Review of the Rail Fares Index

Prepared for the Office of Rail Regulation

Jim O'Donoghue
Methodology Advisory Service,
Office for National Statistics
April 2014
Tel: 020 7592 8641
Email mas@ons.gov.uk

9650119

Review of the ORR's rail fares index

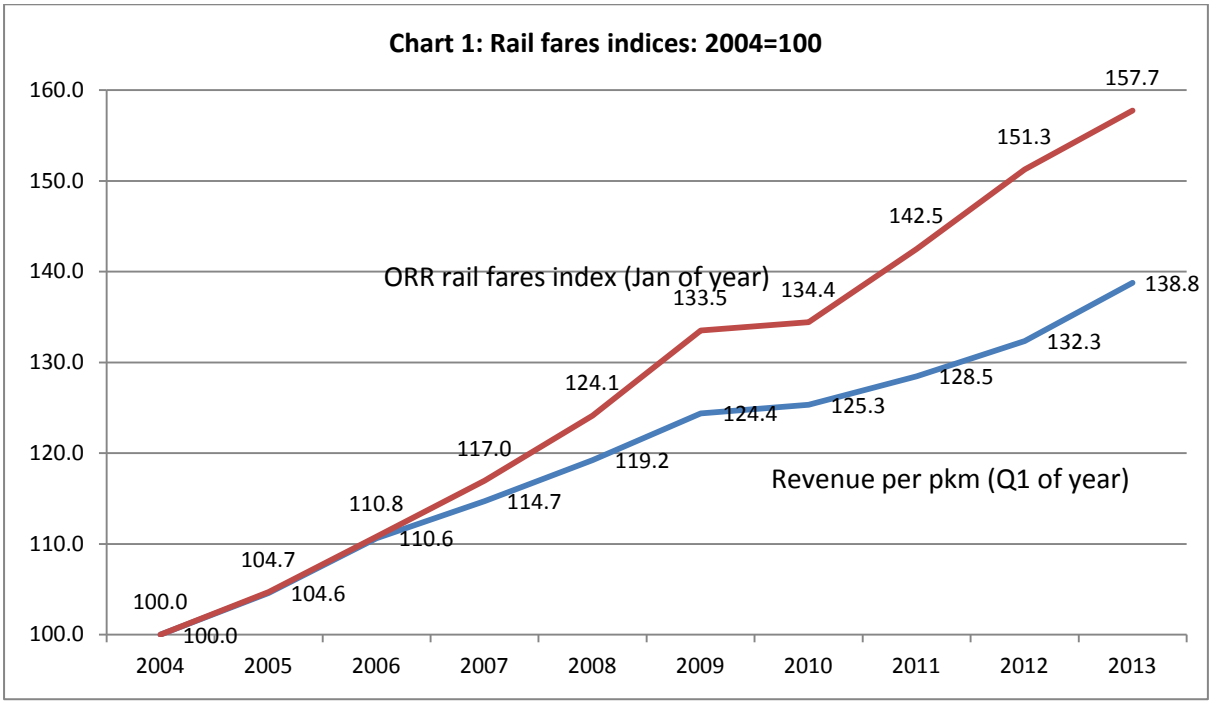
1. Background

The background to this review is set out in the note from ORR, *MAS brief on the rail fares index* (see Annex A) and briefly summarised below.

ORR produces the annual rail fares index which measures the average price change in rail fares from one January to the next (January is the month when fares usually change) by calculating the price change between Jan_x and Jan_{x-1} weighted by earnings.

The data used to compile the ORR index is taken from the rail industry's ticketing and revenue database (LENNON) for the calendar year, detailing the revenue taken by each train operating company (TOC) for each available fare on every flow (origin-destination pair). The same database is used to produce estimates of receipts per passenger journey (pj) and receipts per passenger kilometre (pkm). The fares index has been increasing much faster than unit receipts since 2007, as illustrated in Chart 1.

Chart 1 – Comparison between rail fares index and average revenue per journey index



The divergence shown in the chart is a source of concern to both ORR and major stakeholders, the Department for Transport and Association of Train Operating Companies (ATOC).

ONS' Prices Division, responsible for the production of the consumer prices index and retail prices index, have used the change in regulated fares from 2012 onwards. This was primarily due to concerns over timeliness of the ORR data. Using the capped regulated increase is considered to be a straightforward and complete way of assessing increases across the regulated rail sector.

This report:

- Reviews the methodology used to produce the fares index
- Makes recommendations for changes to the index, with a view to reducing resources expended on producing it, and improving timeliness
- Suggests areas for further investigation
- Considers reasons why the fares index differs from an index of unit revenues

2. Calculation of the fares index

2.1 Scope

The fares index covers franchised and non-franchised operators in Great Britain, but does not include miscellaneous charges associated with railway operations (e.g. car parking charges).

The published results provide a breakdown of the percentage change in fares analysed by:

- Sector – London and the South East; Long distance; and regional; and
- Ticket type – advance; anytime; off-peak; super off-peak; season ticket; and other.

The unit revenues index does not include non-franchised operators but does include miscellaneous charges. These differences are relatively small in terms of earnings, and do not explain the divergence in the two indices that has been witnessed since 2007.

2.2 Data sources

The primary data source for the rail fares index is the rail industry's central ticketing system called LENNON ("Latest Earnings Networked Nationally OverNight"). It allocates revenue to routes and therefore train operators. It is also the basis for statistics on passenger kilometres, journeys and revenue data that are available through the ORR's data portal. LENNON holds information on all national rail tickets purchased in Great Britain. It is owned and operated by RSP Ltd.

LENNON contains two datasets: pre-allocation (sales) and post-allocation (earnings). Passenger usage statistics are generated using a mathematical model which identifies passenger 'opportunities to travel' from an origin station to a destination station using timetable information. This may include one or more changes of train, and one journey is generated for each train used during an opportunity to travel. This results in the number of journeys being inflated by around 5% compared to the pre-allocation dataset that does not assign journeys between TOCs.

The allocation process is also applied to season tickets. For instance, revenue from an annual season ticket is distributed evenly to each of the thirteen 4-week accounting periods for which it is valid. Similarly, allocation factors exist for the number of journeys that are made using a season ticket.

