Review of the OFT's correlation analysis on coach and rail competition

Introduction

1. This appendix uses correlation analysis to provide a preliminary assessment of coach and rail competition in the Greater Anglia franchise area. It begins by summarizing the work of the OFT and then considers the response of NEG to the analysis by the OFT. It concludes that there appears to be a strong positive correlation between the off-peak fares of coach and rail but notes that though at first sight this correlation may be a reason to suppose that there might be competition between coach and rail, there are several factors that may also explain this relationship other than competition.

Coach on rail competition

Price correlation analysis across routes

2. To assess whether there is any coach on rail competition, the OFT performed simple correlation analysis based on 43 initially identified overlap flows to determine how fares and journey times compare across routes. Based on the correlation coefficients shown in the first four rows of Table 1, the OFT concluded that rail fares tended to be higher the less competitive are coach journey times relative to rail journey times.

TABLE 1 **Correlation analysis on flow data**

Variable 1	*Variable 2*	*Correlation*
Rail fare (p per mile)—off-peak	Coach journey time divided by rail journey time	0.35
Rail fare (p per mile)—peak	Coach journey time divided by rail journey time	0.45
Rail fare (p per mile)—off-peak	Coach fare (p per mile)—off-peak	0.55
Rail fare (p per mile)—peak	Coach fare (p per mile—)peak	0.00
Rail fare (p per mile)—off-peak (22 flows where coach is relatively fastest)	Coach fare (p per mile)—off-peak	0.51
Rail fare (p per mile)—off-peak (21 flows where coach is relatively slowest)	Coach fare (p per mile)—off-peak	0.69

Source: OFT economic advice.

Note: 'p per mile' = pence per mile.

3. The OFT also drew attention to the high correlation that existed between coach and rail off-peak fares compared with no correlation between the peak fares. These findings were thought to be suggestive of an element of competition between coach and rail.

4. The OFT acknowledged that these correlation coefficients should be treated with caution since it was possible that other factors, which might be correlated with fares, may be responsible for the observed relationship. As an example, the OFT pointed out that the competitiveness of the car could account for the lower coach and rail fares on the faster routes.

5. To address this possibility, the OFT split the sample into 22 fastest and 21 slowest relative coach journey times. The rationale offered by the OFT for doing so is that in the presence of competition from the car, there would be a weaker correlation between rail and coach fares where coach journey times were relatively slow. As shown in Table 1, the correlations remained and contrary to the suggestion that competition from the car may influence the correlation, the relationship was weaker where coach was faster. This therefore was viewed by the OFT as lending extra weight to the view that the correlations may be indicative of competition between rail and coach. The OFT also considered the possibility that common costs could be another factor driving the correlations but argued that there was no reason to believe that that was the case since the correlations were across individual flows rather than across time.[1]

6. In our investigation, NEG identified more overlaps than in the OFT stage of the investigation (at the time, 58 overlap flows instead of 43). We carried out the correlation analysis on this higher number of flows and found that 50 per cent of the correlation coefficients were statistically insignificant as indicated by the p-values shown in Table 2. More specifically, there was no statistically meaningful correlation between off-peak rail fares and coach journey time relative to rail journey time and, similar to the findings of the OFT, coach and rail peak fares did not appear to be associated. However, there appears to be a strong positive relationship between rail peak fares and relative journey time and between coach and rail off-peak fares.

TABLE 2 **Correlation analysis on flow data for 58 flows**

Variable 1	*Variable 2*	*Correlation coefficient*	*P-value*
Rail fare (p per mile)—off-peak	Coach journey time divided by rail journey time	0.12	P = 0.35
Rail fare (p per mile)—peak	Coach journey time divided by rail journey time	0.44	P = 0.00
Rail fare (p per mile)—off-peak	Coach fare (p per mile)—off-peak	0.67	P = 0.00
Rail fare (p per mile)—peak	Coach fare (p per mile)—peak	0.10	P = 0.46
Rail fare (p per mile)—off-peak (29 flows where coach is relatively fastest)	Coach fare (p per mile)—off-peak	0.44	P = 0.02
Rail fare (p per mile)—off-peak (29 routes where coach is relatively slowest)	Coach fare (p per mile)—off-peak	0.87	P = 0.00
Rail fare (p per mile)—off-peak (29 flows where coach is relatively fastest—route where coach is quicker than rail is excluded)	Coach fare (p per mile)—off-peak	0.29	P = 0.14

Source: CC analysis based on NEG data.

Notes:
1. The p-values state the statistical significance of the correlation coefficient. A p-value of 0.10 and lower indicates statistical significance at conventional levels.
2. 'p per mile' = pence per mile.

7. We also replicated the OFT's analysis with respect to the fastest and slowest relative coach times (see rows 5 and 6 of Table 2). As can be seen, the correlations are consistent with those presented in Table 1. Again, the correlation is strongest for journeys on which coach journey times are relatively slow. This is consistent with the OFT's findings and would imply that competition from the car might not be responsible for the correlation.

[1] Since the analysis is based on a single cross-section, the problems posed by factors present over time, such as inflation, can be largely ignored.

