

## **E ECONOMETRIC ANALYSIS**

### **Introduction**

- E.1 This annexe is in three parts. Part One presents our econometric methodology, which is unchanged from the methodology we adopted in the Proposed Decision. It also includes tables of regression results, which were included in the Proposed Decision but which we subsequently revised in the Additional Note on Price Concentration. We include those original tables of results in Part One only for completeness: we do not rely upon them.
- E.2 Part Two is the Additional Note on Price Concentration, which was published on 9 October but which was provided to Frontier Economics, who reviewed it on behalf of four major bus operators, in advance of its publication. The note includes our revised regressions on which we rely for our findings. Part Two reproduces the note as it was originally published.
- E.3 Part Three contains our response to the report submitted by Frontier Economics (Frontier) in relation to our econometric analysis. It discusses the issues raised by Frontier and presents a number of robustness tests that we have conducted in the light of Frontier's report.
- E.4 Part One includes the tables and text from Annexe C of the Proposed Decision exactly as they were published in the Proposed Decision. For consistency we use the same paragraph and table numbering in Part One as was used in the Proposed Decision. Therefore, references to Annexe C of the Proposed Decision or paragraphs and tables within it can be read as also being references to Part One of this annexe.
- E.5 Similarly, and for the same reason, references to the Additional Note on Price Concentration can be read as being references to Part Two of this annexe.
- E.6 Taken together the three parts of this annexe present our methodology, findings and robustness checks with respect to our

econometric analysis: to that extent they should be read in combination with each other.

# **PART ONE: REPRODUCTION OF ANNEXE C FROM THE PROPOSED DECISION**

## **C Impact of market structure on bus fares**

### **Introduction**

- C.1 This section examines how bus fares vary across local markets in relation to the number and the identity of bus operators providing local bus services. In particular, we test the hypothesis, which was put to us in the course of this study, that the local bus markets are contestable so that an operator which is a sole provider of commercial bus services in a local area is prevented from increasing prices above competitive level by threat of potential entry.
- C.2 The data which we have obtained in the course of this study rejects the strong version of this hypothesis. Our analysis suggests that bus fares are higher when there is only one national operator providing services in a LTA after controlling for a range of factors which can be expected to influence pricing of bus services. These factors include competition from the car, local area socio-demographics, service quality, traffic density and operators' cost. We also find some evidence that the move from monopoly to duopoly may have a stronger negative effect on bus fares than the effect of any additional entry.
- C.3 Because of data limitations we are not able to say much about the nature of the competitive constraint imposed by rival operators on each other when they provide services in the same LTA. In particular, we have not been able to establish whether fares are lower when rivals are engaged in head to head competition, or when they serve overlapping inter-urban services, or indeed, as put to us by some operators, whether a mere presence of rival operators nearby act as an effective constraint on pricing.
- C.4 The operators have raised various issues about the quality of some of the underlying data which we used in our study. We have considered their points very carefully. After carrying out additional analysis we concluded that, on balance, our findings are sufficiently robust for

the purpose of a consulting on our proposed decision to make a market investigation reference to the Competition Commission.

C.5 The remainder of this annex is structured as follows:

- methodology
- data
- exploratory data analysis
- regression results and
- robustness and endogeneity checks.

## **Methodology**

C.6 For the purpose of testing the contestability hypothesis, we use an empirical technique called price-concentration analysis. This technique investigates the relationship between price and the market concentration using multiple regression analysis.

C.7 The basic intuition for using price-concentration analysis in this context is that if bus markets are contestable fares should be independent of market concentration. A finding that fares depend on concentration shows that markets are not perfectly contestable (it does not however rule out the possibility that potential competition exerts some influence on pricing).

C.8 Our measure of price is the average fare charged by the subsidiaries of national operators in LTAs across the country. We relate this variable to the number of national operators providing commercial services in a particular LTA. We also consider the impact of smaller operators on the fares charged by national operators.

C.9 When conducting a price-concentration study, in addition to the impact of market concentration which is the main focus of interest, it is important to control for factors which may affect firms' pricing behaviour. If these factors are omitted from the analysis, and they

are correlated with concentration measures, the regression of price on concentration variables may result in biased parameter estimates.

C.10 In deciding which control variables to consider in our analysis, we were guided by the literature and by bus operators' own views which we have elicited in the course of this study. Specifically, we consider the following factors:

- competition from the car
- quality of bus service
- local cost differences
- local socio-demographics
- traffic density and
- average journey distance.

C.11 A more detailed description of how we measure these variables is provided below at paragraphs C.22 to C.26.

## Data

### Sources

C.12 The main data used in our analysis comes from the STATS100 database held by the DfT. This database is compiled from financial year survey returns to the DfT from bus operators licensed for 20 or more vehicles. The principal aim of the survey is to collect information from the industry in order to monitor national and regional trends and inform policy development.

C.13 Permission was obtained from the four national operators to have access to their survey returns to the DfT.<sup>104</sup> The data contains information on passenger boardings and receipts, revenue support, concessionary fares and vehicle miles which is provided at LTA level.

---

<sup>104</sup> These are First, Stagecoach, Arriva and Go-Ahead.

The data also contains information on operators' fleet size, type of vehicles used, operating and administrative costs, depreciation and profitability.

- C.14 DfT told us that they have made some adjustments to the raw data supplied by the operators to account for under-recording of ticketless passengers by the operators. DfT also told us that although passenger boardings and revenue are collected at LTA level, the survey is designed to provide estimates which are robust at national and regional level and the figures at LTA level are of lesser quality.
- C.15 We have supplemented DfT's data with two additional data sources. We obtained information on competition from the car and socio-demographics at LTA level from the 2001 census data compiled by the Office for National Statistics (ONS) in England and Wales and by the General Register Office for Scotland in Scotland. We also obtained a list of names, addresses and the number of vehicles for all bus operators in the UK from Vehicles & Operator Service Agency (VOSA).

## **Variable definitions**

### **Dependent variable**

- C.16 As mentioned in paragraph C.8, our measure of price is the average fare charged by subsidiaries of national operators in LTAs across the country. We have calculated this measure by dividing total passenger receipts by our estimate of the number of paying passenger boardings which are reported by the operators to the DfT for each LTA in which they operate. We have based our analysis on the data for 2006/07.<sup>105</sup> This is the last year for which the revenue dataset was made available to us by the DfT.
- C.17 We did not have the corresponding paying passenger numbers as these were first reported by the operators in 2007/08. This meant that we had to make an estimate of the paying passenger boardings for 2006/07. We did this by calculating the proportion of the paying

---

<sup>105</sup> In 2006/07, data was reported for 133 LTAs outside London.

passenger boardings in the total passenger boardings in 2007/08 and then applying this proportion onto the total number of passenger boardings reported in 2006/07 to estimate the number of paying passenger boardings in 2006/07.

- C.18 This procedure may introduce measurement error in our analysis which under certain circumstance may cause parameter estimates to be biased.<sup>106</sup> We discuss this possibility, and explain how we dealt with it, at paragraphs C.35 to C.45 below.

### Concentration measures

- C.19 We use two measures of market concentration for national bus operators. One is a binary indicator variable which distinguishes between cases when there is only one national provider of commercial bus services in an LTA area, and cases where there are two or more national operators. The other measure is a count of national operators providing commercial services in a particular LTA. When there are two or more subsidiaries from the same parent company operating in the same LTA, we count this as only one competitor.
- C.20 We also consider whether the presence of smaller operators impacts on the fares charged by the subsidiaries of the four national operators. In order to allow for the possibility that the strength of a competitive constraint may be related to operator size, we use three different classifications of smaller operators:
- operators with more than 20 buses
  - operators with more than 40 buses and
  - operators with more than 100 buses.
- C.21 Within the timeframe available to us for this study, we were not able to establish which smaller operators on our list provide both

---

<sup>106</sup> If the error is correlated with the explanatory variables.

commercial and supported services, and which provide either commercial or supported services. We would expect those operators that provide commercial services to exert more pressure on national operators than those that only provide supported services. We also assumed that each small operator provides services in the LTA where its offices are registered, which may not always be the case. For these reasons, this aspect of our study needs to be treated with caution.

## Control variables

- C.22 As part of our industry consultation, we asked the subsidiaries of the national bus operators to rank various factors that influence their pricing. Although the answers varied across operators, the factors that were mentioned most frequently in addition to competition from other operators were competition from the car, income, local socio-demographics and route density.
- C.23 We took these factors into account by including in our regression the number of paying passenger boardings, the proportion of households with no car, the proportion of unemployed and the proportion of ABs<sup>107</sup> in the population for each LTA. The proportion of households with no car will not only capture the impact of competition from the car but it may also act as a proxy for income, together with unemployment and ABs (all those variables are highly correlated). Therefore, we do not interpret the coefficient associated with the proportion of households with no car as solely measuring the impact of competition from the car.
- C.24 We also considered aspects of service quality which may influence fares. Firstly, we took into account the number of commercial vehicles kilometres per paying passenger operated by national operators in each LTA. This is a commonly used proxy for service levels in the empirical literature. Secondly, we have considered in our regression a range of variables which relate to operators' fleets,

---

<sup>107</sup> ABs is a term used in consumer research for the purpose of socio-economic profiling. It includes persons in senior managerial, administrative or professional roles (group A) and those in a middle managerial role or equivalent (group B).

including the number of buses with a disability access certificate, low floors, GPS and CCTV equipment and the average fleet age. These variables are measured at the subsidiary level of operators rather than LTA level. Because an operator will provide services typically in more than one LTA, we are assuming implicitly that this operator will use the same vehicle types in all LTAs.

- C.25 We allowed for a possibility that operators' costs may vary regionally, and that this may have an impact on how they price their services locally. We took this into account by including the operating expenditure per kilometre operated into our regression. We also consider an alternative measure, namely operating expenditure per passenger boarding. We found that the two measures yield similar results. Again, these variables are measured at the level of operators' subsidiaries rather than LTA level. We do not perceive this to be a problem as the costs are unlikely to vary substantially between the neighbouring LTAs.
- C.26 Finally, the bus operators told us that fares may vary according to journey distance. We could not estimate the average journey distance from our data set. Instead we used several proxies which can be expected to correlate with journey distance, including LTA area, population density, type of LTA (for example, rural, urban) and UK region (for example, North West).

## **Geographic scope of analysis**

- C.27 Our analysis covers the whole of Great Britain except London and the West Midlands. The bus industry in London is not suitable for the price-concentration analysis because it is regulated with competitive forces introduced via a regime of bus route tendering. For the West Midlands, we are missing the biggest operator in the area - National Express - which did not consent for DfT to release their data to us. We were concerned that including the West Midlands in our analysis might bias our results and so we excluded it from our sample.

## Measurement errors

- C.28 In addition to obtaining operators' data from DfT, we also received revenue and traffic data directly from the operators. We could not use this information for our price-concentration analysis because different operators record revenue and traffic data at different levels of aggregation. For example, First's data are split into 18 fairly large regional units (for example, Midlands, Potteries) whereas the data from Stagecoach are reported in a much more granular form (more than 100 reporting units). Within the timeframe which was available to us for this study, it would not be possible for us to express these data at the same level of aggregation.
- C.29 We have used the operators' data to carry out validity checks on the DfT's data. There were several instances where the estimates of the average fare based on DfT's data appeared too high or too low relative to the data provided by the operators. In particular, we excluded from our data set any observations where the average fare was less than 50p or more than £3.50.<sup>108</sup> This reduced the size of our sample from 222 observations to 208 observations. This is a cautious approach as leaving those observations in the data set results in an even higher estimate of the impact of market concentration on fares.
- C.30 We have also asked the operators to examine DfT's data and let us know whether they are any material errors which should be corrected. In addition we sought their views regarding the method which we used to estimate the number of paying passenger boardings for 2006/07 and which is described at paragraph C.17.
- C.31 Stagecoach told us that they were comfortable with the revenue figures that were provided to us by DfT but they expressed concerns about data consistency and the potential impact of changes in concessionary fares scheme on our analysis. This is discussed in more detail below at paragraphs C.35 to C.45. We have not received any comments from Go-Ahead.

---

<sup>108</sup> We used the figures in the Stagecoach dataset to identify these observations as outliers because the reporting units for Stagecoach are more similar to the way that DfT's data are reported.

- C.32 Arriva noted that virtually all of its concessionary boarding numbers for 2007/08 were increased by DfT and that this increase appears to be broadly uniform within DfT's regions. Similar comment was made by First. At paragraph C.14, we note that DfT makes adjustments to the raw data supplied by the operators to account for under-recording of ticketless passengers by the operators. Therefore, in our view the figures adjusted by DfT are likely to be a more accurate representation of the correct number of concessionary passengers than the operators' own unadjusted figures.
- C.33 Arriva told us that there were three data points in DfT's data set purporting to show Arriva's traffic and revenue in 2006/07 which Arriva had not supplied to DfT. Arriva suggested that we remove these data points from our data set. First identified several examples where the dataset which was given to us by DfT may be inaccurate. We controlled for the impact of those observations and re-estimated our regressions. This has not altered our main conclusion materially, as explained at paragraphs C.76, though it had an effect on parameter estimates.
- C.34 First suggested that we should use the total local registered service vehicle kilometres in our analysis rather than the commercial services kilometres. We took First's comments into account and re-estimated our model using total instead of commercial kilometres per passenger. This has resulted in a bigger size of the estimate of the relationship between price and concentration as described at paragraph C.75 below.
- C.35 First, Stagecoach and Arriva expressed concerns over our use of the proportion of concessionary boardings in 2007/08 to estimate the number of paying and concessionary passenger boardings in 2006/07. They pointed out that the proportion of concessionary passenger boardings has been growing substantially in recent years since the scheme first became free in England in 2006.
- C.36 If so, applying the proportion of concessionary passenger boardings in 2007/08 to estimate the number of concessionary passenger boardings in 2006/07 from the total number of passenger boardings in 2006/07 may result in an overestimate of concessionary

passengers boardings (or equivalently an underestimate of the paying passenger boardings). This, in turn, may result in an overestimate of the average fare which is calculated by dividing passenger receipts by the number of paying passenger boardings, as explained at paragraphs C.16 to C.17.

- C.37 First also suggested that we include youth concessionary boardings and revenue into our calculations of the average fare since not all youth concessions are free and different LTAs have different policies in place with respect to young people. First pointed out that the Scotland-wide concessionary scheme for young people started during 2006/07 in January 2007 and that because of that the proportion of youth concessions in 2007/08 will not be representative of 2006/07 for Scotland.
- C.38 We were aware of the change in the concessionary regime travel in England which took place in April 2006 and for this reason we only used data for the years 2006/07 and 2007/08 during which the scheme was already in place throughout the UK so as to minimise any distortion in our calculation of the average fare. Furthermore, for those two years (that is, 2006/07 and 2007/08) we have compared the proportion of the concessionary passenger boardings in the data which was made available to us by the operators. The comparison suggests that overall the proportion of concessionary boardings has changed very little between 2006/07 and 2007/08.
- C.39 Specifically, for First, Stagecoach and Go-ahead<sup>109</sup> there was a one percentage point increase in the proportion of concessionary passenger boardings between 2006/07 and 2007/08 when the data are aggregated across reporting units. For Arriva, the increase is somewhat higher – approximately three percentage points. There is more variation at an individual area level, but we have no reasons to believe that this variation is related to our concentration measures which would cause our estimates to be biased.
- C.40 This notwithstanding, we have investigated what impact any incorrect measurement of the number of concessionary passenger boardings may have on our regression estimates. Specifically, we

---

<sup>109</sup> Go-Ahead was unable to provide data for all their subsidiaries.

included in our regression as an explanatory variable the proportion of population eligible for concessionary fares. We expect that any underestimate of the average fare (by virtue of overestimating the number of concessionary passengers) will be more significant in the areas where a higher proportion of the population is eligible for concessionary fares.

- C.41 When we include this explanatory variable in our regression, we find that the average fare is indeed higher in those areas where a higher proportion of the population is eligible for concessionary fares. This may be due to the error in measurement of concessionary passenger boardings along the lines suggested above. However, it could also be that operators' incentives when setting fares for paying passengers may be affected by the fact that these fares are used as a basis for setting concessionary fares which operators recover from LTAs.
- C.42 Specifically, increasing fares to paying passengers will lead to a reduction in the number of trips which, in turn, may lower the total revenue from paying passengers if demand is price elastic. However, increasing fares to paying passengers may also increase the concessionary revenue from LTAs if this increase is factored into concessionary fares. Arguably, the second effect could be greater as passengers eligible for free travel will not be deterred from travelling by higher costs of non-concessionary travel.
- C.43 Importantly, even when we include this variable (that is, the proportion of passengers eligible for concessionary fares) in our model, we still find that fares are higher in LTAs where there is only one provider of commercial services. This is described at paragraph C.77 below.
- C.44 More generally, the operators told us that they do not record their revenue and traffic data at LTA level, which is how DfT's dataset is structured, and that each subsidiary may use a different method to provide its data to DfT in this format. The operators told us that despite their best endeavours there is inevitably an element of arbitrariness involved in this process which means that we should treat their data with caution when considering it at LTA level.

- C.45 While this may be the case, there is nothing to suggest that any errors from this will be related to the concentration measures. Consequently, we would not expect our estimates to be biased though they may have larger variances.

### **Endogeneity of explanatory variables**

- C.46 A number of econometric issues arise when conducting a price concentration study, the most obvious and difficult of which is the possible endogeneity of some of the explanatory variables. Endogeneity describes a situation where an explanatory variable is correlated with the error term in which case parameter estimate may be biased. In addition to the measurement error bias, which is discussed in the previous section, there are several other potential sources of endogeneity in our study.
- C.47 One is the use of traffic volumes as an explanatory variable. The use of this variable is justified in transport economics with the reference to the effect of economies of density where higher levels of traffic result in lower marginal costs and therefore fares. Failure to find such an effect would indicate that economies of density are inconsequential in magnitude and/or that there are other offsetting effects such as capacity shortage or an exercise of market power on part of operators.
- C.48 Irrespective of how the use of traffic volumes in the price equation is motivated, there is likely to be a feedback effect from fares to traffic in which case the traffic variable will be endogenous. That said, because our data is aggregated at the level of LTA we do not expect this endogeneity to be as problematic as it would be if we used a route level data.
- C.49 The other, potentially more important source of endogeneity is possible correlation of market structure with the error term. In particular, the market structure may be an outcome of strategic decision by firms that evaluate demand and cost conditions as well as potential competitors in their entry decision. As a consequence, there might be unobserved demand and cost shocks in a market that

influence not only prices but also the underlying market structure. If this is not taken into account, the parameter estimates may be biased.

- C.50 The direction of the bias will depend on the nature of the prevailing shocks. For example, the markets with unobserved high costs are likely to have higher prices, but these markets are also likely to attract fewer entrants. In this case, a regression of prices on the number of competitors will overstate the impact of market structure. In contrast, positive demand shocks will understate the impact of market structure on the price as they will result in higher prices and higher number of firms.
- C.51 Our view is that this type of endogeneity may not be serious. As noted in the main body of our report (see paragraph 4.11[of the Proposed Decision]), the current market structure in terms of geographic location, market concentration and the identities of the major players has not changed significantly in the last ten years and largely reflects historic pattern of depot ownership. Thus, an argument can be made that the market structure in the bus industry is largely pre-determined.
- C.52 This notwithstanding, we have explored the issue of endogeneity empirically by employing an instrumental variable (IV) estimator. The IV methodology requires finding instrumental variables that are correlated with endogenous explanatory variables (in our case market concentration and traffic density) but are not directly related to prices. We have identified several variables that may satisfy these conditions though we cannot rule out categorically the possibility that some of them may in fact be covariates in their own right.<sup>110</sup>
- C.53 Specifically we consider two types of variables as potential instruments for both market structure and traffic density. The first are variables that affect the size of the market such as area population. The size of the market should be correlated with bus

---

<sup>110</sup> This issue is generic to IV estimation and not specifically an issue for this study. All IV estimation faces the same trade-off: a 'good' instrument (that is, one that is highly correlated with the endogenous independent variable) should probably be an independent variable; but a 'weak' instrument (that is, one that is not strongly correlated with the endogenous independent variable) may not control for the endogeneity.

operators' desire to enter but should not affect the degree of competition between them once they have. So the size of the market should be correlated with concentration but not with price.