Multi-modal or area wide tickets (such as Rovers or BritRail passes) present a particular problem for the allocation process. Here there is no way to be sure how the ticket has been used and the system uses manual allocations, generally based on survey data, to estimate revenues, journeys and passenger kilometres.

2.3 Index form

The rail fares index is a Laspeyres type price index. Formally, this is expressed as follows:

$$I_t = \frac{\sum_i P_{i,t} Q_{i,0}}{\sum_i P_{i,0} Q_{i,0}}$$

Where *P* and *Q* are the prices and quantities of item *i* in the base period *0* and a later period *t*.

This can be re-arranged as follows:

$$I_t = \frac{\sum_i P_{i,0} Q_{i,0} * \left(P_{i,t} / P_{i,0} \right)}{\sum_i P_{i,0} Q_{i,0}} = \sum w_i * \frac{P_{i,t}}{P_{i,0}}$$

Where w_i is the expenditure weight for category *i* in the base period.

In the rail fares index, the categories are defined as ticket type by sector. Weights are updated annually. The resulting indices are chain-linked together to form a continuous series.

This type of index holds a basket of goods and services constant in a base period and calculates how much it would cost to buy that basket in the base period and later period. The percentage change between the two periods in the cost of the fixed basket is the fares index. An index of this type takes no account of changes in the composition of the basket within the year – e.g. switching between fare types from Anytime to Advance tickets. The reason for this is that the different tickets are not considered to be of the same "quality", reflecting the restrictions that exist on the use of Advance tickets.

Switching between ticket types between years is taken account of through the annual updating of the weights.

2.4 Index calculation

The fares index is calculated by comparing the prices of tickets for each unique combination of origin-destination pair and fare type in successive Januarys. The relative changes in price are weighted together using annual revenue generated between the two Januarys for each relevant route and fare type.

The price of each of these fares is taken from LENNON and/or the National Fares Manual so that there is a price from January in the current year (Jan_x) and a price from January in the previous year (Jan_{x-1}). These prices are based on the pre-allocated dataset for transactions taking place in the first 4-week accounting period following the date of the fare change – ie for season tickets they relate to the price paid to purchase these tickets at the new rate.

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)

Also omitted from the index are those flows with annual earnings under £100. The thresholds for these omissions are arbitrary between years, depending on the volume of records in the dataset.

Finally, only price changes in the range -40% to +60% are taken into the calculation.

The application of these thresholds reduces the number of flows in the fares index dataset from 25.1 million to 6.7 million (by excluding zero fares) and to 1.6 million (by also excluding low revenue routes).

The process of generating and checking the January prices is resource-intensive and time-consuming. Large price changes have to be checked manually. Where there are doubts about the quality of the prices, fares are looked up on a separate fares database, the National Fares Manual, and entries are amended on the fares index dataset.

2.5 Weighting

In the formula for a Laspeyres price index set out in section 2.3, the prices and quantities in the base period are known, and are used to form the weights. In practice, it is difficult to obtain a representative set of quantities at a single specific point in time and it is usual for the expenditure weights to be derived from a period (typically one year) preceding that used for the base prices. For instance, last year's price index for newspapers in the CPI compared prices in Jan 2013 with later months using weights based on expenditures in 2012.

The rail fares index is different in that it uses revenues for the year following the base period (January) but only for flows that are included in the fares index calculation. This means that revenues generated on lightly used routes or from fares newly introduced or dropped during the course of the year are excluded from the calculation of the fares index.

3. Probability sampling

One of the objectives of the review is to consider how to reduce the resources used to produce the index. This section considers the possibility of introducing probability sampling. Testing was done using the 2012 dataset which contained the January prices for 2011 and 2012, and earnings for each flow.

Five scenarios were tested, where all flows with earnings above a pre-specified amount were included with certainty in the sample, together with a random sample of smaller flows. The latter were ranked according to earnings and systematically sampled using a random start point. The detailed results are shown in Annex B and summarised below.

Table 1: 95% confidence intervals for percentage change in fares: 2011

			Per cent			
	Sampling interval (£m)					*Revenue*
	£0.5m	£0.25m	£0.1m	£0.05m	£0.025m	*(£m)*

advance	0.23	0.18	0.10	0.08	0.03	*696*
anytime	0.08	0.06	0.03	0.03	0.01	*1,533*
off-peak	0.16	0.09	0.05	0.04	0.02	*1,299*
other	1.31	1.34	0.69	0.60	0.31	*24*
season	0.07	0.05	0.03	0.03	0.02	*1,297*
super off-peak	0.36	0.26	0.12	0.08	0.04	*254*
London and South East	0.09	0.06	0.03	0.03	0.01	*2,271*
Long distance	0.10	0.07	0.04	0.03	0.01	*2,133*
Regional	0.16	0.12	0.07	0.06	0.03	*700*
All	0.06	0.04	0.02	0.02	0.01	*5,103*
Sample size	*8,962*	*14,820*	*26,862*	*40,513*	*58,801*	

The published fare changes are rounded to one decimal point. It can be seen that a sampling interval of £0.1m yields 95% confidence intervals of no greater than 0.12 per cent for all categories, with the exception of "other fares" which makes up only 0.5 per cent of total revenues. The total number of flows used in this calculation was around 5 per cent of that used for the published indices.

The detailed results, where ticket type is cross-tabulated by sector, are shown in Annex B and it can be seen that these have wider confidence intervals, although for the main categories where expenditure is greatest (anytime, off-peak, seasons, long distance advance), they are no greater than 0.15 per cent. Smaller confidence intervals are obtained by reducing the sampling interval to £0.025m.

It can also be seen that each of the sampling intervals tested results in sample sizes much smaller than the existing method, which in 2011 comprised 1.6 million flows.

The formula for calculating the average percentage change for a particular category is as follows:

$$pc = \frac{\sum_{i=1}^{l} e_i p_i}{e_t} + \left(1 - \frac{\sum_{i=1}^{l} e_i}{e_t}\right) * \frac{1}{s} * \sum_{j=1}^{s} p_j$$

Where e_t is total expenditure, e_i is expenditure on flow *i*, p_i is the percentage change for flow *i*, and there are l flows with earnings greater than the sampling interval and *s* flows with earnings less than the sampling interval.