8. However, once the route on which rail exceeds coach in journey time[2] is excluded from the analysis, the correlation coefficient between coach and rail off-peak fares for the 29 fastest routes is no longer significant (see row 7 of Table 2).

9. Notwithstanding, it is interesting to note from Table 2 that although the correlation for the 29 fastest routes is sensitive to inclusion/exclusion of one route, this is not true for the entire sample. This is not surprising given that in the smaller sub-sample the outlier would have a greater influence on the correlation.

Excluding the Norwich–London flow

10. The Norwich–London flow is the largest of the flows identified and is often thought to be the route where competition between rail and coach would be the heaviest if indeed there was any competition. To determine whether the Norwich–London flow is affecting the correlations in any appreciable way, we have excluded it from the list of flows and re-estimated the correlations based on the remaining 57 flows. The findings are reported in Table 3.

TABLE 3 **Correlation analysis on flow data—excluding the Norwich–London route**

Variable 1	*Variable 2*	*Correlation*	*P-value*
Rail fare (p per mile)—off-peak	Coach journey time divided by rail journey time	0.13	P = 0.35
Rail fare (p per mile)—peak	Coach journey time divided by rail journey time	0.44	P = 0.00
Rail fare (p per mile)—off-peak	Coach fare (p per mile)—off-peak	0.67	P = 0.00
Rail fare (p per mile)—peak	Coach fare (p per mile)—peak	0.11	P = 0.43
Rail fare (p per mile)—off-peak (29 flows where coach is relatively fastest)	Coach fare (p per mile)—off-peak	0.44	P = 0.02
Rail fare (p per mile)—off-peak (29 flows where coach is relatively slowest)	Coach fare (p per mile)—off-peak	0.87	P = 0.00
Rail fare (p per mile)—off-peak (28 flows where coach is relatively fastest—route where coach is quicker than rail is excluded)	Coach fare (p per mile)— off-peak	0.29	P = 0.14

Source: CC analysis based on NEG data.

Note: ‘p per mile’ = pence per mile.

11. It is clear that the correlations are virtually identical to those shown in Table 2. Hence, there is no reason to assume that the Norwich–London flow is an outlier that could be driving the correlations.

A selection of the substantive overlap flows

12. It is interesting as well as necessary to determine whether the correlations are significant when only the substantive flows are considered.

13. In Appendix F, we identified 16 substantive flows. Of those, four were indirect flows and one was classed as a limited direct flow.[3] Since we were not provided with the relevant information for these five flows, we concentrate on the remaining 11 substantive overlap flows. These are:

[2]On the route Alresford–Colchester, the journey time of coach relative to that of rail is 0.74.
[3]Further details can be found in that appendix.

Route	*Start of flow*	*End of flow*
490	Norwich	London
481	Ipswich	London
490/497	Yarmouth	London
484	Colchester	London
497	Lowestoft	London
484	Clacton	London
481	Chelmsford	London
490	Norwich	Stratford
497	Diss	London
484	Clacton	Stratford
484	Clacton	Romford

14. Correlations for a sub-sample consisting only of these 11 flows are considered and we also seek to determine whether with their exclusion the correlations remain robust across the remaining 47 flows. Tables 4 and 5 present the findings.

TABLE 4 **Correlation analysis on flow data for 11 overlaps**

Variable 1	*Variable 2*	*Correlation coefficient*	*P-value*
Rail fare (p per mile)—off-peak	Coach journey time divided by rail journey time	0.85	P = 0.00
Rail fare (p per mile)—peak	Coach journey time divided by rail journey time	0.21	P = 0.54
Rail fare (p per mile)—off-peak	Coach fare (p per mile)—off-peak	0.67	P = 0.02
Rail fare (p per mile)—peak	Coach fare (p per mile)—peak	0.02	P = 0.95

Source: CC analysis based on NEG data.

Note: 'p per mile' = pence per mile.

TABLE 5 **Correlation analysis on flow data excluding 11 overlaps**

Variable 1	*Variable 2*	*Correlation coefficient*	*P-value*
Rail fare (p per mile)—off-peak	Coach journey time divided by rail journey time	0.12	P = 0.43
Rail fare (p per mile)—peak	Coach journey time divided by rail journey time	0.50	P = 0.00
Rail fare (p per mile)—off-peak	Coach fare (p per mile)—off-peak	0.68	P = 0.00
Rail fare (p per mile)—peak	Coach fare (p per mile)—peak	0.15	P = 0.31

Source: CC analysis based on NEG data.

Note: 'p per mile' = pence per mile.

15. According to the correlations in Table 4, rail peak fares and coach journey time divided by rail journey time are unrelated whereas rail off-peak fares and coach journey time divided by rail journey time are strongly correlated. This is in contrast to the findings in Tables 2 and 3 as well as Table 5. Though it is debatable, it seems that one would have more reason to expect a correlation between rail off-peak fares and relative coach journey time than between rail peak fares and relative coach journey time.