- C.54 The second group are the variables which relate to the geography of the UK (UK regions and LTA types). Their use as potential instruments can be motivated by an observation that local geography, administration and historic pattern of depot ownership appears to be conducive to a particular market structure. For example, some areas of the country (the North East and the North West) have a significant presence of three national operators. Other areas, such as the East of England and the South West, are predominately served by one national operator (See Figure 4.1 [in the Proposed Decision]).
- C.55 The results of the IV estimation are discussed at paragraph C.80 to C.83 below. They suggest that our ordinary least squares (OLS) estimates may in fact understate the relationship between market concentration and price by about factor of three. However, we have concerns over both the validity and the strength of our instruments and so we treat this result with caution.

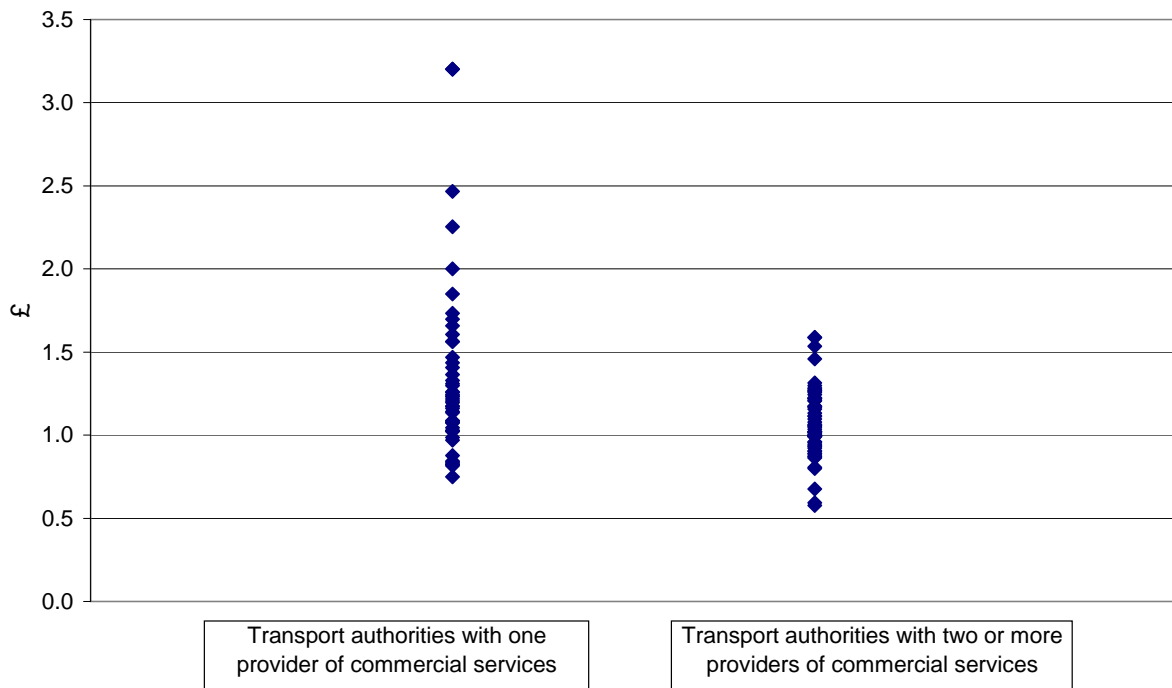
## **Exploratory data analysis**

- C.56 In this section we examine how the average fare paid by passengers varies across LTAs.<sup>111</sup> For the purpose of this exercise we have aggregated operators' revenue and boardings data at LTA level. We have not identified LTAs in the figures below as the operators' data used to calculate the average fare is confidential.
- C.57 Figure C.1 below plots the average fare against an indicator variable which distinguishes between cases where there is only one national operator of commercial services in an LTA and cases where there are two or more national operators.

---

<sup>111</sup> Note that in our regression analysis in the following section we focus on the average fare charged by individual operators.

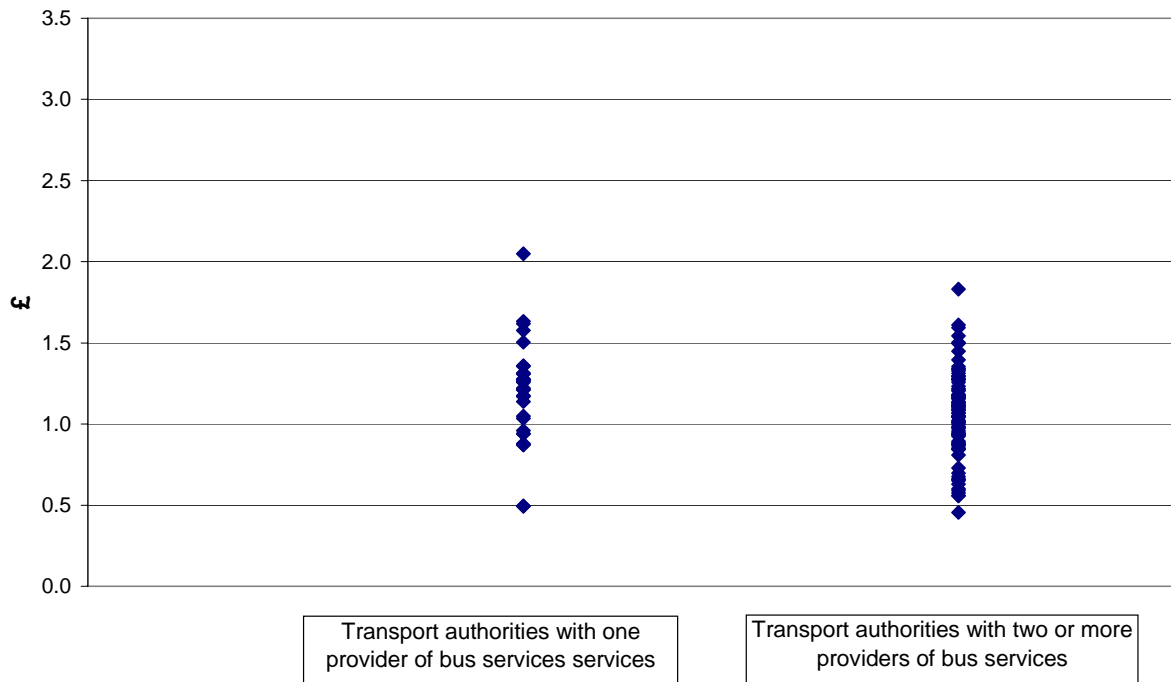
**Figure C.1: Average fare and the number of national operators providing commercial services (based on fares charged by national operators)**



C.58 Overall, the data suggest that the average fare is higher in LTAs where there is only one national operator of commercial services than where there are two or more national operators providing commercial services. There is a statistically significant difference in the mean between the two groups: £1.35 for LTAs with one provider and £1.07 where there are two or more providers.

C.59 Figure C.2 repeats this analysis, this time plotting the total revenue per passenger (that is, including concessionary revenue and passengers in the comparison) against the number of national operators providing either commercial or supported services, or both.

**Figure C.2: Total revenue per passenger and the number of national operators providing local bus services (based on revenue earned by national operators)**



C.60 Even though there is less dispersion in the data than in Figure C.1, Figure C.2 gives the same impression – overall the revenue per passenger appears to be higher in the areas with one national provider than in the areas where bus services are provided by two or more national operators. Again, there is a statistically significant difference in the mean between the two groups: £1.19 for LTAs with one operators and £1.08 where there are two or more operators.

## Regression analysis

### Impact of market structure on bus fares

C.61 In this section we examine how fares charged by national operators vary according to the number of operators providing commercial services in an LTA. We control for a range of factors which may influence pricing of bus services; these factors are discussed at paragraphs C.22 to C.26 above. We start from the following baseline specification:

$$p_{ij} = \alpha + f(\alpha; MS_j) + \beta X_j + \gamma W_{ij} + \delta Z_j + \epsilon_{ij}$$

where  $p_{ij}$  is the average fare charged by operator  $i$  in LTA  $j$ ,  $f(\cdot)$  is the function which captures the effect of the market structure in LTA  $j$ ,  $X$  is a vector containing control variables measured at LTA level (socio-demographics and car ownership),  $W$  is a vector containing control variables measured at operator subsidiary/LTA level (paying passenger boardings and commercial kilometres per paying passenger boarding) and  $Z$  is a vector containing control variables measured at operator subsidiary level (costs and fleet characteristics).

- C.62 We have estimated several regression models using different combinations of variables, data transformations and functional forms. Irrespective of which model specification is used, the parameter estimate associated with the variable indicating that there is only one national provider of commercial services is statistically significant and suggests that a monopoly operator charges higher fares than when there are two or more national operators providing commercial services.
- C.63 Table C.3 below shows the results of one such model (henceforth 'baseline model') which relates the average fare to LTA socio-demographics, the proportion of households with no car, the number of commercial kilometres per paying passenger, operating costs per kilometre, paying passenger boardings and to our two measures of concentration. One is a binary indicator variable which distinguishes between cases where there is one national operator providing commercial services in a LTA and cases where there are two or more such operators. The other measure is a count of small operators with 20 or more buses. All the variables except the ones that measure concentration are transformed logarithmically.

**Table C.3: Impact of market structure on bus fares**

explanatory variable	parameter estimate	Standard error	p-value	95 per cent confidence intervals	
single national operator	0.141**	0.048	0.004	0.046	0.236
count of smaller operators (20 or more buses)	-0.007	0.008	0.382	-0.023	0.009
km per passenger	0.086**	0.025	0.001	0.036	0.135
proportion of households with no car	0.235**	0.106	0.027	0.026	0.444
proportion of ABs	0.085**	0.039	0.031	0.008	0.163
proportion of unemployed	-0.294**	0.117	0.013	-0.525	-0.063
operating cost per km	0.097**	0.048	0.046	0.002	0.192
paying passenger boardings	-0.001	0.008	0.942	-0.017	0.016
constant	-0.687**	0.299	0.023	-1.277	-0.096
R <sup>2</sup> = 0.21					

\*significant at 10 per cent, \*\* significant at 5 per cent

C.64 Our interest is in the parameters associated with the 'single national operator' (henceforth 'monopoly variable') and the 'count of small operators' variables. The 95 per cent confidence intervals for the parameter estimate associate with the monopoly variable are 0.046 and 0.236. The point estimate implies that the average fare is approximately 15 per cent higher when there is only one national provider of commercial services than when there are two or more national providers.<sup>112</sup> The parameter estimate associated with the 'count of small operators' is negative but not statistically significant. All other parameters are statistically significant and have expected signs except the parameter estimate associated with the 'paying passenger boardings' variable.

C.65 We have estimated the same model in the levels (that is, without logarithmic transformation of the data) and this has produced similar results. The parameter estimate associated with the monopoly variable is both statistically significant and sizeable: it suggests that fares are higher by about 20p where there is only one national provider of commercial services (see Table C.6). The estimates of the

<sup>112</sup> The percentage impact is given by inserting the value of the parameter estimate and its variance into the following expression:  $g = 100 (\exp(b - V(b)/2) - 1)$ .

levels regression also suggest that small operators exert competitive pressure on the national operators though the size of the impact is small (each additional competitor reduces fares by about 2p) and only significant at the 10 per cent significance level.

- C.66 We have also estimated several regressions using different ways of measuring the presence of smaller operators which are explained at paragraph C.20 above. However, none of those regressions suggest that smaller operators impose significant constraint on the pricing of the national operators (see Table C.7). One regression suggests that each additional small operator reduces fares charged by national operators by about three per cent. This result is only significant at the 10 per cent significance level.
- C.67 We found that the number of kilometres per passenger, which we use to proxy service levels, is statistically significant (see Table C.3 above). Other proxies for service levels which relate to operators' fleet - the average fleet age, number of buses with low floors, GPS and CCTV equipment - were either of the wrong sign or statistically insignificant. The results of those regressions are reported in the panels of Table C.8. Importantly, including these variables in our regression does not affect materially the parameter estimates related to the concentration variables.
- C.68 The results of the regressions including proxies for journey distance as explanatory variables are shown in the panels of Table C.9). Our preferred proxies for journey distance - area size and population density - were not statistically significant. The variables representing different LTA types (for example, rural, urban) were statistically significant but correlated strongly with the car ownership and so the associated parameter estimates need to be treated with caution. Again, the results with respect to the concentration variables remain broadly similar to those shown in Table C.3 above. The model with UK regions also produced similar results to the model shown in Table C.3.

## Impact of additional competitor

C.69 We have also considered how fares vary in relation to the number of national operators providing commercial services in an LTA area. In the regression described in the previous section we used a binary indicator variable which distinguishes between monopoly market structure and the case where there are two or more national operators providing commercial services. In the example below we use the actual number of competitors while retaining the same control variables as in the regression shown in Table C.3 above. The parameter estimates are presented in Table C.4 below.

**Table C.4: Impact of additional competitor on bus fares – linear functional form in the number of competitors**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
count of national operators	-0.097**	0.031	0.002	-0.157	-0.037
count of smaller operators (20 or more buses)	-0.010	0.008	0.217	-0.026	0.006
km per passenger	0.082**	0.025	0.001	0.032	0.131
proportion of households with no car	0.220**	0.106	0.038	0.012	0.428
proportion of ABs	0.097**	0.040	0.016	0.018	0.176
proportion of unemployed	-0.254**	0.118	0.032	-0.486	-0.022
operating cost per km	0.101**	0.048	0.037	0.006	0.196
paying passenger boardings	-0.003	0.008	0.720	-0.020	0.014
constant	-0.730**	0.300	0.016	-1.323	-0.138
R <sup>2</sup> = 0.18					

\*significant at 10 per cent, \*\* significant at 5 per cent

C.70 Our interest is in the 'count-of-national-operators' variable which shows that each additional national operator reduces prices by about 10 per cent. Other parameter estimates are of the same order of magnitude as the corresponding parameter estimates shown in Table C.3

C.71 The results presented in Table C.4 are produced using the model where local concentration is described by a linear function in the number of service providers. This is a restrictive functional form whereby the rate of change in the dependent variable produced by changes in the independent variable does not vary with the values of the independent variable.

C.72 This approach is not in general supported by the academic literature, which suggests that the effect of the first additional competitor - the move from monopoly to duopoly – typically has a stronger effect than that of any further competitor. To test this empirically, we have created separate binary indicator variables for each competitor count and omitted the one for monopoly. This way the parameter estimates measure the impact of a move from monopoly to the market structure in question (that is, duopoly, three and four competitors). The results are presented in Table C.5 below.

Table C.5: Impact of additional competitor on bus fares – separate dummies for each market structure

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
two national operators	-0.132**	0.048	0.007	-0.227	-0.037
three national operators	-0.224**	0.070	0.002	-0.361	-0.086
four national operators	-0.226**	0.111	0.043	-0.445	-0.008
count of smaller operators (20 or more buses)	-0.009	0.008	0.302	-0.025	0.008
km per passenger	0.080**	0.025	0.002	0.030	0.130
proportion of households with no car	0.235**	0.107	0.029	0.025	0.446
proportion of ABs	0.095**	0.041	0.020	0.015	0.175
proportion of unemployed	-0.267**	0.118	0.025	-0.501	-0.034
operating cost per km	0.102**	0.048	0.035	0.007	0.198
paying passenger boardings	-0.003	0.008	0.726	-0.020	0.014
constant	-0.829	0.327	0.012	-1.474	-0.185
R <sup>2</sup> = 0.19					

\*significant at 10 per cent, \*\* significant at 5 per cent

C.73 The parameter estimates support the notion that the move from monopoly to duopoly has a stronger effect than that of any further

national competitor. Specifically, they imply that duopoly fares are about 14 per cent lower than monopoly fares whereas the move from monopoly to three national operators reduces fares by about 23 per cent (so the entry of a third operator reduces fares by about nine percent). Adding the fourth competitor does not make any difference.

## **Robustness and endogeneity tests**

- C.74 In this section we present the results of the robustness tests which we have carried out in order to address the various concerns about the quality of our data which are discussed at paragraphs C.28 to C.45.
- C.75 First recommended that we use total local registered service vehicle kilometres in our analysis rather than commercial services kilometres. We considered this suggestion and re-estimated our model accordingly. The parameter estimate associated with total kilometres per passenger is not significant, and the parameter estimate associated with the monopoly variable is even bigger in size relative to the estimate shown in Table C.3. It suggests that the average fare is approximately 18 per cent higher when there is only one national provider of commercial services than in the case where there are two or more national providers. The results of this regression are presented in Table C.10.
- C.76 First and Arriva told us that there were several observations where DfT's data may be inaccurate. We have assigned dummy variables to these observations ('first dummy' and 'arriva dummy') and re-estimated the baseline model. The results are shown in Table C.11. It can be seen that the parameter estimate associated with the monopoly variable is statistically significant but smaller than the corresponding parameter estimate shown in Table C.3. It suggests that the average fare in monopoly areas is higher by about 11 per cent than in LTAs where there are two or more national competitors.
- C.77 The results of the regression which includes the proportion of the population eligible for concessionary fare as an explanatory factor are presented in Table C.12 (there are two variables, one relating to people aged over 60, the other to people aged 14 or less). It can be

seen that the fares are higher in LTAs where there is proportionately more passengers eligible for concessionary fares, though the individual parameters are not estimated precisely. Importantly, the parameter estimate associated with the monopoly variable is still significant at five per cent significance level and of the similar size to the one presented in Table C.3.

- C.78 We also estimated the regression where we controlled for the data points which according to the operators may be inaccurate and included the proportion of the population eligible for concessionary fares as explanatory variables. The results of this regression are presented in Table C.13. The parameter estimate associated with the monopoly variable is significant at the 10 per cent significance level (with a p value of 0.061). It suggests that the fares in monopoly areas are higher by about nine per cent than in LTAs where there are two or more national competitors.
- C.79 We then estimated the same model for the specification with the count-of-national-operators variable instead of the monopoly variable. The results which are presented in Table C.14 are similar to the results shown in Table C.5 above. The parameter estimate associated with the count of national operators is statistically significant at the five per cent significance level and suggests that each additional entrant reduces fares by about nine per cent.
- C.80 Finally, we estimated the model described in paragraph C.61 using the instrumental variable approach. Specifically, we consider the possibility that our measures of concentration and our traffic density variable are endogenous, and instrument these variables with population and the selection of dummy variables denoting types of LTA and the UK regions. Because we have more instruments than endogenous variables, the model is estimated using the two-stage least squares method (2SLS). The results together with the output for the first-stage regressions of the endogenous variables on the instruments and other explanatory variables are presented in Table C.15.

- C.81 The results suggest that the prices in monopoly areas are higher by about 30 per cent than in LTAs where there are two or more national operators. This is three times the size of the impact estimated using the OLS regression in Table C.13. The parameter estimates associated with the number of small competitors and traffic density are not statistically significant.
- C.82 However, in the first stage regression results, we cannot reject the null hypothesis of underidentification with respect to traffic density.<sup>113</sup> Furthermore, the Sargan statistics rejects the null hypothesis with a p value of 0.053 that the instruments are uncorrelated with the error term and/or that the model is misspecified.
- C.83 When we exclude traffic density and re-estimate the model, we obtain similar parameter estimates for our concentration variables as described in paragraph C.81. The Sargan statistics is better this time but only marginally so (with the p value of 0.092) which means that we cannot rule out the possibility that our instruments may be correlated with the error term.