Sampling errors were calculated approximately as follows:

$$se = \left(1 - \frac{\sum_{i=1}^{l} e_i}{e}\right) * \frac{1}{\sum_{k=1}^{c} n_k} * \sqrt{\sum_{k=1}^{c} n_k \, \sigma_k^2}$$

Where there are *c* strata (or categories), n_k flows in category *k*, and σ_k^2 is the variance of the percentage changes for category *k*. (Note that there is no variance associated with the flows that are greater than the sampling interval, as they are selected with certainty.)

Conclusions and recommendations to be drawn from this analysis:

R1 Probability sampling should be introduced to select the sample of flows from which the fares index is calculated. This would substantially reduce the number of flows for which January fares need to be checked

R2 The choice of sampling interval should be determined by the desired level of precision for the results

R3 Consideration should be given to dropping the separate publication of the "other" category given its small weight and the size of its sampling error

R4 Similarly, the need for the publication of the detailed sector by ticket type fares indices should be considered

R5 If it is decided to continue with publication of fares indices as at present, the use of unequal sampling fractions should be considered to achieve more precise results for the smaller weighted fares indices

4. Excluded earnings

This section examines the effect of trimming the data taken from Lennon to exclude flows where there are zero fares recorded in either January and to exclude flows which generate less than £100 in annual earnings. The amount of revenue excluded by these two actions is shown in the table below for 2012 earnings.

Table 2: Revenues by sector: excluded earnings: 2012

				£m	
	London and South East	Long Distance	Regional	All	Number of flows (000s)
Revenues from ORR data portal	3,509	2,485	1,069	7,063	
Total earnings on Lennon extract for the fares index	3,043	2,717	944	6,704	25,100
Total exc zero fares	2,400	2,449	868	5,717	6,654
Total exc zero fares and <£100 earnings	2,361	2,418	782	5,561	1,582
Total revenues removed £m	682	299	162	1,143	
as % of total	*22%*	*11%*	*17%*	*17%*	

It can be seen that:

1. Total revenues from the data portal are greater than those used for the fares index Lennon, but Long Distance revenues are less
2. The overall effect of trimming the fares index flows is to result in flows totalling one-sixth of earnings being removed.
3. The excluded flows are greatest for London and South East

4. Overall, nearly £1bn in revenues are lost by excluding zero fares, compared with £150m lost by excluding low revenue flows.

The reason for fares index revenues being lower in aggregate than those available on the data portal is partly because the data portal includes non-fares revenue, such as parking charges and fines. However, this does not account for all the differences, as the "other revenues" category on the data portal, in which this revenue would be recorded totalled only £135m. **R6 The reasons for this discrepancy should be investigated further.**

Long distance earnings are greater for the fares index because historically the fares index has mapped all routes operated by a TOC to one sector, whereas the portal data has been assigned to sectors based on the service group. For example, First Great Western operate in all three sectors but, for fares, all FGW services are allocated to the Long Distance sector. It would be helpful for users to have consistency of definitions for different ORR statistics and it is recommended that **R7 the fares index definitions should be brought into line with those used on the data portal**.

A more detailed examination of the flows with zero fares shows that many have revenues well in excess of £100. For instance, the largest flow had revenues of £20m in 2012. In total, there were 150,000 of these flows, accounting for £653m of revenues. These flows are concentrated in three season ticket types, as shown in the table below.

Table 3: Flows with zero fares but patronage above £100 per annum: 2012

ticket code	Number of flows	Revenues (£m)
J	213	202
M	59,026	89
O	78,544	251
Other	12,235	111
Total	150,018	653

Ticket types J and O represent revenues generated from the use of Oyster/travelcards in the London area. The ability to use these cards on national rail services is a fairly recent development, dating back to 2011. **R8 These flows should be investigated further together to see if more detail can be obtained as to what they represent and whether prices can be obtained for them.** This should be done as a matter of priority, given the volume of revenues involved and the increased usage that is being made of Oyster cards.

There are also substantial revenues associated with other non-Oyster/travelcard flows. **R9 These should also be investigated further to see if it can be established what they represent.**

5. Other issues

5.1 Advance tickets

Advance tickets must be booked in advance and are sold in limited numbers and so are subject to availability. In general, the further ahead the booking is made, the cheaper the Advance fare will be. Advance fares are valid only on the date and train shown on the ticket and are non-refundable.

Changes in the pricing structure for advance tickets on a particular route – e.g. moving from a 3-point to 4-point fares scale – can lead to the introduction of a new product code or dropping of an old code during the course of the year. Under the current procedures for compiling the fares index, the new/dropped code would not be included in the calculation as it would be missing in one of the two Januarys. However, the tickets sold with the new/dropped code can be considered comparable to other advance tickets sold on the same route, as they will have similar restrictions for their usage. It is therefore recommended that **R10 for advance tickets an overall average price, covering all advance ticket sales, should be calculated for each operator on a route (origin-destination pair) for the 4-week accounting period following the price change.**

Advance tickets may also be introduced on a route for the first time during the course of the year. This is likely to be accompanied by some switching of ticket sales from Anytime to cheaper Advance tickets; it may also result in increased patronage on that route. These are compositional changes, involving switching between what can be considered to be tickets of differing quality. (Anytime fares are fully flexible tickets, with no time restrictions on when they can be used. Contrast this with Advance tickets.) Switching of this type correctly does not impact on the fares index.

5.2 Off-peak fares

Off-peak fares are cheaper tickets for travelling on trains that are less busy. These tickets may require passengers to travel at specific times of day, days of the week or on a specific route. Where there is more than one Off-Peak fare for a journey, the cheaper fare with more restrictions is called Super Off-Peak**.**

Off-Peak tickets can be bought any time before travel. The travel restrictions depend on the journey being travelled and customers will be advised of these when buying their ticket.

The restrictions associated with the use of an off-peak or super off-peak ticket may vary – e.g. the time from which an off-peak ticket may be used may change.