16. The correlation between off-peak fares is remarkably consistent. The correlation coefficient is essentially the same across the smaller number of flows and across the larger number of flows: the correlation coefficient for the entire sample is 0.67 (see Table 2), for the sample excluding the Norwich–London flow the correlation coefficient is 0.67 (see Table 3), for the sub-sample consisting of only 11 flows the

correlation coefficient is also 0.67 (see Table 4), and for the sample excluding the 11 flows the correlation coefficient is 0.68 (see Table 5).[4]

17. Should these correlations be read as evidence of competition? Bearing in mind that there are a myriad of factors that may be driving these correlations,[5] it is prudent to regard these results as providing tentative reasons to suppose that there *might* be coach on rail competition.

NEG's response

18. NEG responded to the OFT's correlation analysis by criticizing the method employed. NEG proposed that, to determine whether there was coach on rail competition, it was correlations over time and not single period correlations across routes that ought to be considered. NEG pointed out that the relevant test for the purpose of market definition was one of determining 'whether, in response to a small but significant, non-transitory increase in the price of one mode of transport, a sufficient number of customers would switch to using the other mode of transport in order to render such a price increase unprofitable'. Therefore, NEG contested, the OFT's analysis was not appropriate in so far as it could only provide information on the average coach fare and rail fare for routes of a specific length and it was thereby unable to say how passengers might be expected to respond to a fare change. It is not clear why NEG suggested that inferences on switching behaviour with respect to fare changes could be made based on price correlation across time but maintained that price correlations across routes were mute with regard to the possibility of such behaviour. The correlation coefficients, be it across routes or across time, simply give an indication as to the likely relationship between the two fares. Switching behaviour is not directly borne out by either type of correlation. Therefore, neither correlation across routes nor correlation across time can be used as the sole basis for determining how passenger might respond to changes in fares.

19. A supplementary argument put forward by NEG was that the fare data employed by the OFT consisted of different routes with different characteristics and hence any correlation analysis would be 'mixing apples and pears'. Such an analysis, NEG argued, would not be meaningful.

20. However, NEG has neglected the fact that for the analysis to be meaningful there should be some heterogeneity across the routes. A problem only arises if these characteristics are somehow correlated with both fares. If the routes had the same characteristics, then there should be no reason for variation in fares.[6] The strength of the OFT's findings lay in the fact that despite the differences between routes, a correlation could be found between the coach and rail off-peak fares. In other words, this can be taken to imply that regardless of the route considered, coach and rail can be expected to be competing with one another during off-peak times.

21. NEG went on to argue that fares were materially affected by journey distances and that any correlation between fares was simply reflecting this common influence. To

---

[4]Including or excluding the Alresford–Colchester flow from the samples does not materially alter the conclusion. The exclusion of this flow results in slightly higher correlation coefficients. Thus, the conclusion by the OFT that there may be competition between coach and rail with respect to off-peak fares is robust to the inclusion/exclusion of this outlier though the earlier point with regards to the 29 fastest routes should be noted.

[5]It is conceivable that factors such as general attitudes towards public transport, population characteristics in areas serviced by the flows, and other demand factors could affect the degree of correlation between rail and coach fares. Unless these factors are controlled for, their influence cannot be proved nor disproved.

[6]Indeed, if the routes considered were identical then the analysis would be effectively based on a single route.

test this hypothesis put forward by NEG, we purged journey distance from the correlations.[7] As demonstrated in Table 6, the resulting correlations remained high and, with the exception of the correlation with the peak fares for the 11 selected overlaps, significant.

TABLE 6 **Correlation analysis controlling for the effects of journey distance**

Variable 1	*Variable 2*	*All 58 flows*	*Excluding the 11 overlap flows*	*11 flows*
Rail fares (p per mile)–peak	Coach fares (p per mile)– peak	0.34 p-value=0.01	0.36 p-value=0.01	0.41 p-value=0.21
Rail fares (p per mile)–off-peak	Coach fares (p per mile)–off-peak	0.59 p-value=0.00	0.59 p-value=0.00	0.55 p-value=0.08

Source: CC analysis based on NEG data.

Note: 'p per mile' = pence per mile.

22. Therefore, while it is reasonable to argue that examining price correlations over time is a more favoured approach, the results from the OFT analysis based on cross-section data in a single time period cannot be summarily dismissed as invalid. The strong correlations between off-peak fares, though only for a selected number of routes with selected characteristics, could be suggestive of coach on rail competition but could also be a result of other factors.

23. For example, residual elements of past relationships between coach and rail fares may cause a (spurious) current correlation between them.

24. In summary, the observed correlations between coach and rail fares are not conclusive—while it is possible that the associations might be driven by competition between coach and rail it is also possible that they may be driven by other factors.

[7]This was done by performing a correlation analysis between the residuals from the regression of coach fares on coach journey distance and the residuals from the regression of rail fares on rail journey distance. In a supplementary correlation analysis, the possibility of non-linearity in journey distance was considered by looking at the correlation between the residuals from the separate regressions of coach fares and rail fares on the respective journey distance and its square. The correlation coefficients were very close to those reported in Table 6 and are thus not presented here.