**Table C.6: Impact of market structure on bus fares (levels regression)**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.209	0.059	0.000	0.093	0.324
count of smaller operators (20 or more buses)	-0.018	0.010	0.079	-0.039	0.002
km per passenger	0.032	0.011	0.003	0.011	0.053
proportion of households with no car	2.430E-06	2.430E-06	0.317	0.000	0.000
proportion of ABs	1.360E-06	5.640E-07	0.017	0.000	0.000
proportion of unemployed	-1.170E-05	1.300E-05	0.370	0.000	0.000
operating cost per km	0.057	0.030	0.060	-0.002	0.116
paying passenger boardings	-4.860E-09	2.400E-09	0.045	0.000	0.000
Constant	0.940	0.068	0.000	0.807	1.074
R <sup>2</sup> = 0.18					

<sup>113</sup> The partial R<sup>2</sup> is very small and the F version of the same test suggests that the parameter estimates associated with the instruments are not statistically significant in the equation where traffic density is the dependent variable.

**Table C.7: Impact of smaller operators on fares charged by national operators**

a) Small operators with 40 or more buses

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.140	0.048	0.004	0.046	0.234
count of smaller operators (40 or more buses)	-0.027	0.015	0.074	-0.056	0.003
km per passenger	0.085	0.025	0.001	0.036	0.134
proportion of households with no car	0.228	0.105	0.030	0.022	0.435
proportion of ABs	0.085	0.037	0.024	0.011	0.158
proportion of unemployed	-0.281	0.116	0.016	-0.509	-0.052
operating cost per km	0.096	0.048	0.047	0.001	0.190
paying passenger boardings	-0.002	0.008	0.852	-0.018	0.015
constant	-0.703	0.270	0.010	-1.235	-0.171
R <sup>2</sup> = 0.19					

b) Small operators with 100 or more buses

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.141	0.048	0.004	0.046	0.235
count of smaller operators (100 or more buses)	-0.025	0.028	0.383	-0.081	0.031
km per passenger	0.084	0.025	0.001	0.035	0.134
proportion of households with no car	0.249	0.105	0.018	0.043	0.456
proportion of ABs	0.076	0.037	0.042	0.003	0.149
proportion of unemployed	-0.311	0.115	0.008	-0.538	-0.084
operating cost km	0.095	0.048	0.050	0.000	0.191
paying passenger boardings	-0.001	0.008	0.896	-0.018	0.015
constant	-0.584	0.260	0.026	-1.098	-0.070
R <sup>2</sup> = 0.18					

**Table C.8: Fares and service quality**

a) Proportion of buses with disability access certificate

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.123	0.048	0.011	0.028	0.217
count of smaller operators (20 or more buses)	-0.008	0.008	0.342	-0.024	0.008
km per passenger	0.096	0.025	0.000	0.046	0.145
proportion of households with no car	0.209	0.105	0.047	0.003	0.416
proportion of ABs	0.073	0.039	0.064	-0.004	0.150
proportion of unemployed	-0.254	0.117	0.030	-0.485	-0.024
operating cost per km	0.124	0.049	0.012	0.028	0.220
paying passenger boardings	0.000	0.008	0.992	-0.016	0.016
proportion of buses with disability access certificate	-0.330	0.131	0.013	-0.589	-0.071
constant	-0.570	0.299	0.058	-1.160	0.019
$R^2 = 0.17$					

b) Proportion of buses with low floor access

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.151	0.047	0.002	0.058	0.245
count of smaller operators (20 or more buses)	-0.006	0.008	0.451	-0.022	0.010
km per passenger	0.079	0.025	0.002	0.030	0.128
proportion of households with no car	0.306	0.107	0.005	0.095	0.518
proportion of ABs	0.086	0.039	0.027	0.010	0.163
proportion of unemployed	-0.368	0.118	0.002	-0.602	-0.134
operating cost per km	0.075	0.048	0.122	-0.020	0.170
paying passenger boardings	-0.003	0.008	0.736	-0.019	0.014
proportion of buses with low floors access	-0.368	0.134	0.007	-0.632	-0.103
constant	-0.674	0.295	0.023	-1.255	-0.093
R <sup>2</sup> = 0.20					

c) Proportion of buses with GPS

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.132	0.047	0.005	0.039	0.224
count of smaller operators (20 or more buses)	-0.005	0.008	0.561	-0.020	0.011
km per passenger	0.093	0.025	0.000	0.044	0.141
proportion of households with no car	0.227	0.103	0.029	0.024	0.429
proportion of ABs	0.094	0.038	0.015	0.018	0.170
proportion of unemployed	-0.297	0.114	0.010	-0.521	-0.072
operating cost per km	0.153	0.049	0.002	0.055	0.250
paying passenger boardings	0.003	0.008	0.713	-0.013	0.019
proportion of buses with GPS	-0.257	0.073	0.000	-0.400	-0.114
constant	-0.654	0.291	0.026	-1.228	-0.080
R <sup>2</sup> = 0.23					

d) Proportion of buses with CCTV

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.129	0.047	0.006	0.037	0.222
count of smaller operators (20 or more buses)	-0.008	0.008	0.310	-0.024	0.008
km per passenger	0.081	0.025	0.001	0.033	0.129
proportion of households with no car	0.214	0.103	0.040	0.010	0.417
proportion of ABs	0.070	0.039	0.070	-0.006	0.146
proportion of unemployed	-0.247	0.115	0.033	-0.474	-0.020
operating cost per km	0.097	0.047	0.041	0.004	0.189
paying passenger boardings	-0.001	0.008	0.911	-0.017	0.015
proportion of buses with CCTV	-0.337	0.099	0.001	-0.531	-0.142
constant	-0.610	0.293	0.038	-1.186	-0.033
R <sup>2</sup> = 0.22					

e) Average fleet age

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.142	0.048	0.004	0.047	0.237
count of smaller operators (20 or more buses)	-0.007	0.008	0.382	-0.023	0.009
km per passenger	0.085	0.025	0.001	0.036	0.135
proportion of households with no car	0.233	0.106	0.029	0.024	0.443
proportion of ABs	0.083	0.041	0.043	0.003	0.163
proportion of unemployed	-0.290	0.118	0.015	-0.523	-0.056
operating cost per km	0.096	0.048	0.049	0.000	0.192
paying passenger boardings	-0.001	0.008	0.943	-0.017	0.016
average fleet age	-0.031	0.101	0.759	-0.231	0.168
constant	-0.606	0.398	0.129	-1.391	0.179
R <sup>2</sup> = 0.18					

**Table C.9: Impact of journey distance**

a) Population density

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.141	0.048	0.004	0.046	0.235
count of smaller operators (20 or more buses)	-0.005	0.008	0.550	-0.021	0.011
km per passenger	0.085	0.025	0.001	0.036	0.135
proportion of households with no car	0.218	0.106	0.042	0.008	0.427
proportion of ABs	0.091	0.039	0.023	0.013	0.168
proportion of unemployed	-0.292	0.117	0.013	-0.523	-0.062
operating cost km	0.083	0.049	0.092	-0.014	0.180
population density	-0.002	0.008	0.846	-0.018	0.015
paying passenger boardings	0.019	0.013	0.166	-0.008	0.045
constant	-0.592	0.306	0.055	-1.196	0.012
R <sup>2</sup> = 0.23					

b) Area size

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.141	0.048	0.004	0.046	0.235
count of smaller operators (20 or more buses)	-0.005	0.008	0.551	-0.021	0.011
km per passenger	0.085	0.025	0.001	0.036	0.135
proportion of households with no car	0.221	0.106	0.039	0.012	0.431
proportion of ABs	0.099	0.041	0.016	0.019	0.180
proportion of unemployed	-0.287	0.117	0.015	-0.518	-0.057
operating cost km	0.085	0.049	0.086	-0.012	0.182
area size	-0.002	0.008	0.854	-0.018	0.015
paying passenger boardings	-0.017	0.013	0.210	-0.043	0.009
constant	-0.554	0.317	0.082	-1.179	0.072
R <sup>2</sup> = 0.19					

c) Type of LTA

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.142	0.049	0.004	0.046	0.238
count of smaller operators (20 or more buses)	0.001	0.009	0.895	-0.016	0.018
km per passenger	0.089	0.025	0.001	0.039	0.138
proportion of households with no car	0.116	0.117	0.319	-0.113	0.346
proportion of ABs	0.109	0.043	0.012	0.024	0.195
proportion of unemployed	-0.251	0.124	0.044	-0.495	-0.007
operating cost km	0.094	0.048	0.054	-0.002	0.189
paying passenger boardings	-0.003	0.008	0.742	-0.019	0.014
large urban	-0.190	0.100	0.059	-0.387	0.007
medium urban	-0.267	0.102	0.010	-0.469	-0.066
mixed urban/rural	-0.157	0.092	0.092	-0.339	0.026
mostly rural	-0.312	0.092	0.001	-0.494	-0.131
mostly urban	-0.204	0.088	0.021	-0.377	-0.031
other large conurbations	-0.119	0.091	0.192	-0.298	0.060
constant	0.099	0.390	0.800	-0.669	0.867
R <sup>2</sup> = 0.27					

Baseline transport authority dummy is PTE. The parameter estimates associated with other categories give price relative to the price in PTE.

d) UK regions

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.142	0.052	0.007	0.040	0.244
count of smaller operators (20 or more buses)	-0.005	0.009	0.561	-0.023	0.013
km per passenger	0.079	0.026	0.003	0.028	0.130
proportion of households with no car	0.150	0.111	0.179	-0.070	0.370
proportion of ABs	0.032	0.047	0.496	-0.061	0.125
proportion of unemployed	-0.147	0.127	0.248	-0.397	0.103
operating cost km	0.063	0.053	0.237	-0.042	0.168
paying passenger boardings	-0.002	0.009	0.829	-0.019	0.015
East of England	0.189	0.090	0.038	0.011	0.367
North East	0.027	0.096	0.778	-0.163	0.217
North West	0.094	0.088	0.288	-0.080	0.267
Scotland	0.164	0.087	0.062	-0.008	0.336
South East	0.212	0.086	0.015	0.042	0.381
South West	0.231	0.089	0.010	0.055	0.407
Wales	0.087	0.091	0.338	-0.092	0.266
Yorkshire and the Humber	0.121	0.085	0.158	-0.047	0.289
constant	-0.650	0.323	0.045	-1.286	-0.014
R <sup>2</sup> = 0.23					

\*Baseline regional dummy is East Midlands. The parameter estimates associated with other categories give price relative to the price in PTE.

**Table C.10: Using total km per passenger instead of commercial km per passenger as proxy for service levels**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.178	0.048	0.000	0.083	0.273
count of smaller operators (20 or more buses)	-0.009	0.008	0.278	-0.026	0.007
total km per passenger	0.015	0.032	0.644	-0.048	0.078
proportion of households with no car	0.276	0.108	0.012	0.062	0.489
proportion of ABs	0.105	0.041	0.011	0.024	0.185
proportion of unemployed	-0.337	0.120	0.005	-0.574	-0.101
operating cost per km	0.052	0.049	0.295	-0.046	0.149
paying passenger boardings	-0.008	0.009	0.353	-0.026	0.009
constant	-0.809	0.307	0.009	-1.414	-0.204
R <sup>2</sup> = 0.13					

**Table C.11: Controlling for observations identified by First and Arriva as potentially inaccurate**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.109	0.046	0.020	0.017	0.200
count of smaller operators (20 or more buses)	-0.002	0.008	0.775	-0.018	0.013
km per passenger	0.105	0.024	0.000	0.057	0.153
proportion of households with no car	0.305	0.102	0.003	0.103	0.506
proportion of ABs	0.040	0.039	0.306	-0.037	0.117
proportion of unemployed	-0.329	0.112	0.004	-0.551	-0.108
operating cost per km	0.072	0.046	0.123	-0.020	0.163
paying passenger boardings	0.002	0.008	0.837	-0.014	0.017
first dummy	0.458	0.102	0.000	0.257	0.659
arriva dummy	-0.262	0.243	0.283	-0.741	0.218
constant	-0.651	0.286	0.024	-1.214	-0.087
R <sup>2</sup> = 0.26					

**Table C.12: Including proportion of population eligible for concessionary fares**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.137	0.049	0.005	0.041	0.233
count of smaller operators (20 or more buses)	-0.009	0.008	0.307	-0.025	0.008
km per passenger	0.084	0.025	0.001	0.035	0.133
proportion of households with no car	0.329	0.118	0.006	0.096	0.562
proportion of ABs	0.091	0.040	0.022	0.013	0.169
proportion of unemployed	-0.395	0.130	0.003	-0.651	-0.139
operating cost per km	0.104	0.048	0.033	0.009	0.199
paying passengers boardings	-0.001	0.008	0.946	-0.017	0.016
proportion of population under 15	0.653	0.375	0.084	-0.087	1.393
proportion of population over 60	0.158	0.179	0.380	-0.196	0.511
constant	0.513	0.822	0.533	-1.109	2.135
R <sup>2</sup> = 0.19					

**Table C.13: Including proportion of population eligible for concessionary fares and controlling for observations identified by Arriva and First as potentially inaccurate ('monopoly variable model')**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.089	0.047	0.061	-0.004	0.182
count of smaller operators (20 or more buses)	-0.005	0.008	0.494	-0.021	0.010
km per passenger	0.105	0.024	0.000	0.058	0.153
proportion of households with no car	0.412	0.113	0.000	0.189	0.636
proportion of ABs	0.047	0.039	0.224	-0.029	0.123
proportion of unemployed	-0.446	0.123	0.000	-0.689	-0.202
operating cost per km	0.076	0.046	0.099	-0.014	0.167
paying passenger boardings	0.002	0.008	0.810	-0.014	0.018
proportion of population under 15	0.831	0.357	0.021	0.127	1.536
proportion of population over 60	0.395	0.177	0.026	0.047	0.743
first dummy	0.509	0.105	0.000	0.302	0.717
arriva dummy	-0.277	0.241	0.252	-0.752	0.198
constant	1.198	0.791	0.132	-0.362	2.759
R <sup>2</sup> = 0.28					

**Table C.14: Including proportion of population eligible for concessionary fares and controlling for observations identified by Arriva and First as potentially inaccurate ('count-of-operators model')**

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
number of national operators	-0.086	0.029	0.003	-0.142	-0.029
count of smaller operators (20 or more buses)	-0.009	0.008	0.239	-0.025	0.006
km per passenger	0.097	0.024	0.000	0.050	0.144
proportion of households with no car	0.407	0.112	0.000	0.187	0.627
proportion of ABs	0.067	0.039	0.088	-0.010	0.144
proportion of unemployed	-0.416	0.122	0.001	-0.658	-0.175
operating cost per km	0.080	0.046	0.079	-0.009	0.170
paying passenger boardings	-0.001	0.008	0.891	-0.017	0.015
proportion of population under 15	0.904	0.351	0.011	0.212	1.596
proportion of population over 60	0.473	0.170	0.006	0.138	0.809
first dummy	0.530	0.102	0.000	0.329	0.731
arriva dummy	-0.232	0.239	0.332	-0.703	0.239
constant	1.277	0.755	0.092	-0.211	2.766
R <sup>2</sup> = 0.26					

**Table C.15: Instrumental variable estimation**

Instrumented: single national operator, count of smaller operators (20 or more buses), paying passengers boardings.

Instruments: population, medium/urban LTA, urban/rural LTA, PTE, mostly urban, North East, North West.

a) First-stage regression for 'single national operator'

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
km per passenger	0.117	0.034	0.001	0.051	0.183
proportion of households with no car	0.104	0.173	0.551	-0.239	0.446
proportion of ABs	0.034	0.241	0.887	-0.441	0.509
proportion of unemployed	0.168	0.220	0.446	-0.266	0.601
operating cost km	-0.057	0.069	0.415	-0.194	0.080
proportion of population under 15	4.393	3.212	0.173	-1.943	10.729
proportion of population over 60	4.575	1.346	0.001	1.919	7.230
first dummy	0.415	0.154	0.008	0.112	0.719
arriva dummy	-0.196	0.357	0.584	-0.899	0.508
population	-0.532	0.402	0.188	-1.326	0.262
medium/urban LTA	0.088	0.096	0.360	-0.101	0.277
urban/rural LTA	0.114	0.074	0.123	-0.031	0.260
PTE	0.098	0.119	0.411	-0.137	0.333
mostly urban	0.005	0.072	0.949	-0.137	0.147
North East	-0.388	0.097	0.000	-0.579	-0.196
North West	-0.001	0.097	0.991	-0.193	0.191
constant	2.248	1.191	0.061	-0.102	4.598
R <sup>2</sup> = 0.42		Partial R <sup>2</sup> = 0.10		F = 3.03 (p=0.0048)	

b) First-stage regression for 'count of smaller operators'

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
km per passenger	-0.102	0.183	0.578	-0.463	0.259
proportion of households with no car	-2.620	0.947	0.006	-4.487	-0.753
proportion of ABs	-3.680	1.313	0.006	-6.270	-1.091
proportion of unemployed	0.987	1.199	0.411	-1.378	3.353
operating cost km	-0.631	0.379	0.097	-1.378	0.116
proportion of population under 15	0.324	17.526	0.985	-34.246	34.894
proportion of population over 60	5.434	7.345	0.460	-9.055	19.922
first dummy	-0.106	0.839	0.900	-1.760	1.549
arriva dummy	1.745	1.945	0.371	-2.092	5.582
population	7.981	2.196	0.000	3.649	12.313
medium/urban LTA	0.146	0.523	0.781	-0.886	1.178
urban/rural LTA	-0.168	0.403	0.677	-0.962	0.627
PTE	-1.551	0.650	0.018	-2.834	-0.268
mostly urban	1.069	0.393	0.007	0.294	1.844
North East	-1.740	0.529	0.001	-2.783	-0.696
North West	0.253	0.530	0.634	-0.793	1.299
constant	-40.997	6.500	0.000	-53.818	-28.176
R <sup>2</sup> = 0.63		Partial R <sup>2</sup> = 0.17		F = 8.18 (p = 0.000)	

c) First-stage regression for 'paying passengers boardings'

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
km per passenger	-0.875	0.206	0.000	-1.282	-0.469
proportion of households with no car	2.118	1.065	0.048	0.017	4.220
proportion of ABs	0.733	1.478	0.621	-2.182	3.648
proportion of unemployed	-0.445	1.350	0.742	-3.108	2.218
operating cost km	-0.014	0.426	0.974	-0.855	0.826
proportion of population under 15	8.001	19.728	0.686	-30.912	46.915
proportion of population over 60	3.507	8.268	0.672	-12.802	19.816
first dummy	-0.271	0.944	0.774	-2.134	1.591
arriva dummy	0.967	2.190	0.659	-3.352	5.286
population	-1.749	2.472	0.480	-6.625	3.127
medium/urban LTA	0.346	0.589	0.558	-0.816	1.507
urban/rural LTA	0.895	0.453	0.050	0.001	1.790
PTE	-0.377	0.732	0.607	-1.821	1.068
mostly urban	0.060	0.442	0.892	-0.812	0.933
North East	0.687	0.596	0.250	-0.488	1.861
North West	0.400	0.597	0.503	-0.777	1.578
constant	7.340	7.317	0.317	-7.092	21.772
R <sup>2</sup> = 0.21		Partial R <sup>2</sup> = 0.036		F = 1.01 (p = 0.427)	

d) 2SLS estimates

explanatory variable	parameter estimate	standard error	p-value	95 per cent confidence intervals	
single national operator	0.316	0.160	0.049	0.001	0.632
paying passenger boardings count of small operator (20 or more buses)	-0.013	0.051	0.798	-0.114	0.088
km per passenger	0.067	0.058	0.249	-0.047	0.181
proportion of households with no car	0.425	0.143	0.003	0.144	0.706
proportion of ABs	0.078	0.063	0.218	-0.047	0.203
proportion of unemployed	-0.449	0.139	0.001	-0.723	-0.176
operating cost km	0.088	0.052	0.094	-0.015	0.191
proportion of population under 15	3.318	2.196	0.132	-1.013	7.648
proportion of population over 60	0.827	1.109	0.457	-1.360	3.015
first dummy	0.413	0.133	0.002	0.150	0.675
arriva dummy	-0.229	0.267	0.394	-0.756	0.299
constant	-1.919	0.838	0.023	-3.572	-0.266
R <sup>2</sup> = 0.18					

## **PART TWO: ADDITIONAL NOTE ON PRICE CONCENTRATION ANALYSIS**

1. The purpose of this note is to set out two issues. First we have identified an inadvertent error in the labelling of our socio-demographic variables in the computer code for some of the price concentration regressions contained in the market study report.<sup>114</sup> As a result we have estimated amended regressions and this note sets out the results of those regressions. As is clear below the qualitative results from these amended regressions are largely unchanged. The second purpose of this note is to set out some further results from examining alternative model specifications that were not contained in the market study report.
2. Subsequent to publication of the report we identified an inadvertent error in the labelling of our socio-demographic variables in the STATA code which was used to estimate the price-concentration regressions presented in Annexe C ('Impact of market structure of bus fares'). As a result of this error, instead of using the proportion of households with no car, the proportion of unemployed and the proportion of ABs as explanatory socio-demographic variables, as stated in our report, we used the actual number of households with no car, unemployed and ABs in each LTA.
3. While a model specification containing absolute numbers rather than proportions may be of interest in its own right, we have re-estimated all the regressions presented in Annexe C using the variable definitions as originally intended (that is, where our socio-demographic variables are defined as the proportion of total population).<sup>115</sup> We consider these definitions to be more appropriate in trying to capture the impact of competition from the car and income on bus fares.