The differing restrictions in the availability of off-peak and super off-peak tickets mean that they cannot be considered to be of comparable quality. **R11 The off-peak and super off-peak indices should therefore continue to be compiled separately, as at present.**

5.3 Weighting

The weights used in the calculation of the rail fares index are based on revenues for flows in the reduced data set where flows with low patronage or zero fares have been filtered out. ORR report that the effect of excluding low patronage routes is very small, and calculations for 2012 confirm this, showing that their exclusion has minimal impact on the overall index (0.002 per cent), with small effects on the detailed ticket type by sector fares indices (most differ by less than 0.1 per cent).

Analysis of the 2012 data also shows that the proportion of excluded revenues varies by category and sector. This does not matter when the percentage changes for the detailed sub-indices do not differ by very much, which tends to be the case. However, this may not always be so and it is recommended that **R12 the detailed sub-indices are weighted together using revenues that include revenues generated on low patronage routes**.

5.4 Valid price change range

Only flows with a percentage change between -40% and +60% are included in the calculation of the fares index. The sensitivity of the results to changes in these parameters was examined. The results are shown in the table below.

Table 4: Percentage change in fares index by valid percentage change range: 2012

	Percentage change range					
	No limits	-40 to 60	-33 to 50	-25 to 33	-20 to 25	-20 to 20
Overall % change	6.18	6.13	6.11	6.08	6.04	5.99
London & South East	6.10	6.09	6.05	6.02	5.97	5.94
Long Distance	6.14	6.08	6.07	6.05	6.01	5.96
Regional	6.55	6.38	6.38	6.34	6.30	6.28
advance	5.73	5.55	5.54	5.47	5.36	5.24
anytime	6.39	6.38	6.38	6.38	6.38	6.35
off-peak	6.37	6.28	6.22	6.17	6.15	6.11
other	6.51	6.50	6.56	6.46	5.78	5.44
season	5.98	5.99	5.99	5.98	5.97	5.96
super off- peak	6.03	5.98	5.96	5.90	5.54	5.51
Earnings in range (£m)	*5,561*	*5,557*	*5,551*	*5,547*	*5,532*	*5,520*

It can be seen that:

- Narrowing the valid range reduces the overall change in the fares index
- The effect varies by sector and ticket type
- Season tickets and Anytime tickets are unaffected by changes in the valid range

The introduction of probability sampling, as recommended above, will reduce substantially the number of flows used in the calculation of the fares index and should permit relatively more of the outliers with high percentage change to be checked. It is recommended that **R13 all flows with percentage changes outside the range -20% to +20% should be checked.**

6. Comparison of the fares index and unit revenues index

The rail fares index and the unit revenues index are conceptually different. The fares index measures the changing cost of a fixed basket of journeys. The unit revenue index by contrast is affected by switching between fare types (compositional change).

The difference between the two approaches is illustrated in the following example.

	No. jnys	Price per jny	Actual revenue	Base weighted revenue
Base period				
Ticket type A	100	20	2000	2000
Ticket type B	50	30	1500	1500
All Ticket types	150	23.3	3500	3500
Later period				
Ticket type A	110	22	2420	2200
Ticket type B	45	33	1485	1650
All Ticket types	155	25.2	3905	3850
% change				
Ticket type A		10%		
Ticket type B		10%		
All Ticket types		8.0%	11.6%	10.0%

The price per journey on both ticket types has increased by 10 per cent; so the fares index also goes up by 10 per cent (bottom right hand cell in the table). However, in the later period there are relatively more journeys made at the cheaper price, so the average revenue per journey goes up by less (8 per cent) than the change in fares (10 per cent).

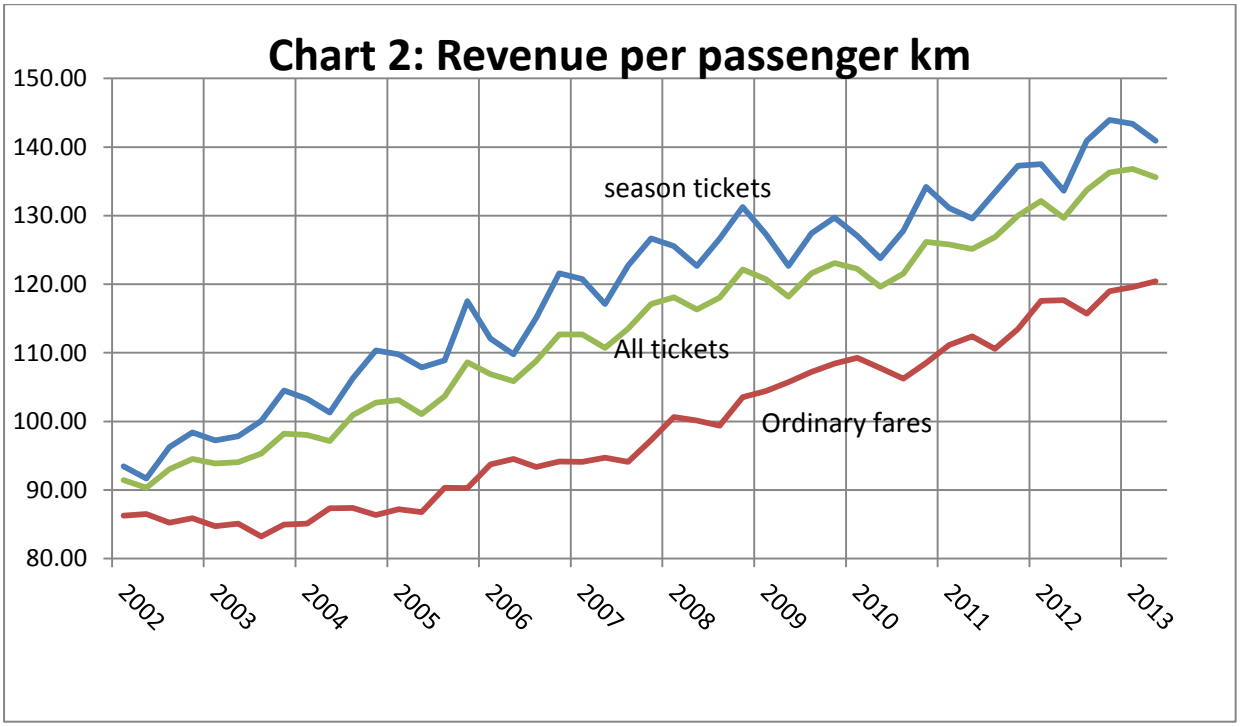
Statistics published on the ORR's data portal shows an increase in the proportion of revenues generated by Advance ticket sales – up from 15.6 per cent of all franchised ordinary fare ticket receipts in 2010 to 18.7 per cent in 2013 – mainly at the expense of full price Anytime tickets. This will help to pull down the unit revenues index relative to the fares index, as would any switching away from First Class to Standard fares that might have taken place as a result of the recession.

