measures of inflation is available on the Office for National Statistics website under [Inflation and price indices](#).

ORR continues 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.

In 2014, ORR undertook a review of the rail fares index, which was carried out by the Office for National Statistics (ONS) Methodology Advisory Service. They made a number of recommendations which were implemented in time for the 2014 index calculation. A summary of the changes we made are below. The full report can be found on the ORR data portal titled [Review of the rail fares index](#).

Sampling

In 2014 and 2015, ORR 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 in excess of 97,000, approximately 3% of the population dataset.

In 2016, ORR reverted back 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 ORR 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, ORR should calculate an overall average price covering all advance ticket sales for each origin destination pair. This has been implemented in the index 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 ORR 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 both within the industry and in comparison to the wider cost of living.

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

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

Accuracy and reliability

The proximity between an estimate and the unknown true value.

The rail fares index is calculated from transaction data held in the LENNON database, the rail industry's ticketing and revenue database. The index has been constructed so that, as far as possible, it covers the cost of travel only. This is done by excluding fares that include 'extras' so as not to distort the index, such as the following:

- Additional services such as multi modal tickets for urban areas, bus and car parking tickets, entrance fees to attractions etc
- Short-term temporary fares/promotions;
- Passengers 'switching' ticket types following the introduction/deletion of certain tickets are not taken into account immediately; and
- Flows for which we were unable to find price information for either of the two reference years, for example a ticket type that is introduced after the first reference date.

London Travelcard purchases are included in the index because such tickets are important in the earnings of train operators and purchases by rail passengers. In addition, TOCs influence price changes associated with these tickets.

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.

The statistics contained within the Rail Fares Index are published annually on the ORR data portal approximately four months after the new fares for the year take effect during the first week in January.

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

These publication dates are determined by availability of the information and are the earliest possible dates which we can publish the information.

More detailed information on timeliness and effectiveness of the statistical output is available on our [UKSA Code of Practice page](#).

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 data tables can be accessed on the [ORR Data Portal](#) free of charge.

The procedures and policy used to ensure sound confidentiality, security and transparent practices.

ORR is fully compliant with the Statistics and Registration Service Act 2008 and the pillars and principles of the Code of Practice for Official Statistics.

The Rail Fares Index is calculated from transaction data held in the LENNON database. These are not disaggregated by TOC as it is commercially sensitive information. The statistics are stored in a data warehouse (ORRbit). Only selected members of staff in ORR have access to the data warehouse and access is password protected. The data provision and storage processes have been independently assessed by external consultants Amour Group to ensure they are secure. Internal and external IT vendors also conduct periodic assessments of our systems

ORR has systems and processes in place to safeguard personal identities/details and commercially restricted information. The statistics are internally validated by two teams before it is granted a 'publication' status and viewable on the data portal. Connections to remotely hosted databases are within a secure network and penetration testing has confirmed that the data portal is secure against external attacks.

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.

Since 2013, ORR has included revenue per passenger journey by sector alongside the rail fares index. This information is calculated from the passenger revenue and passenger journeys tables, available in the passenger rail usage section of the ORR data portal. The rail fares indices measure price change from one January to the next, so in order to make the information comparable, the equivalent index for revenue per journey has been calculated using Q4 data, which covers January to March.

It is also important to understand the differences between the rail fares indices and revenue per passenger journey when making comparisons between the two measures. Given the exclusions above, in general we would expect revenue per passenger journey to grow at a slower rate than a matched price index although this is not always the case. For example, there are instances when revenue from particular products are recorded in the most recent quarter, but not recorded in the equivalent quarter last year, or vice versa. Therefore any comparisons between the two should be treated with caution.

The revenue per journey metric reflects changes in customer behaviour and reflects 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.

Comparability to European statistics

The difference in the structure of internal rail markets in European countries means that finance statistics are difficult to compare across member states. The statistical office of the European Union, Eurostat, collects no financial statistics on the rail market. Limited financial information is collected by the [Independent Regulators' Group \(IRG-Rail\)](#) for their Market Monitoring Report, including information on passenger and freight revenues although data are not supplied by all European countries.



© Crown copyright 2020

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 [ORR data portal](#)

Any media enquiries regarding this publication should be sent to us at orr.gov.uk/contact-us.