---

<sup>114</sup> Local Bus Services – Report on the market study and proposed decision to make a market investigation reference OFT1112con.

<sup>115</sup> The unemployment was calculated as the proportion of working age population and the ABs as the proportion of all persons aged 16 or over (as per ONS methodology).

4. One of our proxies for income - the proportion of ABs – now has a negative sign in each model which is economically unintuitive as it suggests that income and fares are negatively related. On account of this, we have excluded this variable from the regression analysis.<sup>116</sup> The other proxy for income – the proportion of unemployed – is statistically significant and has the expected sign. We have therefore retained this variable in the regression analysis. We have also retained the proportion of households with no car in the model as this variable is also statistically significant and has the expected sign.
5. As is the case with the original results presented in Annexe C, the parameter estimate associated with the variable indicating that there is only one national provider of commercial services is statistically significant in each model specification. It suggests that a monopoly operator charges higher fares than when there are two or more national operators providing commercial services. The parameter estimates across corresponding equations have similar magnitudes.
6. In particular, the model specification which was used to test the robustness of our results in light of the comments from the operators (as reported in paragraph 78 in Annexe C) suggests that fares in monopoly areas are higher by about eight per cent than in LTAs where there are two or more national competitors. This parameter estimate is statistically significant at the 10 per cent significance level ( $p = 0.059$ ).<sup>117</sup>
7. When we exclude from this model specification the two statistically insignificant variables with parameter estimates close to zero<sup>118</sup> (the number of small operators and the number of paying passengers), the parameter estimate associated with the monopoly variables implies that fares are higher by about nine per cent in monopoly areas, which is the impact we highlighted in Annexe C based on the original

---

<sup>116</sup> Some parameter estimates and their standard errors were affected by this though the overall thrust of the results was not.

<sup>117</sup> See Table C.13A.

<sup>118</sup> Overfitting the model with irrelevant variables causes a loss of efficiency.

estimates. This parameter estimate is statistically significant at the five per cent significance level ( $p = 0.035$ ).<sup>119</sup>

8. For completeness, we have also estimated several additional regression models with the basic model specification as described in paragraph 6 above, but using different functional forms and measures of market structure (that is, using count of competitors instead of a binary indicator for monopoly and semi-log form in addition to logarithmic form). We have also augmented the model described in paragraph 6 using the variables representing different types of LTAs.<sup>120</sup> These results are set out in Tables N.1-N.8.

---

<sup>119</sup> See Table N.1.

<sup>120</sup> All models are estimated using robust standard errors.

9. **Replication of models presented in Annexe C**

10. Table numbers correspond to those set out in Annexe with the addition of an A suffix, that is, Table C.3A corresponds to Table C.3 in the published report.

**Table C.3A: Impact of market structure on bus fares**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.122	0.043	0.005	0.037	0.206
count of smaller operators (20 or more buses)	-0.003	0.006	0.579	-0.015	0.008
km per passenger	0.089	0.034	0.010	0.022	0.156
proportion of households with no car	0.255	0.107	0.018	0.044	0.466
proportion of unemployed	-0.381	0.122	0.002	-0.622	-0.140
operating cost km	0.094	0.042	0.026	0.011	0.176
paying passenger boardings	0.002	0.008	0.787	-0.013	0.017
Constant	-0.932	0.313	0.003	-1.550	-0.314
R <sup>2</sup> = 0.18					

**Table C.4A: Impact of additional competitor on bus fares – linear functional form in the number of competitors**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
number of national operators	-0.059	0.026	0.022	-0.110	-0.008
count of smaller operators (20 or more buses)	-0.002	0.006	0.745	-0.014	0.010
km per passenger	0.092	0.034	0.008	0.025	0.159
proportion of households with no car	0.256	0.109	0.020	0.041	0.470
proportion of unemployed	-0.369	0.122	0.003	-0.610	-0.128
operating cost per pax	0.097	0.042	0.023	0.014	0.181
paying passenger boardings	0.002	0.008	0.787	-0.013	0.017
Constant	-0.735	0.293	0.013	-1.312	-0.158
R <sup>2</sup> = 0.17					

**Table C.5A: Impact of additional competitor on bus fares – separate dummies for each market structure**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
two national operators	-0.110	0.045	0.016	-0.200	-0.021
three national operators	-0.143	0.050	0.005	-0.241	-0.045
four national operators	-0.138	0.144	0.340	-0.421	0.146
count of smaller operators (20 or more buses)	-0.001	0.006	0.803	-0.013	0.010
km per passenger	0.089	0.034	0.010	0.022	0.156
proportion of households with no car	0.266	0.109	0.015	0.052	0.481
proportion of unemployed	-0.385	0.123	0.002	-0.627	-0.143
operating cost per pax	0.097	0.042	0.023	0.014	0.180
paying passenger boardings	0.002	0.008	0.792	-0.013	0.017
Constant	-0.808	0.299	0.007	-1.398	-0.219
R <sup>2</sup> = 0.18					

**Table C.6A: Impact of market structure on bus fares (levels regression)**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.182	0.063	0.004	0.059	0.305
count of smaller operators (20 or more buses)	-0.001	0.007	0.851	-0.015	0.012
km per passenger	0.030	0.026	0.246	-0.021	0.080
proportion of households with no car	1.409	0.520	0.007	0.383	2.435
proportion of unemployed	-15.337	4.658	0.001	-24.522	-6.152
operating cost km	0.041	0.026	0.123	-0.011	0.093
paying passenger boardings	0.000	0.000	0.078	0.000	0.000
Constant	1.170	0.099	0.000	0.974	1.366
R <sup>2</sup> = 0.17					

**Table C.7A: Impact of smaller operators on fares charged by national operators**

c) Small operators with 40 or more buses

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.114	0.043	0.009	0.029	0.198
count of smaller operators (40 or more buses)	-0.019	0.012	0.132	-0.043	0.006
km per passenger	0.090	0.034	0.008	0.023	0.156
proportion of households with no car	0.252	0.108	0.020	0.040	0.464
proportion of unemployed	-0.379	0.122	0.002	-0.618	-0.139
operating cost km	0.091	0.041	0.029	0.009	0.172
paying passenger boardings	0.002	0.008	0.780	-0.013	0.017
constant	-0.915	0.311	0.004	-1.529	-0.300
R <sup>2</sup> = 0.19					

d) Small operators with 100 or more buses

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.127	0.043	0.004	0.041	0.213
count of smaller operators (100 or more buses)	-0.022	0.028	0.436	-0.077	0.033
km per passenger	0.086	0.034	0.012	0.019	0.154
proportion of households with no car	0.260	0.108	0.017	0.047	0.472
proportion of unemployed	-0.381	0.123	0.002	-0.623	-0.140
operating cost km	0.091	0.042	0.031	0.009	0.174
paying passenger boardings	0.001	0.008	0.884	-0.014	0.016
constant	-0.919	0.312	0.004	-1.534	-0.303
R <sup>2</sup> = 0.18					

**Table C.8A: Fares and service quality**

f) Proportion of buses with disability access certificate

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.103	0.042	0.015	0.021	0.186
count of smaller operators (20 or more buses)	-0.004	0.006	0.558	-0.015	0.008
km per passenger	0.099	0.034	0.004	0.033	0.165
proportion of households with no car	0.229	0.109	0.036	0.015	0.444
proportion of unemployed	-0.336	0.125	0.008	-0.584	-0.089
operating cost km	0.120	0.044	0.007	0.034	0.206
paying passenger boardings	0.003	0.008	0.721	-0.013	0.018
proportion of buses with disability access certificate	-0.323	0.122	0.009	-0.563	-0.083
constant	-0.751	0.327	0.023	-1.395	-0.106
R <sup>2</sup> = 0.20					

g) Proportion of buses with low floor access

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.133	0.042	0.002	0.049	0.217
count of smaller operators (20 or more buses)	-0.003	0.006	0.663	-0.014	0.009
km per passenger	0.082	0.035	0.019	0.014	0.151
proportion of households with no car	0.325	0.106	0.003	0.116	0.535
proportion of unemployed	-0.452	0.124	0.000	-0.696	-0.208
operating cost per km	0.072	0.041	0.078	-0.008	0.152
paying passenger boardings	0.000	0.008	0.989	-0.015	0.015
proportion of buses with low floors access	-0.364	0.121	0.003	-0.602	-0.125
constant	-0.975	0.311	0.002	-1.587	-0.362
R <sup>2</sup> = 0.21					

#### h) Proportion of buses with GPS

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.113	0.041	0.006	0.033	0.193
count of smaller operators (20 or more buses)	-0.001	0.006	0.814	-0.012	0.010
km per passenger	0.096	0.031	0.002	0.035	0.156
proportion of households with no car	0.240	0.101	0.019	0.040	0.439
proportion of unemployed	-0.389	0.118	0.001	-0.622	-0.157
operating cost per km	0.150	0.045	0.001	0.062	0.237
paying passenger boardings	0.006	0.008	0.438	-0.009	0.021
proportion of buses with GPS	-0.266	0.080	0.001	-0.424	-0.109
constant	-0.964	0.310	0.002	-1.576	-0.353
R <sup>2</sup> = 0.23					

#### i) Proportion of buses with CCTV

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.105	0.041	0.011	0.024	0.185
count of smaller operators (20 or more buses)	-0.002	0.006	0.685	-0.014	0.009
km per passenger	0.086	0.034	0.011	0.020	0.152
proportion of households with no car	0.237	0.106	0.027	0.027	0.447
proportion of unemployed	-0.330	0.120	0.007	-0.567	-0.093
operating cost per km	0.095	0.042	0.024	0.013	0.177
paying passenger boardings	0.002	0.008	0.768	-0.013	0.018
proportion of buses with CCTV	-0.312	0.092	0.001	-0.493	-0.132
constant	-0.691	0.311	0.027	-1.304	-0.078
R <sup>2</sup> = 0.22					

j) Average fleet age

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.123	0.043	0.005	0.038	0.207
count of smaller operators (20 or more buses)	-0.003	0.006	0.576	-0.015	0.008
km per passenger	0.089	0.034	0.010	0.021	0.156
proportion of households with no car	0.253	0.108	0.020	0.039	0.466
proportion of unemployed	-0.377	0.126	0.003	-0.625	-0.129
operating cost per km	0.093	0.042	0.027	0.011	0.175
paying passenger boardings	0.002	0.008	0.788	-0.013	0.017
average fleet age	-0.017	0.077	0.822	-0.170	0.135
constant	-0.882	0.400	0.029	-1.672	-0.093
R <sup>2</sup> = 0.18					

**Table C.9A: Impact of journey distance**

e) Population density

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.128	0.043	0.003	0.044	0.213
count of smaller operators (20 or more buses)	-0.003	0.006	0.607	-0.015	0.009
km per passenger	0.087	0.036	0.015	0.017	0.157
proportion of households with no car	0.206	0.118	0.082	-0.026	0.437
proportion of unemployed	-0.360	0.127	0.005	-0.610	-0.110
operating cost km	0.076	0.043	0.082	-0.010	0.162
paying passenger boardings	0.000	0.008	0.952	-0.015	0.016
population density	0.022	0.014	0.121	-0.006	0.050
Constant	-0.935	0.315	0.003	-1.556	-0.314
R <sup>2</sup> = 0.19					

f) Area size

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.117	0.043	0.007	0.032	0.202
count of smaller operators (20 or more buses)	0.000	0.007	0.940	-0.013	0.012
km per passenger	0.090	0.035	0.010	0.022	0.158
proportion of households with no car	0.236	0.111	0.034	0.017	0.455
proportion of unemployed	-0.381	0.123	0.002	-0.624	-0.138
operating cost km	0.083	0.044	0.059	-0.003	0.169
paying passenger boardings	0.002	0.008	0.812	-0.013	0.017
area size	-0.015	0.013	0.277	-0.041	0.012
constant	-0.793	0.346	0.023	-1.475	-0.112
R <sup>2</sup> = 0.18					

g) Type of LTA

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.156	0.047	0.001	0.063	0.249
count of smaller operators (20 or more buses)	-0.003	0.006	0.595	-0.016	0.009
km per passenger	0.085	0.033	0.011	0.020	0.151
proportion of households with no car	-0.001	0.121	0.993	-0.240	0.238
proportion of unemployed	-0.253	0.117	0.031	-0.483	-0.023
operating cost km	0.080	0.043	0.065	-0.005	0.165
paying passenger boardings	-0.002	0.008	0.757	-0.017	0.013
large urban	-0.172	0.086	0.047	-0.342	-0.002
medium urban	-0.262	0.084	0.002	-0.428	-0.097
mixed urban/rural	-0.168	0.083	0.044	-0.331	-0.004
mostly rural	-0.331	0.088	0.000	-0.504	-0.158
mostly urban	-0.210	0.074	0.005	-0.357	-0.063
other large conurbations	-0.088	0.077	0.257	-0.240	0.064
constant	-0.605	0.304	0.048	-1.204	-0.006
R <sup>2</sup> = 0.25					

## h) UK regions

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.127	0.050	0.012	0.028	0.226
count of smaller operators (20 or more buses)	0.000	0.006	0.943	-0.011	0.012
km per passenger	0.083	0.034	0.015	0.016	0.150
proportion of households with no car	0.189	0.104	0.071	-0.016	0.393
proportion of unemployed	-0.231	0.139	0.098	-0.505	0.043
operating cost km	0.065	0.045	0.152	-0.024	0.153
paying passenger boardings	0.001	0.008	0.926	-0.015	0.016
East of England	0.185	0.066	0.006	0.054	0.316
North East	0.045	0.093	0.626	-0.138	0.229
North West	0.111	0.085	0.192	-0.056	0.279
Scotland	0.166	0.077	0.033	0.013	0.318
South East	0.208	0.076	0.007	0.058	0.358
South West	0.223	0.076	0.004	0.074	0.372
Wales	0.078	0.071	0.272	-0.062	0.217
Yorkshire and the Humber	0.139	0.070	0.047	0.002	0.277
constant	-0.631	0.384	0.102	-1.389	0.127
R <sup>2</sup> = 0.22					

**Table C.10A: Using total km per passenger instead of commercial km per passenger as proxy for service levels**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.149	0.047	0.002	0.056	0.243
count of smaller operators (20 or more buses)	-0.002	0.006	0.711	-0.015	0.010
total km per passenger	0.015	0.031	0.619	-0.046	0.076
proportion of households with no car	0.315	0.120	0.009	0.078	0.553
proportion of unemployed	-0.457	0.138	0.001	-0.729	-0.185
operating cost per km	0.046	0.044	0.297	-0.041	0.134
paying passenger boardings	-0.005	0.008	0.549	-0.021	0.011
constant	-0.998	0.335	0.003	-1.659	-0.338
R <sup>2</sup> = 0.13					

**Table C.11A: Controlling for observations identified by First and Arriva as potentially inaccurate**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.100	0.041	0.015	0.019	0.181
count of smaller operators (20 or more buses)	0.000	0.006	0.946	-0.012	0.011
km per passenger	0.106	0.035	0.003	0.037	0.175
proportion of households with no car	0.298	0.108	0.006	0.086	0.510
proportion of unemployed	-0.371	0.123	0.003	-0.613	-0.130
operating cost per km	0.068	0.041	0.096	-0.012	0.148
paying passenger boardings	0.004	0.008	0.607	-0.011	0.019
first dummy	0.434	0.089	0.000	0.257	0.610
arriva dummy	-0.272	0.036	0.000	-0.342	-0.202
Constant	-0.874	0.312	0.006	-1.490	-0.258
R <sup>2</sup> = 0.26					

**Table C.12A: Including proportion of population eligible for concessionary fares**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.115	0.044	0.010	0.028	0.203
count of smaller operators (20 or more buses)	-0.005	0.006	0.396	-0.017	0.007
km per passenger	0.087	0.034	0.010	0.021	0.154
proportion of households with no car	0.346	0.130	0.009	0.089	0.602
proportion of unemployed	-0.475	0.144	0.001	-0.759	-0.191
operating cost per km	0.101	0.043	0.020	0.016	0.187
paying passengers boardings	0.002	0.008	0.791	-0.013	0.017
proportion of population under 15	0.620	0.357	0.084	-0.085	1.325
proportion of population over 60	0.201	0.180	0.264	-0.153	0.556
Constant	0.229	0.709	0.747	-1.168	1.626
R <sup>2</sup> = 0.19					

**Table C.13A: Including proportion of population eligible for concessionary fares and controlling for observations identified by Arriva and First as potentially inaccurate ('monopoly variable model')**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.076	0.040	0.059	-0.003	0.156
count of smaller operators (20 or more buses)	-0.004	0.006	0.494	-0.016	0.008
km per passenger	0.107	0.034	0.002	0.040	0.174
proportion of households with no car	0.429	0.128	0.001	0.178	0.681
proportion of unemployed	-0.501	0.141	0.000	-0.779	-0.222
operating cost per km	0.073	0.041	0.077	-0.008	0.153
paying passenger boardings	0.004	0.007	0.620	-0.011	0.018
proportion of population under 15	0.809	0.354	0.023	0.111	1.507
proportion of population over 60	0.465	0.167	0.006	0.135	0.795
first dummy	0.501	0.082	0.000	0.340	0.662
arriva dummy	-0.293	0.038	0.000	-0.367	-0.218
Constant	0.956	0.704	0.176	-0.432	2.344
$R^2 = 0.28$					

**Table C.14A: Including proportion of population eligible for concessionary fares and controlling for observations identified by Arriva and First as potentially inaccurate ('count-of-operators model')**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
number of national operators	-0.052	0.023	0.026	-0.097	-0.006
count of smaller operators (20 or more buses)	-0.002	0.006	0.739	-0.014	0.010
km per passenger	0.106	0.034	0.002	0.039	0.173
proportion of households with no car	0.454	0.129	0.001	0.200	0.707
proportion of unemployed	-0.515	0.141	0.000	-0.792	-0.237
operating cost per km	0.077	0.041	0.064	-0.004	0.158
paying passenger boardings	0.003	0.007	0.640	-0.011	0.018
proportion of population under 15	0.867	0.349	0.014	0.178	1.556
proportion of population over 60	0.524	0.165	0.002	0.198	0.850
first dummy	0.524	0.088	0.000	0.350	0.699
arriva dummy	-0.276	0.040	0.	-0.356	-0.196
Constant	1.255	0.677	0.065	-0.080	2.590
R <sup>2</sup> = 0.29					

### Table C.15A: Instrumental variable estimation

Instrumented: single national operator, count of smaller operators (20 or more buses), paying passengers boardings.