The reason why the fares index does not take account of switching between ticket types (e.g. between anytime and advance tickets) is that the different tickets are not considered to be of the same "quality", reflecting the restrictions that exist on the use of advance tickets.

Because the unit revenue index is affected by compositional change, it will vary throughout the year. The fares index, by contrast, will only change when prices change. It would therefore be misleading to think of a unit revenue index as a direct measure of the change in rail fares as the latter does not change continuously throughout the year.

Chart 2 shows the change in average revenues. There is a pronounced seasonal pattern, particularly for season tickets, which tend to have peaks in the 1st quarter of the calendar year and troughs in

the 3rd quarter. This is most likely because there are relatively fewer weekly and monthly season tickets bought during the summer months, with people going away on holiday. These tend to be more expensive per journey than annual tickets, so the average falls.

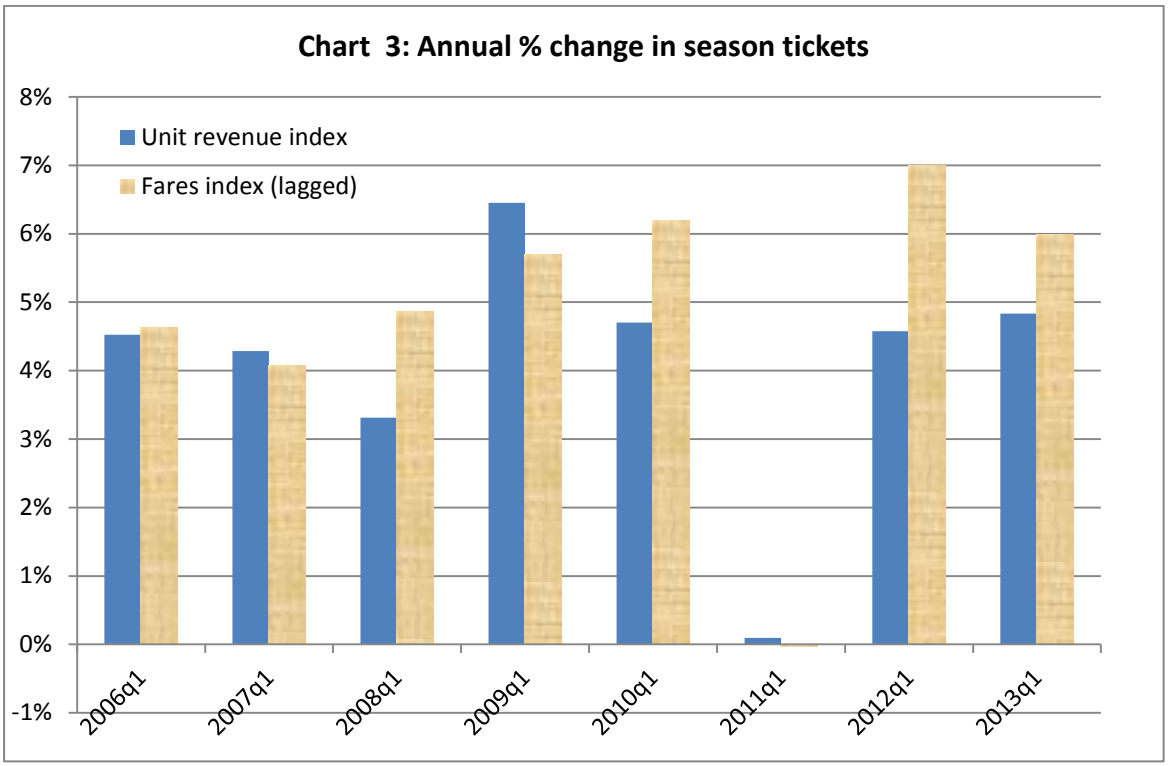


There is also a timing issue when comparing unit revenues with changes in fares, which mainly impacts on season tickets; consider annual season tickets. The price of a season ticket usually changes once a year, let's say on 1 January. So the percentage change from one year to the next is reflected in the fares index for January. However, commuters renew their season tickets throughout the year, so the impact on revenues and the unit revenue index feeds through gradually during the course of the year. It is only when all commuters have renewed their season tickets that the full impact of the change in price on January 1 is reflected in the unit revenue index – i.e. one year later than for the fares index.

This can be seen to some extent in chart 3, which compares the fares index lagged by one year against the unit revenues index for the first quarter of the calendar year. The correspondence between the two indices will not be exact because there are other frequencies of season tickets (e.g. weeklies and monthlies) whose full impact will be felt sooner. For instance, the change in the unit revenue index for January will contain a mixture of the previous year's increases for annual season tickets and the most recent year's increases for weeklies.

One final point that needs to be noted is that the unit revenues index captures all revenue and ticket types, whereas the fares index only captures those fares that exist in both Januarys. Consider a situation where a TOC introduces a new set of fares during the course of the year (e.g. Advance tickets on a new set of routes) – because they are not available in the January at the start of the year, there will be no price against which to compare them in the January following their

introduction, so they do not impact on the fares index. However, their introduction will be captured in the unit revenues index.



7. Conclusions

The basic methodology being used for the fares index is sound, although there is a question mark over the treatment of zero-fare high-patronage routes which needs to be investigated as a matter of priority, particularly with the growth in Oyster card usage.

There are some relatively minor aspects of the methodology that could be improved, e.g. the use of average fares by route for Advance tickets; and re-weighting of sub-indices to include revenues generated on low patronage routes.