Instruments: population, medium/urban LTA, urban/rural LTA, PTE, mostly urban, North East, North West.

c) First-stage regression for 'single national operator'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	0.111	0.034	0.001	0.044	0.178
proportion of households with no car	0.163	0.190	0.394	-0.213	0.538
proportion of unemployed	0.101	0.210	0.633	-0.314	0.516
operating cost km	-0.055	0.069	0.429	-0.190	0.081
proportion of population under 15	0.852	0.566	0.134	-0.265	1.968
proportion of population over 60	1.025	0.273	0.000	0.487	1.563
first dummy	0.447	0.152	0.004	0.147	0.746
arriva dummy	-0.198	0.355	0.578	-0.899	0.503
population	-0.225	0.038	0.000	-0.300	-0.151
medium/urban LTA	0.110	0.094	0.242	-0.075	0.295
urban/rural LTA	0.114	0.075	0.129	-0.034	0.261
PTE	0.086	0.119	0.468	-0.148	0.321
mostly urban	0.005	0.072	0.939	-0.136	0.147
North East	-0.390	0.093	0.000	-0.574	-0.206
North West	-0.005	0.096	0.958	-0.195	0.184
constant	6.693	1.197	0.000	4.331	9.055
$R^2 = 0.40$					

d) First-stage regression for 'count of smaller operators'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.156	0.191	0.414	-0.533	0.220
proportion of households with no car	-1.951	1.070	0.070	-4.062	0.161
proportion of unemployed	2.451	1.183	0.040	0.117	4.784
operating cost km	-0.661	0.387	0.089	-1.426	0.103
proportion of population under 15	3.758	3.184	0.239	-2.522	10.038
proportion of population over 60	2.692	1.535	0.081	-0.335	5.719
first dummy	-0.391	0.854	0.647	-2.076	1.293
arriva dummy	1.791	1.998	0.371	-2.151	5.732
population	2.605	0.214	0.000	2.183	3.026
medium/urban LTA	-0.100	0.527	0.850	-1.139	0.940
urban/rural LTA	-0.157	0.420	0.709	-0.986	0.671
PTE	-1.655	0.669	0.014	-2.975	-0.335
mostly urban	0.935	0.404	0.022	0.138	1.732
North East	-2.089	0.523	0.000	-3.121	-1.056
North West	0.066	0.540	0.902	-0.999	1.132
constant	-13.654	6.733	0.044	-26.936	-0.372
R <sup>2</sup> = 0.60					

c) First-stage regression for 'paying passengers boardings'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.911	0.209	0.000	-1.324	-0.498
proportion of households with no car	1.574	1.174	0.182	-0.742	3.889
proportion of unemployed	-0.097	1.298	0.940	-2.657	2.462
operating cost km	0.008	0.425	0.986	-0.831	0.846
proportion of population under 15	-1.430	3.492	0.683	-8.319	5.458
proportion of population over 60	-0.231	1.683	0.891	-3.552	3.089
first dummy	-0.288	0.937	0.759	-2.136	1.559
arriva dummy	1.086	2.191	0.621	-3.237	5.409
population	0.736	0.234	0.002	0.274	1.198
medium/urban LTA	0.465	0.578	0.422	-0.675	1.605
urban/rural LTA	0.998	0.461	0.032	0.089	1.907
PTE	-0.384	0.734	0.601	-1.833	1.064
mostly urban	0.180	0.443	0.685	-0.694	1.054
North East	0.712	0.574	0.216	-0.420	1.845
North West	0.463	0.593	0.436	-0.706	1.632
constant	3.137	7.385	0.671	-11.431	17.705
		R <sup>2</sup> = 0.21			

d) 2SLS estimates

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.300	0.137	0.030	0.029	0.570
paying passenger boardings	0.055	0.045	0.220	-0.033	0.143
count of small operator (20 or more buses)	0.015	0.014	0.281	-0.012	0.041
km per passenger	0.114	0.045	0.012	0.026	0.203
proportion of households with no car	0.182	0.172	0.291	-0.157	0.520
proportion of unemployed	-0.311	0.158	0.050	-0.622	0.001
operating cost km	0.103	0.056	0.065	-0.006	0.213
proportion of population under 15	0.204	0.451	0.652	-0.686	1.094
proportion of population over 60	0.045	0.265	0.867	-0.478	0.567
first dummy	0.430	0.128	0.001	0.176	0.683
arriva dummy	-0.288	0.281	0.306	-0.841	0.265
constant	-1.222	1.311	0.352	-3.808	1.363

**Additional results (excluding number of small operators and paying passengers)**

**Table N.1: Impact of monopoly market structure on bus fares – logarithmic form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.085	0.040	0.035	0.006	0.164
Km per passenger	0.103	0.032	0.001	0.040	0.167
Proportion of households with no car	0.432	0.126	0.001	0.184	0.680
Proportion of unemployed	-0.493	0.139	0.000	-0.767	-0.220
operating cost per km	0.075	0.040	0.066	-0.005	0.155
Proportion of population under 15	0.769	0.342	0.026	0.094	1.444
Proportion of population over 60	0.439	0.162	0.007	0.120	0.759
first dummy	0.501	0.081	0.000	0.342	0.661
arriva dummy	-0.296	0.036	0.000	-0.367	-0.226
Constant	0.915	0.678	0.179	-0.422	2.251
R <sup>2</sup> = 0.28					

**Table N.2: Impact of monopoly market structure on bus fares – semi-log form<sup>121</sup>**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.104	0.041	0.013	0.022	0.185
km per passenger	0.023	0.016	0.151	-0.008	0.054
proportion of households with no car	1.581	0.466	0.001	0.662	2.501
Proportion of unemployed	-15.037	3.965	0.000	-22.856	-7.217
operating cost per km	0.024	0.019	0.212	-0.014	0.062
Proportion of population under 15	4.647	2.030	0.023	0.644	8.650
Proportion of population over 60	2.388	0.873	0.007	0.667	4.109
first dummy	0.445	0.077	0.000	0.294	0.597
arriva dummy	-0.227	0.030	0.000	-0.286	-0.168
Constant	-1.294	0.532	0.016	-2.344	-0.244
R <sup>2</sup> = 0.24					

<sup>121</sup> Only dependent variable transformed logarithmically.

**Table N.3: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – logarithmic form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.101	0.043	0.019	0.017	0.186
km per passenger	0.104	0.031	0.001	0.043	0.164
proportion of households with no car	0.102	0.133	0.441	-0.159	0.364
proportion of unemployed	-0.328	0.130	0.012	-0.585	-0.072
operating cost km	0.052	0.039	0.185	-0.025	0.128
proportion of population under 15	0.607	0.362	0.095	-0.107	1.322
proportion of population over 60	0.656	0.211	0.002	0.240	1.073
first dummy	0.467	0.092	0.000	0.285	0.648
arriva dummy	-0.179	0.050	0.000	-0.277	-0.080
large urban	-0.120	0.085	0.158	-0.286	0.047
medium urban	-0.174	0.081	0.033	-0.333	-0.014
mixed urban/rural	-0.135	0.084	0.109	-0.301	0.030
mostly rural	-0.327	0.097	0.001	-0.518	-0.136
mostly urban	-0.198	0.077	0.011	-0.351	-0.046
other large conurbations	-0.007	0.084	0.933	-0.173	0.159
constant	1.245	0.751	0.099	-0.237	2.727
R <sup>2</sup> = 0.34					

**Table N.4: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – semi-log form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.116	0.045	0.010	0.028	0.204
km per passenger	0.024	0.015	0.110	-0.005	0.053
proportion of households with no car	0.365	0.465	0.434	-0.553	1.283
proportion of unemployed	-10.428	3.642	0.005	-17.611	-3.245
operating cost km	0.015	0.018	0.400	-0.020	0.051
proportion of population under 15	3.508	2.015	0.083	-0.466	7.482
proportion of population over 60	3.241	1.046	0.002	1.178	5.304
first dummy	0.423	0.084	0.000	0.258	0.588
arriva dummy	-0.116	0.048	0.017	-0.211	-0.021
large urban	-0.131	0.087	0.133	-0.302	0.040
medium urban	-0.199	0.083	0.017	-0.363	-0.036
mixed urban/rural	-0.154	0.082	0.063	-0.316	0.009
mostly rural	-0.346	0.098	0.000	-0.538	-0.153
Mostly urban	-0.221	0.080	0.007	-0.379	-0.062
other large conurbations	-0.017	0.089	0.849	-0.192	0.158
constant	-0.920	0.553	0.098	-2.010	0.170
R <sup>2</sup> = 0.30					

**Table N.5: Impact of additional competitor on bus fares – logarithmic form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
number of national operators	-0.054	0.022	0.016	-0.099	-0.010
km per passenger	0.103	0.032	0.001	0.040	0.166
Proportion of households with no car	0.461	0.127	0.000	0.211	0.711
Proportion of unemployed	-0.515	0.138	0.000	-0.787	-0.243
operating cost per km	0.077	0.041	0.057	-0.002	0.157
Proportion of population under 15	0.854	0.342	0.013	0.180	1.527
Proportion of population over 60	0.517	0.162	0.002	0.198	0.836
first dummy	0.524	0.088	0.000	0.351	0.697
arriva dummy	-0.275	0.040	0.000	-0.354	-0.197
Constant	1.278	0.665	0.056	-0.033	2.588
R <sup>2</sup> = 0.29					

**Table N.6: Impact of additional competitor on bus fares – semi-log form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
number of national operators	-0.062	0.023	0.007	-0.108	-0.017
km per passenger	0.023	0.016	0.149	-0.008	0.053
Proportion of households with no car	1.724	0.473	0.000	0.791	2.657
Proportion of unemployed	-15.855	3.967	0.000	-23.678	-8.032
operating cost per km	0.025	0.020	0.217	-0.015	0.064
Proportion of population under 15	5.186	2.028	0.011	1.187	9.186
Proportion of population over 60	2.827	0.882	0.002	1.088	4.566
first dummy	0.472	0.084	0.000	0.305	0.638
arriva dummy	-0.203	0.035	0.000	-0.271	-0.135
Constant	-1.341	0.527	0.012	-2.380	-0.302
R <sup>2</sup> = 0.25					

**Table N.7: Impact of additional competitor on bus fares (separate dummies for each market structure) – logarithmic form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
two national operators	-0.043	0.044	0.321	-0.129	0.043
three national operators	-0.124	0.046	0.008	-0.216	-0.032
four national operators	-0.100	0.124	0.424	-0.345	0.146
km per passenger	0.104	0.031	0.001	0.043	0.166
Proportion of households with no car	0.494	0.127	0.000	0.244	0.744
Proportion of unemployed	-0.538	0.136	0.000	-0.806	-0.269
operating cost per pax	0.079	0.041	0.052	-0.001	0.159
Proportion of population under 15	0.949	0.359	0.009	0.240	1.657
Proportion of population over 60	0.558	0.172	0.001	0.219	0.898
first dummy	0.527	0.085	0.000	0.359	0.695
arriva dummy	-0.259	0.043	0.000	-0.343	-0.175
Constant	1.412	0.735	0.056	-0.037	2.861
R <sup>2</sup> = 0.29					

**Table N.8: Impact of additional competitor on bus fares (separate dummies for each market structure) – semi-log form**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
two national operators	-0.062	0.045	0.166	-0.150	0.026
three national operators	-0.147	0.050	0.003	-0.244	-0.049
four national operators	-0.115	0.118	0.333	-0.347	0.118
km per passenger	0.022	0.016	0.151	-0.008	0.053
Proportion of households with no car	1.874	0.486	0.000	0.916	2.832
Proportion of unemployed	-16.813	3.964	0.000	-24.631	-8.995
operating cost per pax	0.025	0.020	0.211	-0.014	0.064
Proportion of population under 15	5.771	2.183	0.009	1.467	10.076
Proportion of population over 60	3.018	0.950	0.002	1.145	4.891
first dummy	0.467	0.082	0.000	0.305	0.629
arriva dummy	-0.184	0.040	0.000	-0.263	-0.105
Constant	-1.557	0.591	0.009	-2.724	-0.391
R <sup>2</sup> = 0.25					

## **PART THREE: RESPONSE TO REPORT BY FRONTIER ECONOMICS**

### **Introduction**

- E.7 This section reviews the response to our empirical work in relation to commercial services. We presented the findings from two econometric studies in the Proposed Decision. One study looked into the relationship between bus fares and market concentration, and the other, into the relationship between competition and quality of bus fleets. We found that that high concentration is associated with higher prices and lower fleet quality.
- E.8 Four national bus operators, namely First, Stagecoach, Arriva and Go-Ahead commissioned Frontier Economics to review our econometric work. In what follows we describe, and respond to, the comments made by Frontier about our analysis of the relationship between price and concentration.<sup>122</sup> We also respond to the comments made by the operators on the basis of the work carried out by Frontier.

### **Data issues**

- E.9 Frontier has made two overarching arguments in relation to the data which underpin our findings. First, that our results are sensitive to a small number of data points, and second, that the data is subject to substantial measurement errors which create noise. We address each of those arguments in turn.

---

<sup>122</sup> Frontier made no comments about our analysis of the relationship between market concentration and fleet quality.

## Sensitivity of results to a small number of data points

- E.10 Frontier argues that our results are driven by a small number of observations. Specifically, it suggests that the removal of 15 observations from our data set results in the 'monopoly effect' no longer being statistically significant in any of our regression specifications. From this, Frontier infers that our results are not robust and that, therefore, caution must be taken in placing weight on the findings of the regression analysis.
- E.11 We note that the hypothesis being tested in our study is whether prices are higher in 'monopoly' areas. There are 55 'monopoly' observations in our data set. The 15 observations identified by Frontier are virtually all in 'monopoly' areas (14 out of 15, or a quarter of all the 'monopoly' observations), and the fares estimates for those areas are all at the higher end of the fares range (11 of the highest fares estimates for 'monopoly' areas). In this context, we would not characterise 15 observations as a 'small number of observations'.
- E.12 Furthermore, in using 55 monopoly observations for the analysis we had already pared down the sample substantially as part of the process of data checking which also involved asking operators to validate the data. In doing so, as explained at paragraphs C29 and C33 in the Proposed Decision, we acted cautiously and removed observations where we suspected that the data may not be accurate.
- E.13 Taking the dataset at the face value (that is, as reported to the DfT by the operators<sup>123</sup>), and estimating the price concentration regression using all 222 available observations results in a much higher estimate of the impact of market concentration on price than the one reported in the Proposed Decision. Specifically, it suggests that the price is higher by about 20% when there is only one national provider of commercial services in an LTA than in the case when

---

<sup>123</sup> DfT makes a small adjustment to data to account for passenger boardings that may be underreported by the operators.

there are two or more such providers. This is shown in the table below.

**Table E.1: Regression estimates based on the complete data set**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.202	0.060	0.001	0.083	0.320
count of smaller operators (20 or more buses)	0.007	0.007	0.310	-0.007	0.021
km per passenger	0.101	0.042	0.016	0.019	0.184
proportion of households with no car	0.158	0.156	0.315	-0.151	0.466
proportion of unemployed	-0.320	0.178	0.073	-0.670	0.030
operating cost per km	0.174	0.054	0.001	0.068	0.281
paying passenger boardings	0.012	0.011	0.253	-0.009	0.034
Constant	-1.110	0.473	0.020	-2.042	0.178
$R^2 = 0.15$					

E.14 Given the concerns that were expressed to us about DfT’s STATS100 data set by some of the operators, we have carefully examined our data set in order to determine the extent to which our results may be sensitive to potential data errors and/or different model specifications. This was also explained in detail in Annexe C of the Proposed Decision.

E.15 We excluded from our data set all observations where our fare estimates were less than 50p or more than £3.50. As explained at paragraph C.29 in the Proposed Decision, we used the data which was supplied to us by Stagecoach to identify these observations as outliers. This is because the reporting units for Stagecoach are more similar to the way that DfT’s data are reported than the data which was provided to us by other operators.

E.16 This reduced the size of our sample from 222 observations to 208 observations. Using this reduced data set we then estimated various

model specifications (that is, different functional forms, control variables and measures of concentration). All of the results were statistically significant and suggested that, depending on the exact model specification, prices in 'monopoly' areas are higher by between 12 per cent and 14 per cent.

- E.17 This was a cautious approach. Another possibility, arguably less intrusive on the data, was to cap the outlier observations at the selected cut-off points (50p and £3.50) and then to re-estimate our models. This method yields parameter estimates associated with the 'monopoly' variable that are higher than the estimates which we obtained from the reduced data set.
- E.18 An alternative option was to preserve the data set as reported to the DfT by the operators, and use regression techniques that are robust to the presence of outliers. Using this approach results in parameter estimates of similar magnitude to the ones we obtained from the reduced data set. This is shown in the table below which presents the results obtained using a Quantile Regression estimation technique which is less affected by outliers than an Ordinary Least Squares (OLS) estimation technique.

**Table E.2: Quantile regression estimation results**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.130	0.047	0.006	0.037	0.224
count of smaller operators (20 or more buses)	0.000	0.007	0.945	-0.013	0.014
km per passenger	0.128	0.027	0.000	0.075	0.181
proportion of households with no car	0.155	0.121	0.203	-0.084	0.394
proportion of unemployed	-0.309	0.131	0.019	-0.567	-0.052
operating cost km	0.095	0.052	0.072	-0.009	0.198
paying passenger boardings	0.013	0.009	0.134	-0.004	0.031
Constant	-1.004	0.327	0.002	-1.648	-0.359
Pseudo R <sup>2</sup> = 0.07					

- E.19 The regression results suggest that the price is higher by about 13 per cent when there is only one national provider of commercial

services in an LTA than in the case when there are two or more such providers. This is similar to the results that we obtained using the reduced data set (see Table C.3A in the Additional note on price concentration).

- E.20 Prior to publication of the report we had also asked the operators to validate the data which they supplied to the DfT. We sent them the data which DfT had disclosed to us and wrote to them describing the methodology and the data which we had used to calculate average fares in our sample. We invited the operators to comment on our methodology and to report to us any concerns that they may have about data accuracy and/or our methodology. The operators' responses, and the way we dealt with their concerns, were described in detail at paragraphs C.30 – C.45 in the Proposed Decision.
- E.21 We addressed operators' concerns about our estimates of concessionary passenger boardings by including in the regression analysis the proportion of the population eligible for concessionary fares. We have also dealt with the specific points made by two operators in relation to the accuracy of their own data.
- E.22 Arriva told us that there were three data points in DfT's data set purporting to show Arriva's traffic and revenue in 2006/07 which Arriva had not supplied to DfT.<sup>124</sup> Arriva suggested that we remove these data points from our data set. First identified several examples where the dataset which was given to us by DfT may be inaccurate.
- E.23 We have effectively removed the impact of those observations (8 in total) from our regression analysis by assigning them dummy variables. We found that our main finding, namely that the fares are higher in 'monopoly' areas, was unaffected by changes to the data set although the estimated price difference was smaller than that which was estimated using the larger data set. We reported both sets of estimates.

---

<sup>124</sup> These 3 observations relating to Arriva were not in the data which we used in our regression analysis but we identified 1 LTA in our data set in which Arriva operates and which may have been affected by this.

E.24 To sum up, our findings proved to be robust to two rounds of observation removal - one of which was implemented on the basis of a data validation exercise with the operators - and a whole range of model specifications (29 models in total). Therefore, we do not accept Frontier's characterisation of our findings as being sensitive to a small number of observations. Rather, it would appear to us that it is a selective nature of Frontier's exercise – discounting a quarter of our 'monopoly' observations with some of the highest fares – which gives this impression.

### **General points about measurement error**

E.25 Frontier notes that the explanatory power of our regressions (the  $R^2$  term) is relatively low. It ranges from 0.13 to 0.30 depending on model specifications.<sup>125</sup> According to Frontier, this suggests that our analysis is missing variables that are important drivers of price difference or that it is based on data that is subject to substantial measurement errors which create noise in the data.

E.26 In our work we have considered carefully both possibilities. Our model specification follows the best practice in the conduct of price concentration analysis. We have included all theoretically relevant explanatory variables (for example, quality and cost differences) as well as the variables that were suggested to us by the operators (for example, traffic volumes and income). We note that Frontier has not indicated which specific variables may be missing from our analysis, or why our models may be misspecified.

E.27 We were aware of data limitations and this is why we conducted extensive robustness checks. These checks were described in detail in the section entitled 'Measurement error' in the Proposed Decision (paragraphs C28 – C45). We also sent the dataset to the bus companies for validation before we used it as explained at paragraph E.20 above. The bus companies found very few errors with the

---

<sup>125</sup> In fact, the highest  $R^2$  reported is 0.34 (see Table N.3 in the Additional note on price concentration).

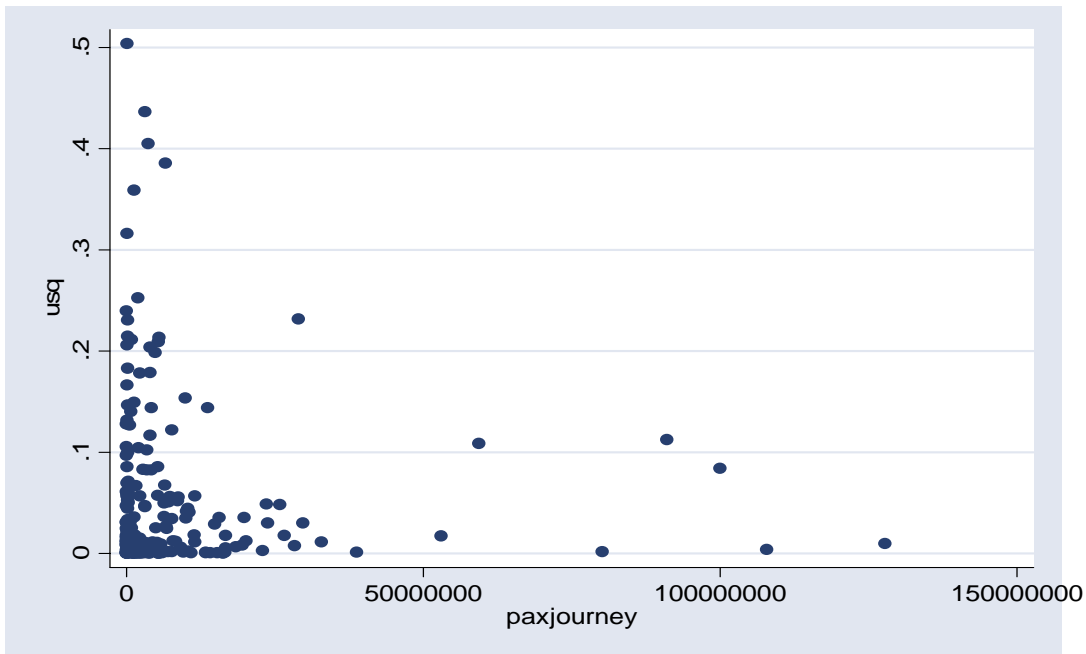
dataset and where errors were found we effectively excluded those data points.

- E.28 It is possible that despite all the checks there may still be some errors in that dataset. This is mainly to do with the fact that operators do not record their revenue and traffic data at LTA level, which is how DfT's dataset is structured, and that each subsidiary may use a different method to provide its data to DfT in this format. The operators told us that despite their best endeavors there is inevitably an element of arbitrariness involved in this process which means that we should treat their data with caution when considering it at LTA level.
- E.29 While we accept this, we have no reasons to suspect that these errors are anything but random. Random mistakes in a dataset just mean that our results are likely to be a diluted version of reality. Random mistakes are like static or white noise affecting a TV screen - it doesn't produce the opposite picture, just a less clear version of the real picture. We believe that we observe our estimates of higher prices in monopoly areas despite the noise, not because of it.

### Measurement error for small LTAs

- E.30 According to Frontier, a number of the operators expressed particular concerns over the accuracy of data for very small LTAs, given that the data used in the analysis has been allocated to LTAs, rather than recorded specifically for LTAs. However, no other reasons have been given by either Frontier or the operators as to why this may be the case, and none of the operators raised this particular issue when we spoke to them as part of the data validation exercise.
- E.31 In order to test for this effect, Frontier produced a scatter plot of squared residuals from one of our regressions against the number of passenger journeys reported by a particular operator in a particular LTA. This figure is reproduced below.

**Figure E.1: Scatter plot of squared residuals against passenger journeys**



- E.32 According to Frontier, the figure shows that the squared residuals are substantially larger for small LTAs than for large LTAs. Frontier argues that this proves that there are data errors for small LTAs.
- E.33 In other to correct for this problem, Frontier used a Weighted Least Squares (WLS) estimation technique. Specifically, Frontier multiplied the equation through by the square root of the number of passenger journeys and then re-estimated the model using the OLS method. Frontier found that the 'monopoly' variable was no longer statistically significant.
- E.34 We have a number of issues with Frontier's analysis. The first point to note is that the pattern shown in Figure E.1 above indicates that the variance of the error term may be greater for smaller LTAs than for large LTAs, which reflects the fact that there is a greater variability of price for smaller LTAs than for large LTAs. However, as we show below, this is likely to be a genuine feature of the data due

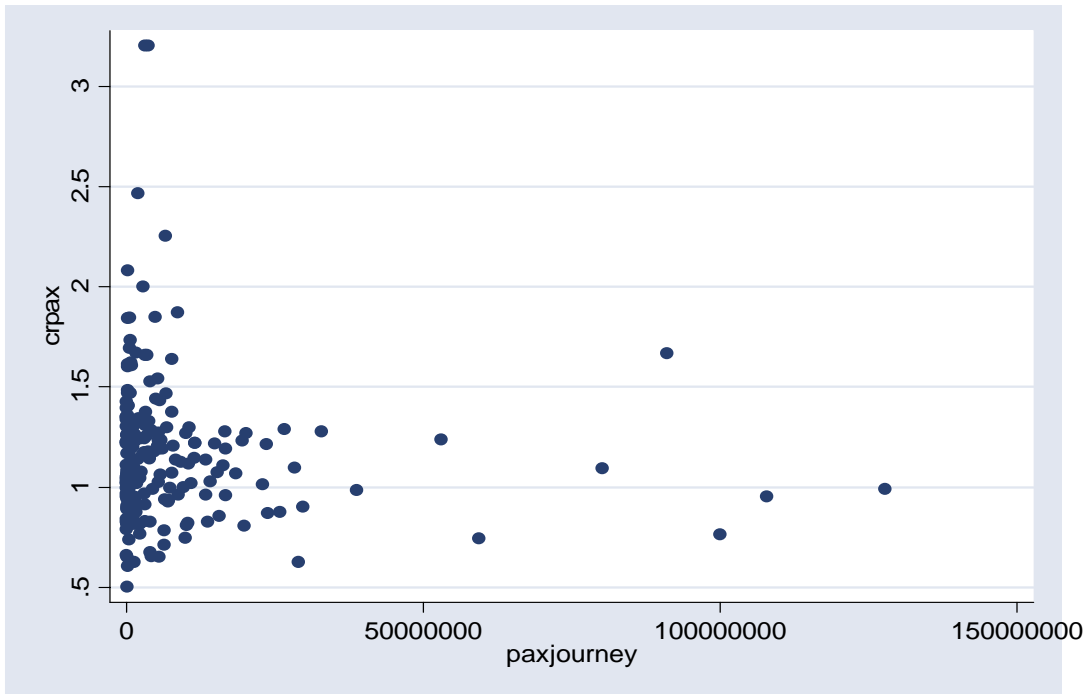
to a greater heterogeneity of smaller LTAs and not necessarily due to any measurement error for small LTAs as suggested by Frontier.

E.35 Specifically, a number of variables that affect bus pricing show much greater variation at the level of small LTAs than at the level of large LTAs. These include, but are not limited to, the number of kilometres per passenger, the proportion of retired population and the proportion of households with no car. Therefore, we would expect there to be a greater variation in prices for smaller LTAs, and hence the greater variation in the variance of the regression residual for these LTAs which is the pattern shown in Figure E.1 above.

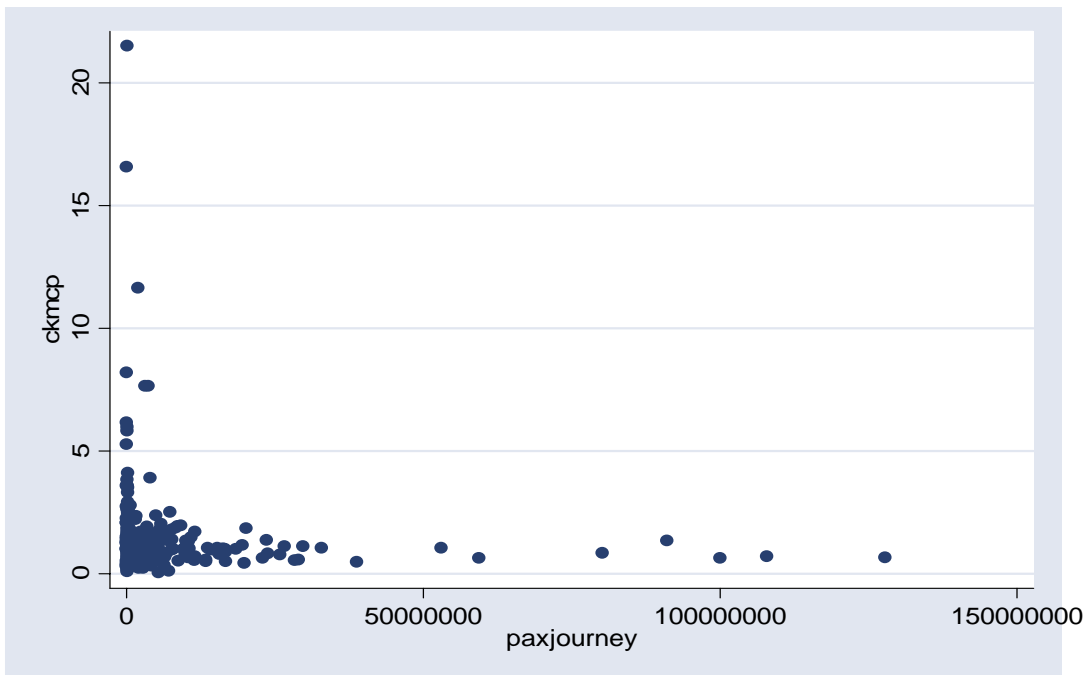
E.36 This is illustrated in the panels of Figure E.2 below:

- figure E.2a shows that there is greater variability of prices for smaller LTAs
- figure E.2b shows that there is greater variability in kilometres per passenger in smaller LTAs (in fact, this figure shows a very similar pattern to the scatter plot of the squared residuals against the number of passenger journeys shown in Figure E.1 above)
- figure E.2c shows that there is a greater variability in the proportion of retired population in smaller LTAs
- figure E.2d shows that there is a greater variability in the proportion of households with no car in smaller LTAs.

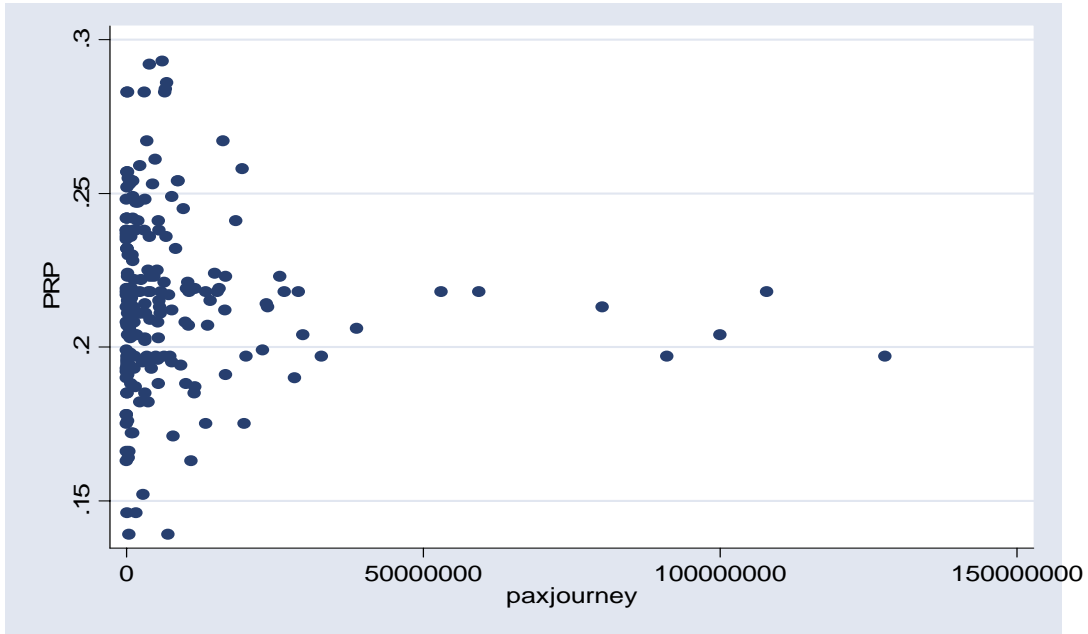
**Figure E.2a: Scatter plot of bus fares against passenger journeys**



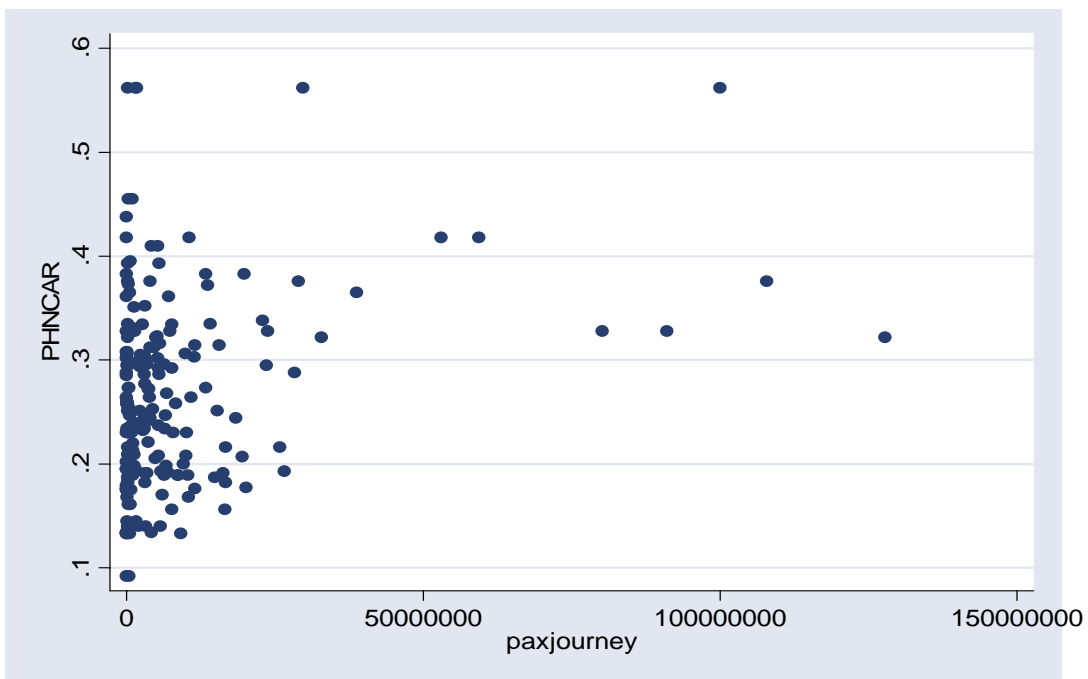
**Figure E.2b: Scatter plot of commercial km per passenger journey against passenger journeys**



**Figure E.2c: Scatter plot of the proportion of retired population against passenger journeys**

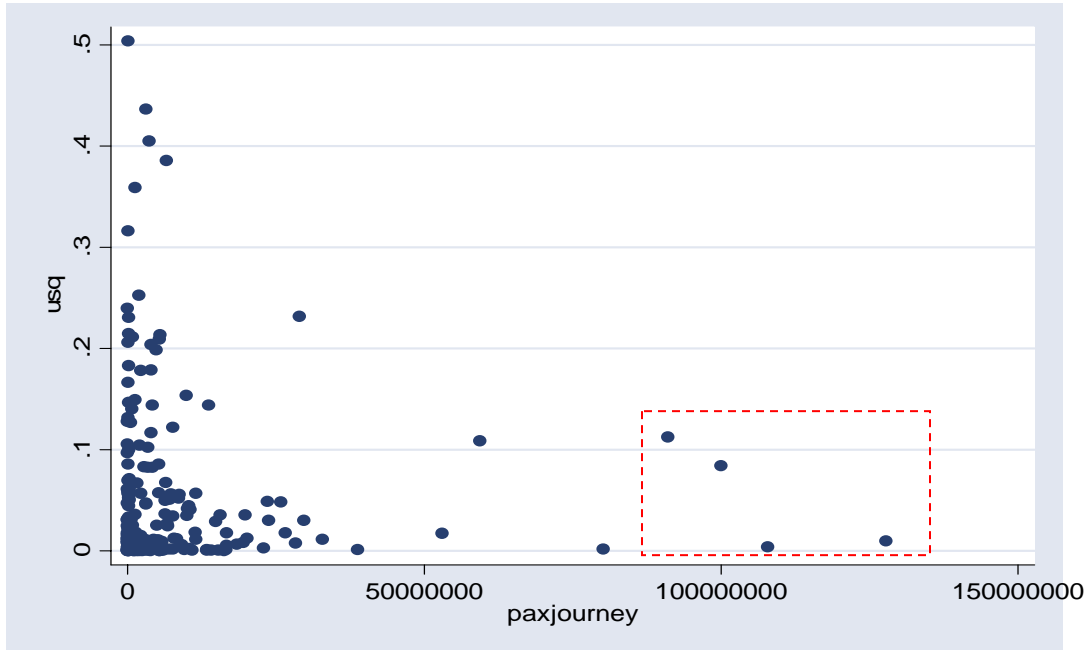


**Figure E.2d: Scatter plot of the proportion of households with no car against passenger journeys**



- E.37 Therefore, we do not accept Frontier's assertion that the data shown in Figure E.1 is evidence that there are measurement errors in data for small LTAs. Rather, as we demonstrated above, it is likely to reflect the greater heterogeneity of smaller LTAs.
- E.38 If the larger variance of the error term for smaller LTA is due to greater variability of the dependent variable and the explanatory variables for these LTAs, and not due to the measurement error as argued by Frontier, our parameter estimates are unbiased but the conventional estimates of standard errors of the coefficients may be incorrect. However, the results of our analysis are not affected by this problem. In the Additional note on the price concentration we have presented the corrected standard errors that are robust to the problem of a non-constant error variance.
- E.39 This notwithstanding, we have considered the WLS procedure which was implemented by Frontier.
- E.40 It is known that weighting data when the variances are suspected to be non-constant is only recommended if the weights are known, or if there is justification that they follow a particular form. However, we have seen no evidence that Frontier has consider the appropriateness of the weights that they applied other than from eyeballing the data shown in Figure E.1. On the basis of the visual inspection of this data, it is far from clear to us that the relationship depicted in Figure E.1 is best captured by using the square root of passenger numbers as the appropriate weight.
- E.41 Indeed, because of the large difference in the size of LTAs, Frontier's choice of the weights resulted in the handful of very large LTAs being given very large weights, so much so that the outcome of the WLS analysis that the 'monopoly' variable is not significant is driven entirely by just four biggest LTAs in the sample. These four LTAs are highlighted in Figure E.3 below.

**Figure E.3: Scatter plot of squared residuals against passenger journeys**



E.42 To illustrate this, we have excluded these four LTAs from the analysis and re-estimated Frontier's model. The results are shown in the table below.