Probability sampling could be introduced with minimal impact on the headline fares index and the main sub-indices. This would reduce the resources and time involved in producing the fares index. Resources freed up could be used to check the largest percentage changes in fares, as the fares index seems to be quite sensitive to outliers.

It would be helpful for users if definitions used in the fares index and the data portal were harmonised.

Conceptually, the fares index and the unit revenues index are measuring different things and, with the growth in the sales of discount tickets (i.e. Advance), some divergence between the two is to be expected.

ANNEX A : BACKGROUND BRIEFING DOCUMENT FROM ORR

ORR rail fares index: Is the current methodology of calculation appropriate and, if not, what alternative methodology should we apply?

Background

ORR produces the annual rail fares index which measures the average price change in rail fares from one January to the next.

The rail fares index is disaggregated by sector (London & South East, Regional and Long Distance) and also by ticket type (Advance, Anytime, Off Peak, Season, Super Off Peak and Others). The current base year for the index is 2004 = 100.

The data are also available by sector, regulated/unregulated fare[1] and ticket class in an index with a base year of 1995 = 100. This is not published but is available on request.

The rail fares index is also included within the basket of goods and services for the Consumer Prices Index so it is important that, where possible, we measure the change in prices using methodology that is consistent with that used in the compilation of the CPI/RPI.

Current methodology

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 approximately 10 million records.

To calculate the index itself, the price of each of these fares is taken from LENNON so that we have a price from 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:

3) The flow and/or fare were introduced in the calendar year so there is no price for Jan_{x-1}
4) 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)

Also omitted from the index are those flows with earnings under £100 and flows where the number of journeys is less than 100. The thresholds for these omissions are arbitrary between years, depending on the volume of records in the dataset though checks carried out on the dataset ensure that the impact on the final index itself are minimal (<0.01%).

Each record within the final dataset is assigned to a sector based on the train operating company (TOC) running the service, and assigned to a ticket type based on the product code and description.

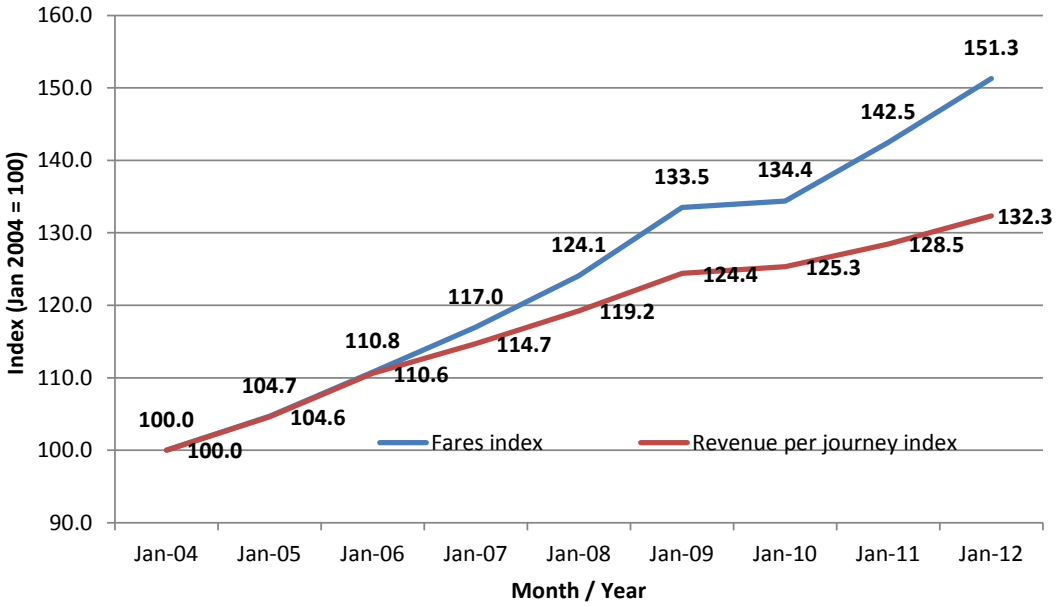
[1] Regulated fares are subject to a 'cap' applied to a weighted average of the relevant fares for each train operating company (TOC). This 'cap' is currently RPI+1%. Unregulated fares are determined commercially by the TOC

The index is produced by calculating the price change between Jan_x and Jan_{x-1} weighted by earnings.

Issues

The Association of Train Operating Companies (ATOC) has raised the issue that the current methodology of excluding records from the index calculation leads to the actual increase in fares being overstated. Chart 1 compares the index for revenue per passenger journey with that for average change in rail fares. This demonstrates the divergence in the two measures.

Chart 1 – Comparison between rail fares index and average revenue per journey index



There are some differences in scope between the two measures:

- Rail fares index includes non-franchised operators
- Rail fares index does not include miscellaneous charges associated with railway operations (e.g. car parking charges)

However, this would not explain the large differences that we have witnessed since 2007 when the difference between the two measures has widened.

Also, previous checks on the index have demonstrated that those flows omitted due to having low revenue levels or low passenger demand could not account for the difference in the two indices.

Therefore, a large part of the divergence may be down to the fact that fares that are introduced or discontinued during the year are not included in the final index calculations.

Although we have developed a series of options for calculating a rail fares index which are detailed in the following section, this is not an exhaustive list and we are happy to consider alternative suggestions depending on our resource constraints. We welcome the thoughts of the Methodology Advisory Service and would ask for their recommendation on which methodology to use bearing in mind our desire to be consistent with methodology employed in calculating other components of the CPI/RPI.

Option 1 – Discontinue current methodology and use revenue per journey as a proxy for the change in rail fares

The current fares index is a significant resource burden on ORR and one option is to discontinue the annual fares index and switch to using revenue per journey as a proxy for the change in rail fares.

This would have the benefit of taking into account when people switch between different tickets. For example, in the latest January 2013 fares index table, the weights data refers to the 2012 calendar year. If we were to switch to using the revenue per journey measure, this would mean the data was more reflective of the current position and would have the additional benefit of being updated more regularly as revenue per journey is published quarterly.

One of the potential downsides is the impact of discontinuing the time series on our users and whether they would be satisfied with using revenue per journey. We would need to consult our stakeholders and, ultimately make a decision based on their responses.