**Table E.3: Weighted Least Squares Regression**

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
Constant	-0.476	0.600	0.429	-1.660	0.708
single national operator	0.075	0.042	0.073	-0.007	0.158
km per passenger	0.206	0.029	0.000	0.150	0.263
proportion of households with no car	0.091	0.118	0.442	-0.141	0.322
proportion of unemployed	-0.314	0.122	0.011	-0.555	-0.073
operating cost per km	0.090	0.039	0.023	0.012	0.167
proportion of population under 15	0.127	0.290	0.662	-0.445	0.698
proportion of population over 60	0.181	0.167	0.279	-0.148	0.510
first dummy	0.578	0.193	0.003	0.197	0.959
arriva dummy	-0.427	0.385	0.268	-1.186	0.331
$R^2 = 0.40$					

E.43 It can be seen that the ‘monopoly’ variable is now statistically significant at the 10 per cent significance level. It suggests that prices are higher by about eight per cent in monopoly areas.

### Data Accuracy

E.44 Frontier told us that they considered the accuracy of the data underlying the 15 high leverage observations in our data set. Specifically, they asked the operators to comment on the accuracy of the price and the concentration data used in our analysis. According to Frontier, the operators identified errors in the way we measured both price and concentration in our study.

E.45 Frontier amended the data set to reflect the comments made by the operators and re-estimated all 29 regressions which were presented in our Additional note on price concentration. Frontier reported that the outcome of their analysis was that our result that prices were

higher in 'monopoly' areas was no longer statistically significant in any of our regressions.

## Price data

- E.46 The operators told Frontier that the prices in our data set were incorrect for seven out of the 15 high leverage observations.
- E.47 We asked Frontier and the operators to provide us with an explanation as to what data sources and methods they used in their process of data validation. When we examined the operators' responses, we found that they had used a different method to calculate the price variable from the one which we had used in our analysis, and that the geographical units, and in some case the time frame covered, were different from those used in our study.
- E.48 In Annexe C of the Proposed Decision we explained that the focus of our analysis was on the average price charged by national operators in LTAs across Great Britain<sup>126</sup> in 2006/07 which was the latest year for which data was available to us. We also explained that we calculated the average price for the year 2006/07 by dividing operators' commercial revenue by our estimate of the number of paying passenger boardings, using the data which the operators provided to the DfT for each LTA in Great Britain. We applied this method consistently across all the operators and all the LTAs in our data set.
- E.49 Our examination of the data used by Frontier revealed the following:
- Operators have used a variety of different methods to estimate the average price, in most cases based on some type of a single fare (for example, 'for a typical route', 'highest', 'average').<sup>127</sup>

---

<sup>126</sup> Excluding London and W. Midlands.

<sup>127</sup> Only in one case we were told that the average price was calculated using data on all ticket types.

- Most areas for which the operators provided alternative estimates of the average price do not correspond closely to the LTA areas in our data set.
- E.50 Virtually all bus operators provide a range of fare types, including single fares, day fares, as well as weekly and monthly bus passes. Even single fares charged by an operator within the same LTA may vary considerably for reasons such as stage length. This being the case, basing the calculation of the average price on listed fares involves a large degree of subjectivity which can produce different results depending on which fare, or a group of fares, is used to estimate the average price.
- E.51 Furthermore, without knowing how many passengers buy different fare types, there is no way of knowing whether an estimate based on single fares is an approximation of what passengers purchasing single fares pay on average, let alone of the overall average price for all fares types which is the measure used in our price concentration analysis. For this reason, where there are multiple fares on offers and no data on the quantity purchased, the appropriate method to calculate the average price should be based on data on revenues and journeys.
- E.52 Another problem with the validation exercise carried out by Frontier and the operators is that their analysis is based on areas which do not correspond to the LTA areas in our data set. For example, in the case of one LTA, the operator's estimate of the average price is based on a flat fare in the main urban area in this LTA. However, the same operator provides about 30 services in this LTA many of which are outside its main urban area. Even if the flat fare was representative of the average price paid by passengers in the main urban area, it may not be for the whole of this LTA.
- E.53 On this basis, our view is that the operators have not actually identified that any of our numbers are wrong. They have instead offered alternative estimates of the average price using a different method which, for the reasons given above, is unlikely to yield an

accurate measure of what passengers pay on average in the LTAs in our data set.

- E.54 Furthermore, even if there were some merits to the method used by the operators to calculate the average fare, it would be wrong in principle to apply it selectively to only part of the dataset in the way that Frontier has done. Using two different methodologies to define the same variable could itself be a source of bias, especially if one methodology is targeted at a specific set of observations, in this case the LTAs classified as monopolies with high prices.
- E.55 We asked Frontier to tell us whether it asked the operators to examine the fares data only for the 15 leverage observations or for the whole sample. Frontier told us that the operators identified problems with the average price for a number of LTAs, and not just for the 15 leverage points (14 of which are in monopoly areas), but that these were not considered further as they were not relevant to the analysis that it carried out and because the limited time available to carry out the analysis did not allow it to validate all of the LTAs with the operators.

### **Truncation point**

- E.56 Frontier argued that we had provided no clear explanation for our choice of cut-off for the fares data, and that the truncation point chosen by us has a substantial impact on the results. According to Frontier, lowering the cut-off point from £3.50 to £3 results in just two observations being dropped, but the 'monopoly effect' being no longer statistically significant. Frontier also told us that its validation exercise suggested that our fare estimates for those two observations were due to measurement error.
- E.57 As explained at paragraph C.29 in the Proposed Decision at paragraph E.9 above, we used the data which was supplied to us by Stagecoach to identify as outliers and exclude from our data set any observations where the average fare was less than 50p or more than £3.50. This is because the reporting units for Stagecoach are more

similar to the way that DfT's data are reported than is the case with the data which we received from other operators.<sup>128</sup>

E.58 Our reasoning for not accepting the alternative fares estimates provided by the operators is explained at paragraphs E.49-E.55 above. This notwithstanding, even if we lower the truncation point to £3, or use the alternative estimates provided by the relevant operator, we find that 20 out of 22 of our 'monopoly' regressions<sup>129</sup> presented in the Additional note on price concentration yield the statistically significant result that the fares are higher in 'monopoly' areas. In our view, this is at odds with Frontier's argument that the truncation point chosen by us has a substantial impact on the results.

## Concentration data

E.59 Two issues were raised in relation to our measurement of concentration. First, Frontier argued that our measure of concentration failed to take into consideration large regional or municipal operators. According to Frontier, in our data set an operator may be identified as the monopoly provider in an LTA even where it faces a regional or municipal operator that is the largest provider of services in the LTA. Second, according to Frontier the operators have identified numerous LTAs where they have been classified by the OFT as a 'monopoly' but where they face competition from other national operators.

E.60 In our original analysis we have included municipal and regional operators in the category of small operators which were assigned a separate variable in our regression analysis. Overall, our findings suggest that small operators do not exert significant competitive pressure on the national operators. That said, we accept that there may be a case for treating the municipal and regional operators as a separate category though we expect that an omission to do so would

---

<sup>128</sup> We are not able to give further detail on this for the reasons of confidentiality.

<sup>129</sup> In the Additional Note on price concentration we present the results of 29 regressions of which 22 measured the impact of 'monopoly' on bus fares whereas seven measured the impact of an additional competitor.

most likely lead to a weaker estimate of the relationship between monopoly market structure and bus fares.

- E.61 This notwithstanding, we have considered the hypothesis about the importance of municipal operators, which was put forward by Frontier and the operators, by assigning a separate variable to municipal operators and re-estimating all 22 of our 'monopoly' regression specifications.<sup>130</sup> The variable representing the presence of a municipal operator is not statistically significant in any of our models and the parameter estimates on other variables remain unchanged. At the end of this section we present the results of four regressions (Tables E.4-E.7 which are based, respectively, on specifications in Tables C.3A, C13A, N.1, and N.2 in the Additional note on the price concentration).
- E.62 In our Proposed Decision, we have classified as monopoly areas those LTAs in which there was only one national provider of commercial bus services in 2006. This classification used the data on the number of commercial kilometres which the operators report to the DfT. Specifically, we classified as monopolies those LTAs where only one national operator reported commercial kilometres in 2006/07.
- E.63 The operators have now suggested that some of that data, which they submitted to the DfT and which we had asked them to validate before we undertook the analysis, is wrong. As a result, the operators and Frontier told us that there are a number of LTAs in our data set which are incorrectly classified as monopoly areas. This includes 5 out of 15 leverage observations which were taken into account by Frontier for the purpose of its analysis.

---

<sup>130</sup> We have included the following municipal operators: Blackpool Transport in Blackpool, Islwyn Borough Transport in Caerphilly, Cardiff Bus in Cardiff, Lothian Buses in Edinburgh, Halton Transport in Halton, Ipswich Buses in Suffolk, Newport Transport in Newport, Nottingham City Transport in Nottingham, Plymouth Citybus in Plymouth, Reading Transport in Reading, Thamesdown Transport in Swindon, Warrington Transport in Warrington and Rossendale Transport in Lancashire. Because our analysis is based on 2006/07 data, we have also included Preston Bus in Lancashire and Eastbourne Bus in East Sussex even though they have subsequently become part of Stagecoach Group.

- E.64 We asked Frontier and the operators to provide us with further detail on the services provided by competitors in the areas which, according to them, are incorrectly classified as monopolies in our data set. In one case the competing services in 2006 were, in fact, not commercial services. As our analysis is based on commercial and not supported services, we do not accept that this LTA is wrongly classified as monopoly area.
- E.65 In four other cases we satisfied ourselves that all but one of the competing services listed by the operators were in fact commercial services in 2006. We therefore tested whether reclassifying these four LTAs as having more than one national operator had an impact on our results. We have also considered the impact of reclassifying six other LTAs as having more than one national operator on the basis of a submission received from one of the bus operators.<sup>131</sup>
- E.66 With the revised data set we have obtained the following results (see Tables E.8 – E.35 at the back of this section):
- a. When we reclassify the four high leverage monopoly LTAs as areas where there are two or more national operators, we find that the monopoly effect is statistically significant (at least at the 10% level) in 20 out of 22 ‘monopoly’ regressions.
  - b. When we reclassify 10 LTAs (that is, using the information supplied in the submission from the bus operator) we find that the monopoly effect is statistically significant in all 22 ‘monopoly’ regressions.

On the basis of the models described at paragraph 60.b. above, we note that our conservative estimate is that fares are higher in ‘monopoly’ areas by about nine per cent, which is the result we highlighted in our Proposed Decision.

---

<sup>131</sup> The bus operator identified a number of additional LTAs where it faced competition from another national operator that had not been reclassified by Frontier for the purpose of its analysis.

- E.67 We have also estimated the regressions described at the paragraph 60.b. above with an explanatory variable representing the presence of a municipal operator (as defined at paragraph E.61 above). The regression results remain unchanged (that is, the monopoly impact is statistically significant in all 22 'monopoly' regressions).
- E.68 Therefore, we acknowledge that there may be some errors in our measurement of concentration notwithstanding the fact that we validated the data with the operators before using it. Nevertheless, in our view, these errors do not undermine the finding that fares are higher in monopoly areas.

### **Comments made by the operators**

- E.69 One operator told us that it could not reconcile any of the data (fares, revenue, costs) that we used with its records, or the STATS100 data, and that the data cited as fares in our report are, in fact, not fares but an indication of passenger revenue. The same operator provided its own estimates of fares, passengers and costs for a number of LTAs in order to illustrate that the data which we used was inaccurate. This operator also told us that it was not reasonable for us to use STATS100 data when the operators provided us with their own data.
- E.70 As we explain at paragraph E.20 above, in the course of our study we told the operators that we would use STATS100 data for the purpose of our analysis. At paragraph C.28 in the Proposed Decision we explained our reasons for using STATS100 data in preference to the operators' own data, namely that STATS100 was the only consistent data set which allowed us to compare the average fare charged by national operators at the level of local transport authorities.

- E.71 In the course of our study, we sent the STATS100 data to the operators and asked them to comment on its validity,<sup>132</sup> as well as on the methodology which we used to calculate average fares across different LTAs. The operators found very few errors with the datasets that we sent them and, as explained at paragraphs E.22 and E.23 above, where errors were found we effectively excluded those data points from our analysis.
- E.72 DfT told us that they have made some adjustments to the raw passenger data supplied by the operators to account for under-recording of ticketless passengers by the operators. Therefore, in our view the figures adjusted by DfT are likely to be a more accurate representation of the correct number of concessionary passengers than the operators' own unadjusted figures.
- E.73 At paragraphs E.50 – E.53 above we explain why we consider that the alternative fares estimates provided by the operators are not an accurate representation of the average price.
- E.74 Two operators told us that there is no correlation between costs and prices in our regression models, and that this suggests that our costs are mis-measured. This is not correct. The cost variable is statistically significant in almost all our regression models. The cost variable is not significant though in the model estimated by Frontier. However, as we explain at paragraphs E.90 – E.102 below where we discuss the issue of costs in a more detail, our view is that the cost variable constructed by Frontier is flawed.

## **Control for service quality**

- E.75 Frontier argues that the variables used in the OFT's analysis are unlikely to pick up adequately differences in service quality. The

---

<sup>132</sup> Specifically, we asked the operators to validate their revenue, passengers and kilometres operated data. This is because some of the operators told us that the quality of this data may be questionable due to the fact that the operators do not record their data at LTA level. We did not ask them specifically to comment on the validity of the cost data because this data was provided at an operator rather than LTA level, and because none of the operators expressed any concern to us about this data.

reason given by Frontier is that our main proxy for service quality – the average number of kilometres operated per passenger – is not a good measure of service quality. Frontier attributes the fact that this variable is highly statistically significant in all our regressions to the possibility that it is most likely picking up differences in the costs of serving different LTAs.

- E.76 The first point to note is that we have used the average number of kilometres operated per passenger as a proxy for frequency. Frequency is just one of several aspects of service provided by bus operators. Other aspects of service relate to reliability, cleanliness, customer service and fleet quality. In our regressions we have controlled comprehensively for fleet quality by including several relevant variables, namely the average age of the fleet, the number of buses with a disability access certificate, low floors, GPS and CCTV equipment. Overall we did not find that these variables account for variation in bus fares across different LTAs.
- E.77 However, in our related econometric work on the relationship between fleet quality and competition, which is presented in Annexe F, we have shown that the quality of the fleet is related to the degree of competition faced by the operators. Specifically, the operators which face less competition tend to operate older fleets with less GPS and CCTV equipment. Consequently, higher prices in monopoly areas cannot be justified by better fleet quality. We note that Frontier has not contested this finding.
- E.78 Turning to the question of frequency, Frontier argues that our proxy for this aspect of service, namely the average number of kilometres operated per passenger, will merely reflect differences in LTA geography. For example, large rural areas with low service frequencies may have a higher average number of kilometres operated per passenger than high density urban LTAs. Therefore, according to Frontier, a higher number of average kilometres operated per passenger journey in a particular LTA tells us nothing about service levels provided by the operators.

- E.79 We accept that the average number of kilometres per passenger is not a perfect proxy for frequency, but we disagree with Frontier’s untested argument that it merely reflects LTA geographies. In all likelihood it will reflect both: the stage length and the frequency of service. All else being equal, for the same number of passengers, higher frequencies will result in a higher average number of kilometres per passenger. Similarly, for any given number of passengers, longer stage lengths will result in a higher average number of kilometres per passenger.
- E.80 This notwithstanding, we have examined the data to test the claim made by Frontier in respect to our proxy for frequency – the number of kilometres per passenger journey. If Frontier is right that this variable merely reflects different LTA geographies, we would expect to find a relationship between the average number of kilometres per passenger and local authority type and a relationship between the average number of kilometres per passenger and area size. In fact, we could not find any. In one regression, we related the average number of kilometres per passenger journey to the variables indicating local authority type (for example, urban, rural, PTE) and, in another regression, to the area size. There were no statistically significant parameters in either regression (see Tables E.36 and E.37 at the back of this section).
- E.81 Therefore, while we accept that there may be some noise in how we measure one (out of several) aspects of service quality, although not on the scale suggested by Frontier, we disagree with Frontier’s characterisation that in our regressions we do not control adequately for service quality.

### **Consistency of approach to control variables**

- E.82 Frontier has criticised our approach to control variables. According to Frontier, in a number of cases the regressions drop control variables that have been highlighted as important by the parties and that were included in the OFT’s original analysis. Frontier argues that this is an inconsistent approach which has potentially important effects on the analysis.

- E.83 The issue mainly concerns our treatment of a single control variable – the proportion of ABs. Frontier points out that our original specification included the proportion of ABs in each LTA and that in the amended regression this variable has been inappropriately dropped from the analysis. Frontier reintroduced this variable into the analysis and found that our results change. Specifically, Frontier reports that the finding that fares in monopoly areas are higher is no longer statistically significant in the baseline regression.
- E.84 We explained in Annexe C of the Proposed Decision (paragraph C.22 -C.23) that we wanted to allow for the argument put to us by the operators that fares may be higher in areas where the population have higher incomes. However, since an income variable was not available to us at the level of LTAs we needed to find an adequate proxy. We considered two proxies: the proportion of the population which is unemployed, which we expect to be negatively correlated with income, and the proportion of ABs which should be positively correlated with income.
- E.85 Of the two, only the unemployment variable was both statistically significant and had the expected sign. The latter indicates that this variable has performed the role for which it was introduced in the regression, namely to proxy for the effect of income. The proportion of ABs was significant, but was not of the expected sign. It suggested that income has a negative impact on what people pay for bus services. This is the opposite effect to the one that the bus operators said that we should expect to find. We see no compelling reason for keeping in the model a proxy variable which is not performing the role for which it was considered.
- E.86 Frontier offered an alternative rationale for the use of ABs, namely that that higher income is correlated with car ownership. Since, according to the operators, the car is a key competitor of local bus services, this could readily explain the finding that prices are lower in areas where the proportion of ABs is higher. This fails to appreciate that we control for the impact of car travel on bus fares by including in our regression the proportion of households with no car – clearly the

more relevant measure of the competition from car travel than the proportion of ABs. This variable is highly statistically significant in all model specifications and of the expected sign.

- E.87 In any case, the omission of the proportion of ABs from the analysis does not make a material difference to our results. As indicated above, Frontier reports that when they reintroduced the proportion of ABs as an explanatory variable, the monopoly variable is no longer significant (p value = 0.113) in one of our monopoly regressions (out of 22 presented in the Additional note on price concentration). We also note that when this regression is estimated in a semi-logarithmic form, the parameter estimate associated with the 'monopoly' variable is statistically significant and suggests that prices are higher by about 8% in LTAs with only one national provider of bus services.
- E.88 Frontier also notes that in the Additional note on price concentration we have omitted the number of paying passengers and the number of small competitors. There is nothing unusual in this. These two variables were not statistically significant in any of the regression specifications and we reported both the estimates with and without these two variables. As is apparent from the Additional note on price concentration, the only effect of omitting these two variables from the analysis was to increase the efficiency with which we estimated the parameter associated with the monopoly variable which is the key effect we are interested in.
- E.89 Therefore we disagree with Frontier's claim that our 'key results can be substantially affected by the control variables included in the analysis'.

### **Measurement of the cost variable**

- E.90 Frontier argues that our cost variable is subject to considerable measurement error. According to Frontier, this is because our costs are measured only at a regional level and not at the level of LTA. Frontier is concerned that a region may have a number of larger metropolitan areas with low operating costs and a small number of

small rural LTAs where the costs of operating bus routes are high. At a regional level, because the rural LTAs are small, the high costs associated with the rural LTAs are not fully reflected in the data because of the impact of the low cost metropolitan area.