Option 2 – Include estimates of prices where there is only one price reference, either Jan_x or Jan_{x-1}

Where Jan_x is not available but Jan_{x-1} is available, we have a number of options on estimating a price:

a) We estimate a price for Jan_x based on the average change for that ticket type in the sector. For example, if the record refers to an advance ticket in the Regional sector, we will use the average price change for that group in the matched price index to calculate an estimated price for Jan_x based on Jan_{x-1}
b) Estimate a price for Jan_x based on an alternative set of variables other than ticket type and sector (e.g. TOC and ticket type, regulated fare 'cap')
c) Calculate the average price change for that ticket type in the sector and apply that to the individual records, not including any estimate of price though this would mean there would be no price reference in future datasets.
d) Continue to omit these from the index as if a price does not exist for Jan_x this is because the fare has been discontinued so is unlikely to re-appear in the index.

Similarly, where Jan_{x-1} is not available but Jan_x is available

a) We impute a base price for Jan_{x-1} based on Jan_x and the average change for that ticket type in the sector.
b) Calculate the average price change for that ticket type in the sector and apply that to the individual records, not including any estimate of price. Alternatively, use a different set of variables to derive the price change.
c) Estimate a base price based on an alternative set of variables other than ticket type and sector (e.g. TOC and ticket type, regulated fare 'cap').
d) Continue to omit these from the index.

Analysis comparing the fares index using option a) with the current methodology shows there is very little difference in the overall outcome. Due to there being such a large number of unmatched prices, using the sector/ticket type average change, in the most part only reinforces the current numbers. Therefore, if either option a) or option b) are deemed to be the preferred method, a more

disaggregated set of variables will need to be used to derive the average price changes in order to make any real change to the index.

Sector	Ticket Type	Current methodology	Unmatched replaced with average change in ticket & sector
London and South East	**advance**	8.03	8.06
	anytime	4.60	4.61
	offpeak	4.34	4.39
	other	6.04	6.04
	season	4.15	4.18
	superoffpeak	4.64	4.64
Long Distance	**advance**	4.24	4.26
	anytime	3.54	3.56
	offpeak	4.25	4.23
	other	1.91	2.81
	season	4.55	4.51
	superoffpeak	4.41	4.40
Regional	**advance**	3.88	3.84
	anytime	4.44	4.42
	offpeak	4.43	4.45
	other	4.11	4.19
	season	4.46	4.51
	superoffpeak	7.15	7.15

Option 3 - Include estimates of prices where there are no price references in either Jan_x or Jan_{x-1}?

Where neither Jan_{x-1} nor Jan_x are available, we have considered the following possible options for estimating prices:

a) We use the average change for that ticket type in the sector and apply that price change to the individual records, not including any estimate of price. Alternatively, use a different set of variables to derive the price change.
b) Continue to omit these from the index. Given that there is no price in either Jan_x or Jan_{x-1} it is likely that they have been introduced as a short-term special fare during the calendar year and are, therefore, unlikely to re-appear in the index.

Other information

Any changes to the index would result in a break in series as we do not store the omitted records so it would not be possible to back fit any changes in methodology.

For further information on the fares index, please see the latest publication available on ORR's data portal http://dataportal.orr.gov.uk/displayreport/report/html/7cff3127-a5cc-4173-ac78-016db2339811

And the accompanying statistical release and quality report: http://orr.gov.uk/statistics/published-stats/statistical-releases

ANNEX B: THE EFFECT OF THE INTRODUCTION OF PROBABILITY SAMPLING

	sampling interval	no. in sample
Estimates	£0.5m	8962

Row Labels	advance	anytime	offpeak	other	season	superop	Grand Total
Lon SE	3.57	6.37	6.12	7.53	5.96	5.71	6.08
Long D	5.42	6.27	6.54	6.16	5.95	6.14	6.06
Rgiona	5.27	6.62	6.11	6.97	6.03	7.59	6.29
Grand Total	**5.33**	**6.38**	**6.31**	**6.68**	**5.96**	**6.10**	**6.10**

Published

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	3.04	6.43	6.15	4.90	5.96	5.86	6.09
Long D	5.71	6.21	6.40	6.68	5.94	5.89	6.08
Rgiona	5.50	6.61	6.25	6.83	6.24	8.14	6.38
All	**5.56**	**6.38**	**6.28**	**6.50**	**5.98**	**5.98**	**6.13**

Difference

	advance	anytime	offpeak	other	season	superoffpeak	all
Lon SE	0.53	-0.06	-0.03	2.62	-0.01	-0.15	-0.02
Long D	-0.29	0.06	0.14	-0.52	0.02	0.25	-0.01
Rgiona	-0.23	0.01	-0.14	0.14	-0.20	-0.55	-0.09
All	**-0.23**	**0.00**	**0.02**	**0.18**	**-0.02**	**0.11**	**-0.03**

95% confidence interval

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	1.06	0.13	0.31	3.97	0.08	1.11	0.09
Long D	0.25	0.12	0.20	1.61	0.20	0.26	0.10
Rgiona	0.86	0.21	0.30	2.29	0.30	1.77	0.16
All	**0.23**	**0.08**	**0.16**	**1.31**	**0.07**	**0.36**	**0.06**

	sampling interval	no. in sample
Estimates	£0.25m	14820

Row Labels	advance	anytime	offpeak	other	season	superop	Grand Total
Lon SE	2.90	6.46	6.19	7.26	5.97	6.03	6.12
Long D	5.66	6.12	6.46	7.11	5.77	5.84	6.04
Rgiona	6.11	6.67	6.12	7.89	6.04	8.83	6.39
Grand Total	**5.54**	**6.37**	**6.30**	**7.42**	**5.95**	**6.02**	**6.13**

Published

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	3.04	6.43	6.15	4.90	5.96	5.86	6.09
Long D	5.71	6.21	6.40	6.68	5.94	5.89	6.08
Rgiona	5.50	6.61	6.25	6.83	6.24	8.14	6.38
All	**5.56**	**6.38**	**6.28**	**6.50**	**5.98**	**5.98**	**6.13**

Difference

	advance	anytime	offpeak	other	season	superoffpeak	all
Lon SE	-0.14	0.02	0.04	2.35	0.01	0.17	0.03
Long D	-0.05	-0.08	0.06	0.43	-0.17	-0.05	-0.03
Rgiona	0.61	0.06	-0.13	1.06	-0.20	0.70	0.01
All	**-0.01**	**-0.01**	**0.02**	**0.92**	**-0.03**	**0.04**	**0.00**

95% confidence interval

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	0.70	0.09	0.19	5.53	0.05	0.76	0.06
Long D	0.18	0.07	0.10	1.92	0.19	0.14	0.07
Rgiona	0.83	0.15	0.20	1.72	0.23	1.20	0.12
All	**0.18**	**0.06**	**0.09**	**1.34**	**0.05**	**0.26**	**0.04**

	sampling interval		no. in sample				
Estimates	£0.1m		26862				
Row Labels	**advance**	**anytime**	**offpeak**	**other**	**season**	**superop**	**Grand Total**
Lon SE	3.35	6.47	6.20	5.85	5.97	5.91	6.13
Long D	5.66	6.16	6.46	6.80	5.91	6.04	6.08
Rgiona	5.47	6.59	6.29	6.50	6.14	7.58	6.35
Grand Total	**5.53**	**6.38**	**6.33**	**6.56**	**5.98**	**6.06**	**6.14**

Published							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**All**
Lon SE	3.04	6.43	6.15	4.90	5.96	5.86	6.09
Long D	5.71	6.21	6.40	6.68	5.94	5.89	6.08
Rgiona	5.50	6.61	6.25	6.83	6.24	8.14	6.38
All	**5.56**	**6.38**	**6.28**	**6.50**	**5.98**	**5.98**	**6.13**

Difference							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**all**
Lon SE	0.31	0.04	0.06	0.94	0.01	0.05	0.04
Long D	-0.05	-0.05	0.06	0.12	-0.03	0.15	0.00
Rgiona	-0.03	-0.02	0.04	-0.34	-0.10	-0.55	-0.03
All	**-0.03**	**0.00**	**0.05**	**0.07**	**0.00**	**0.08**	**0.01**

95% confidence interval							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**All**
Lon SE	0.42	0.04	0.11	2.02	0.04	0.34	0.03
Long D	0.10	0.04	0.06	1.03	0.11	0.09	0.04
Rgiona	0.42	0.10	0.11	1.01	0.15	0.82	0.07
All	**0.10**	**0.03**	**0.05**	**0.69**	**0.03**	**0.12**	**0.02**

	sampling interval		no. in sample				
Estimates	£0.05m		40513				
Row Labels	**advance**	**anytime**	**offpeak**	**other**	**season**	**superop**	**Grand Total**
Lon SE	2.91	6.49	6.20	6.10	5.96	5.70	6.11
Long D	5.66	6.16	6.43	6.43	5.91	5.89	6.05
Rgiona	5.58	6.65	6.29	6.84	6.27	8.58	6.43
Grand Total	**5.51**	**6.40**	**6.32**	**6.54**	**5.98**	**5.95**	**6.13**

Published							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**All**
Lon SE	3.04	6.43	6.15	4.90	5.96	5.86	6.09
Long D	5.71	6.21	6.40	6.68	5.94	5.89	6.08
Rgiona	5.50	6.61	6.25	6.83	6.24	8.14	6.38
All	**5.56**	**6.38**	**6.28**	**6.50**	**5.98**	**5.98**	**6.13**

Difference							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**all**
Lon SE	-0.13	0.06	0.05	1.19	-0.01	-0.16	0.02
Long D	-0.05	-0.05	0.02	-0.25	-0.03	0.01	-0.02
Rgiona	0.08	0.04	0.04	0.00	0.03	0.44	0.05
All	**-0.04**	**0.02**	**0.03**	**0.05**	**-0.01**	**-0.03**	**0.01**

95% confidence interval							
	advance	**anytime**	**offpeak**	**other**	**season**	**superoffpeak**	**All**
Lon SE	0.38	0.04	0.08	2.22	0.03	0.22	0.03
Long D	0.09	0.03	0.04	1.05	0.10	0.05	0.03
Rgiona	0.34	0.07	0.10	0.89	0.13	0.68	0.06
All	**0.08**	**0.03**	**0.04**	**0.60**	**0.03**	**0.08**	**0.02**

	sampling interval		no. in sample				
Estimates	£0.025m		58801				
Row Labels	**advance**	**anytime**	**offpeak**	**other**	**season**	**superop**	**Grand Total**

Lon SE	3.04	6.48	6.19	5.82	5.98	5.83	6.12
Long D	5.67	6.17	6.44	6.64	5.91	5.92	6.07
Rgiona	5.66	6.63	6.29	6.93	6.23	8.20	6.42
Grand Total	**5.54**	**6.39**	**6.32**	**6.63**	**5.99**	**5.99**	**6.14**

Published

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	3.04	6.43	6.15	4.90	5.96	5.86	6.09
Long D	5.71	6.21	6.40	6.68	5.94	5.89	6.08
Rgiona	5.50	6.61	6.25	6.83	6.24	8.14	6.38
All	**5.56**	**6.38**	**6.28**	**6.50**	**5.98**	**5.98**	**6.13**

Difference

	advance	anytime	offpeak	other	season	superoffpeak	all
Lon SE	0.00	0.05	0.04	0.91	0.01	-0.03	0.03
Long D	-0.03	-0.04	0.04	-0.04	-0.03	0.04	-0.01
Rgiona	0.16	0.03	0.05	0.09	-0.01	0.07	0.04
All	**-0.01**	**0.01**	**0.04**	**0.13**	**0.00**	**0.01**	**0.02**

95% confidence interval

	advance	anytime	offpeak	other	season	superoffpeak	All
Lon SE	0.20	0.02	0.03	0.95	0.02	0.10	0.01
Long D	0.03	0.02	0.02	0.40	0.05	0.03	0.01
Rgiona	0.18	0.04	0.05	0.56	0.07	0.38	0.03
All	**0.03**	**0.01**	**0.02**	**0.31**	**0.02**	**0.04**	**0.01**