- E.91 The first point to note is that our price concentration regression is a reduced form of the structural model and that, as such, it does not require the use of firms' actual costs.<sup>133</sup> Rather, it needs to control for the cost drivers that are outside their control. There are likely to be two main sources of exogenous cost variation which we need to control for. One is due to the differences in regional economic development within the UK which are well documented. For example, other than the cost of fuel, an operator based in the South East is likely to have different costs from an operator providing services in Wales. We control for these differences comprehensively through the combination of relying on the operators' own cost estimates (at the level of subsidiaries), fleet characteristics and the use of region specific dummy variables.
- E.92 The cost may also vary for services provided by the same operator in different areas of its operations, which is the point made by Frontier. We note though that operators often provide services in two or more LTAs from the same depot in which case the cost differences should not be significant. If there are cost differences between LTAs served by the same operator which are reflected in prices charged, these should be picked up by the LTA-type variables which are included in our regressions (see Tables C.9A, N.3 and N.4 in the Additional note on the price concentration).
- E.93 Specifically, we control for the following LTA types: PTE, large urban areas, medium urban areas, mixed urban/rural areas, mostly rural, mostly urban and other larger conurbations. We note that the LTA-type variables are highly statistically significant in most regressions as indeed is our main cost proxy – the operating cost per kilometre of service.

---

<sup>133</sup> If the analysis were to include proxies for cost that the firm controls, the estimated model would no longer be a valid reduced form.

- E.94 We also note that if there is any impact on the price from the local wage related cost differences, which is not controlled for adequately by the variables discussed at paragraphs E.92-93 above, this may be captured in part by our unemployment variable.
- E.95 Therefore, we do not accept Frontier's argument that our results may be biased because we failed to control for costs in our regression.
- E.96 Frontier also put forward an alternative cost measure, namely cost per passenger journey. Specifically, Frontier transformed the data from cost per kilometre (commercial and subsidized) which we calculated for the operating groups into an LTA-level cost per passenger measure. Frontier argues that transforming the data in this way highlights concerns regarding the accuracy of the data cost. To illustrate this, Frontier produced a graph of the distribution of operating cost per passenger showing that in a number of LTAs cost per passenger journey exceeds £10 whereas in other LTAs cost per passenger journey is less than 25p.
- E.97 Frontier used the cost per passenger data to calculate price cost margins earned by the operators in all the LTAs in our sample. Frontier then carried out a margin concentration analysis by regressing price cost margins on the same explanatory variables that we used in our price concentration regressions. Frontier suggests that if our finding that prices are higher in monopoly areas is correct, and the cost data is accurate, then it would expect to see the 'monopoly' variable having a positive impact on the margins the operators make in different LTAs. However, Frontier found that the monopoly effect is not statistically significant in its margins regression.
- E.98 We were, and remain, of the view the cost per kilometre is a more appropriate unit cost than the cost per passenger. This is because the two biggest variable cost components – drivers' wages and fuel costs<sup>134</sup> – are evidently closer related to the number of kilometres

---

<sup>134</sup> During the consultation process, the operators told us that drivers' wages and fuel cost are two biggest variable cost components.

operated than to the number of passengers carried. Indeed, the cost per kilometre measure is widely adopted as measure of unit cost for the bus industry and indeed for other transport sectors.<sup>135</sup> We also note that unlike in the case of revenue and passenger kilometres data, the operators have not expressed reservations about the accuracy of the cost data which they supplied to the DfT.

- E.99 Furthermore, the method used by Frontier to transform the cost per kilometre, which we calculated for each bus operator in our sample, into an LTA cost per passenger measure is, in our view, flawed. Specifically, we calculated the cost per kilometre by dividing total operating expenditure by **total** number of kilometres operated. The latter includes both subsidized and commercial kilometres when an operator provides both commercial and supported services. Omitting either of the two would inflate the cost per kilometre measure since both commercial and supported services affect operating expenditure.
- E.100 However, in transforming our cost per kilometre operator based measure to an LTA cost per passenger measure Frontier failed to take into account subsidized kilometres in its calculations. Specifically, Frontier scaled our cost per kilometre measure for each operator by the ratio of the commercial kilometres to all passenger boardings (on both subsidized and commercial services) reported by each operator at the level of an LTA.
- E.101 This introduced a major distortion in the data especially for those LTAs where operators provide both subsidized and commercial kilometres. Indeed, in most cases where Frontier reports that costs per passenger are below 25p, and therefore implausibly low, the number of subsidized kilometres is greater than the number of commercial kilometres. The low estimates in most of these cases reflect the fact that subsidized kilometres are not included in the calculations and not some inherent problem with the dataset.

---

<sup>135</sup> For example, see DfT's publication Transport Statistics for Great Britain available at [www.dft.gov.uk/pgr/statistics/datatablespublications/tsgb/2008edition/](http://www.dft.gov.uk/pgr/statistics/datatablespublications/tsgb/2008edition/)

E.102 Therefore, in our view, the results of Frontier's margin concentration analysis are not meaningful as they are based on a distorted dataset.

## **Conclusion**

E.103 Overall, having considered Frontier's analysis very carefully, we do not accept its general conclusion that our own analysis is subject to such serious flaws that it cannot be relied upon.

**E.104** We by no means consider that our work in this area is perfect, we use our findings cautiously and consider them in the context of the other evidence, particularly on short run and long run fares elasticities. Our view is that the findings presented here build on and are supported by the work done by numerous independent academic researchers and that, notwithstanding the arguments raised to the contrary by the bus operators that there is a dimension to competition between rival bus operators which is stronger than the competition which arises from other modes of transport.

## The analysis of the impact of municipal operators

Table E.4: Impact of market structure on bus fares (replication of table C.3A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.124	0.045	0.007	0.035	0.213
count of smaller operators (20 or more buses)	-0.003	0.006	0.606	-0.015	0.009
municipal operator	-0.016	0.059	0.790	-0.131	0.100
km per passenger	0.088	0.034	0.010	0.021	0.155
proportion of households with no car	0.262	0.116	0.025	0.033	0.491
proportion of unemployed	-0.388	0.132	0.004	-0.648	-0.128
operating cost per km	0.093	0.042	0.027	0.011	0.176
paying passenger boardings	0.002	0.008	0.805	-0.013	0.017
constant	-0.944	0.328	0.004	-1.592	-0.297
R <sup>2</sup> = 0.18					

Table E.5: Including proportion of population eligible for concessionary fares and controlling for observations identified by First and Arriva as potentially inaccurate (replication of Table C.13A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.082	0.042	0.053	-0.001	0.165
count of smaller operators (20 or more buses)	-0.004	0.006	0.557	-0.015	0.008
municipal operator	-0.048	0.064	0.448	-0.174	0.077
km per passenger	0.105	0.034	0.002	0.038	0.172
proportion of households with no car	0.459	0.137	0.001	0.189	0.729
proportion of unemployed	-0.530	0.151	0.001	-0.828	-0.232
operating cost per km	0.072	0.041	0.080	-0.009	0.153
paying passenger boardings	0.003	0.007	0.669	-0.011	0.018
proportion of population under 15	0.850	0.365	0.021	0.131	1.570
proportion of population over 60	0.483	0.168	0.005	0.151	0.814
first dummy	0.506	0.079	0.000	0.351	0.661
arriva dummy	-0.298	0.038	0.000	-0.373	-0.224
constant	1.001	0.712	0.161	-0.404	2.406
R <sup>2</sup> = 0.28					

Table E.6: Impact of monopoly market structure on bus fares (replication of table N.1 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.090	0.042	0.032	0.008	0.172
municipal operator	0.102	0.032	0.002	0.039	0.165
km per passenger	-0.054	0.064	0.400	-0.180	0.072
proportion of households with no car	0.464	0.135	0.001	0.197	0.731
proportion of unemployed	-0.527	0.149	0.001	-0.821	-0.233
operating cost per km	0.074	0.041	0.070	-0.006	0.154
proportion of population under 15	0.822	0.355	0.022	0.122	1.522
proportion of population over 60	0.463	0.163	0.005	0.142	0.784
first dummy	0.506	0.077	0.000	0.354	0.659
arriva dummy	-0.302	0.036	0.000	-0.372	-0.232
constant	0.972	0.688	0.160	-0.386	2.329
R <sup>2</sup> = 0.28					

Table E.7: Impact of monopoly market structure on bus fares – semi-log form (replication of table N.2 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.109	0.043	0.012	0.024	0.194
municipal operator	0.022	0.016	0.163	-0.009	0.053
km per passenger	-0.054	0.065	0.408	-0.181	0.074
proportion of households with no car	1.687	0.502	0.001	0.697	2.676
proportion of unemployed	-15.992	4.272	0.000	-24.417	-7.567
operating cost per km	0.023	0.019	0.226	-0.014	0.061
proportion of population under 15	4.955	2.120	0.020	0.775	9.136
proportion of population over 60	2.515	0.882	0.005	0.776	4.254
first dummy	0.448	0.074	0.000	0.303	0.594
arriva dummy	-0.233	0.030	0.000	-0.291	-0.174
constant	-1.368	0.550	0.014	-2.452	-0.284
R <sup>2</sup> = 0.24					

## The analysis using reclassified concentration variables

### Data used in the Frontier report

Table E.8: Impact of market structure on bus fares (replication of Table C.3A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.083	0.043	0.054	-0.001	0.167
count of smaller operators (20 or more buses)	-0.005	0.006	0.412	-0.017	0.007
km per passenger	0.092	0.034	0.007	0.025	0.158
proportion of households with no car	0.241	0.110	0.029	0.024	0.458
proportion of unemployed	-0.366	0.126	0.004	-0.614	-0.117
operating cost per km	0.093	0.042	0.026	0.011	0.176
paying passenger boardings	0.002	0.008	0.768	-0.013	0.017
constant	-0.883	0.322	0.007	-1.519	-0.248
R <sup>2</sup> = 0.16					

Table E.9: Impact of market structure on bus fares (levels regression) (replication of Table C.6A in the additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.138	0.066	0.037	0.008	0.267
count of smaller operators (20 or more buses)	-0.003	0.007	0.656	-0.018	0.011
km per passenger	0.030	0.026	0.236	-0.020	0.081
proportion of households with no car	1.362	0.531	0.011	0.316	2.408
proportion of unemployed	-14.990	4.798	0.002	-24.450	-5.530
operating cost per km	0.039	0.026	0.142	-0.013	0.091
paying passenger boardings	0.000	0.000	0.057	0.000	0.000
constant	1.193	0.100	0.000	0.995	1.391
R <sup>2</sup> = 0.15					

Table E.10: Impact of smaller operators on fares charged by national operators (replication of Table C.7A in the additional note on the price concentration)

e) Small operators with 40 or more buses

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.075	0.043	0.081	-0.009	0.159
count of smaller operators (40 or more buses)	-0.022	0.013	0.089	-0.046	0.003
km per passenger	0.092	0.033	0.006	0.027	0.158
proportion of households with no car	0.238	0.111	0.033	0.020	0.456
proportion of unemployed	-0.360	0.126	0.005	-0.608	-0.113
operating cost km	0.091	0.041	0.029	0.009	0.172
paying passenger boardings	0.002	0.008	0.783	-0.013	0.017
constant	-0.856	0.321	0.008	-1.490	-0.222
R <sup>2</sup> = 0.17					

f) Small operators with 100 or more buses

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.092	0.043	0.034	0.007	0.177
count of smaller operators (100 or more buses)	-0.024	0.028	0.402	-0.079	0.032
km per passenger	0.088	0.034	0.010	0.022	0.155
proportion of households with no car	0.247	0.111	0.027	0.029	0.466
proportion of unemployed	-0.365	0.127	0.004	-0.615	-0.115
operating cost km	0.092	0.042	0.030	0.009	0.175
paying passenger boardings	0.001	0.008	0.904	-0.014	0.016
constant	-0.864	0.322	0.008	-1.500	-0.228
R <sup>2</sup> = 0.16					

Table E.11: Fares and fleet quality (replication of Table C.8A in the additional note on the price concentration)

k) Proportion of buses with disability access certificate

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.070	0.043	0.106	-0.015	0.155
count of smaller operators (20 or more buses)	-0.005	0.006	0.421	-0.017	0.007
km per passenger	0.102	0.033	0.002	0.037	0.168
proportion of households with no car	0.216	0.111	0.053	-0.003	0.435
proportion of unemployed	-0.320	0.129	0.014	-0.573	-0.066
operating cost km	0.123	0.044	0.006	0.036	0.209
paying passenger boardings	0.003	0.008	0.701	-0.012	0.018
proportion of buses with disability access certificate	-0.355	0.125	0.005	-0.602	-0.109
Constant	-0.693	0.335	0.040	-1.354	-0.032
R <sup>2</sup> = 0.19					

l) Proportion of buses with low floor access

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.100	0.043	0.021	0.015	0.184
count of smaller operators (20 or more buses)	-0.004	0.006	0.497	-0.016	0.008
km per passenger	0.084	0.035	0.016	0.016	0.152
proportion of households with no car	0.314	0.109	0.004	0.099	0.528
proportion of unemployed	-0.439	0.127	0.001	-0.690	-0.188
operating cost per km	0.072	0.041	0.075	-0.007	0.152
paying passenger boardings	0.000	0.008	0.997	-0.015	0.015
proportion of buses with low floors access	-0.364	0.123	0.004	-0.608	-0.121
constant	-0.934	0.319	0.004	-1.564	-0.305
R <sup>2</sup> = 0.19					

m) Proportion of buses with GPS

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.083	0.041	0.047	0.001	0.164
count of smaller operators (20 or more buses)	-0.003	0.006	0.645	-0.014	0.008
km per passenger	0.098	0.030	0.002	0.038	0.158
proportion of households with no car	0.228	0.103	0.028	0.025	0.432
proportion of unemployed	-0.378	0.121	0.002	-0.616	-0.140
operating cost per km	0.152	0.045	0.001	0.063	0.241
paying passenger boardings	0.006	0.008	0.418	-0.009	0.021
proportion of buses with GPS	-0.277	0.081	0.001	-0.438	-0.116
constant	-0.929	0.318	0.004	-1.556	-0.302
$R^2 = 0.22$					

n) Proportion of buses with CCTV

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.075	0.042	0.078	-0.008	0.158
count of smaller operators (20 or more buses)	-0.004	0.006	0.541	-0.015	0.008
km per passenger	0.088	0.033	0.009	0.022	0.154
proportion of households with no car	0.226	0.109	0.039	0.012	0.440
proportion of unemployed	-0.315	0.123	0.011	-0.558	-0.072
operating cost per km	0.095	0.042	0.024	0.012	0.178
paying passenger boardings	0.002	0.008	0.754	-0.013	0.018
proportion of buses with CCTV	-0.334	0.094	0.000	-0.520	-0.148
constant	-0.639	0.319	0.047	-1.268	-0.009
$R^2 = 0.21$					

o) Average fleet age

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.084	0.042	0.048	0.001	0.167
count of smaller operators (20 or more buses)	-0.005	0.006	0.412	-0.017	0.007
km per passenger	0.091	0.034	0.007	0.025	0.158
proportion of households with no car	0.240	0.111	0.032	0.021	0.459
proportion of unemployed	-0.362	0.130	0.006	-0.618	-0.107
operating cost per km	0.093	0.042	0.027	0.011	0.175
paying passenger boardings	0.002	0.008	0.770	-0.013	0.017
average fleet age	-0.014	0.078	0.855	-0.168	0.139
constant	-0.843	0.411	0.042	-1.654	-0.032
$R^2 = 0.16$					

Table E.12: Impact of journey distance (replication of Table C.9A in the additional note on the price concentration)

i) Population density

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.090	0.043	0.036	0.006	0.175
count of smaller operators (20 or more buses)	-0.005	0.006	0.438	-0.017	0.007
km per passenger	0.090	0.035	0.012	0.020	0.159
proportion of households with no car	0.195	0.120	0.108	-0.043	0.432
proportion of unemployed	-0.346	0.130	0.009	-0.603	-0.089
operating cost km	0.077	0.044	0.080	-0.009	0.162
paying passenger boardings	0.001	0.008	0.928	-0.014	0.016
population density	0.021	0.014	0.141	-0.007	0.049
constant	-0.888	0.323	0.007	-1.525	-0.251
$R^2 = 0.17$					

j) Area size

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.078	0.043	0.074	-0.008	0.164
count of smaller operators (20 or more buses)	-0.002	0.007	0.783	-0.015	0.011
km per passenger	0.093	0.034	0.008	0.025	0.160
proportion of households with no car	0.221	0.114	0.054	-0.004	0.445
proportion of unemployed	-0.366	0.127	0.004	-0.616	-0.115
operating cost km	0.081	0.044	0.064	-0.005	0.168
paying passenger boardings	0.002	0.008	0.797	-0.013	0.017
area size	-0.016	0.014	0.234	-0.043	0.011
constant	-0.729	0.358	0.043	-1.434	-0.024
$R^2 = 0.17$					

k) Type of LTA

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.117	0.050	0.021	0.018	0.216
count of smaller operators (20 or more buses)	-0.005	0.007	0.478	-0.018	0.008
km per passenger	0.087	0.033	0.009	0.022	0.152
proportion of households with no car	-0.004	0.123	0.974	-0.247	0.239
proportion of unemployed	-0.244	0.119	0.042	-0.479	-0.009
operating cost km	0.081	0.043	0.060	-0.004	0.167
paying passenger boardings	-0.002	0.008	0.771	-0.017	0.013
large urban	-0.147	0.091	0.108	-0.328	0.033
medium urban	-0.250	0.084	0.003	-0.416	-0.084
mixed urban/rural	-0.153	0.085	0.072	-0.320	0.014
mostly rural	-0.316	0.090	0.001	-0.493	-0.139
mostly urban	-0.201	0.075	0.008	-0.349	-0.052
other large conurbations	-0.082	0.078	0.299	-0.236	0.073
constant	-0.575	0.312	0.067	-1.191	0.041
$R^2 = 0.23$					

I) UK regions

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.083	0.051	0.107	-0.018	0.184
count of smaller operators (20 or more buses)	-0.001	0.006	0.841	-0.013	0.011
km per passenger	0.086	0.034	0.011	0.020	0.152
proportion of households with no car	0.178	0.107	0.097	-0.032	0.389
proportion of unemployed	-0.210	0.144	0.148	-0.494	0.075
operating cost km	0.063	0.045	0.166	-0.026	0.152
paying passenger boardings	0.001	0.008	0.878	-0.014	0.017
East of England	0.185	0.066	0.006	0.055	0.316
North East	0.026	0.094	0.781	-0.158	0.211
North West	0.108	0.084	0.203	-0.058	0.274
Scotland	0.166	0.076	0.031	0.015	0.316
South East	0.211	0.077	0.006	0.060	0.363
South West	0.226	0.076	0.003	0.077	0.375
Wales	0.097	0.070	0.168	-0.041	0.235
Yorkshire and the Humber	0.133	0.069	0.053	-0.002	0.269
constant	-0.560	0.398	0.161	-1.346	0.226
R <sup>2</sup> = 0.22					

Table E.13: Using total km per passenger instead of commercial km per passenger as proxy for service levels (replication of Table C.10A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.118	0.050	0.019	0.020	0.217
count of smaller operators (20 or more buses)	-0.004	0.006	0.554	-0.017	0.009
total km per passenger	0.019	0.031	0.530	-0.041	0.080
proportion of households with no car	0.306	0.122	0.013	0.064	0.547
proportion of unemployed	-0.446	0.141	0.002	-0.724	-0.168
operating cost per km	0.049	0.044	0.276	-0.039	0.136
paying passenger boardings	-0.005	0.008	0.583	-0.021	0.012
constant	-0.965	0.344	0.005	-1.643	-0.287
R <sup>2</sup> = 0.11					

Table E.14: Controlling for observations identified by First and Arriva as potentially inaccurate (replication of Table C.11A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.097	0.043	0.024	0.013	0.181
count of smaller operators (20 or more buses)	0.000	0.006	0.952	-0.012	0.012
km per passenger	0.106	0.035	0.003	0.037	0.175
proportion of households with no car	0.301	0.107	0.006	0.089	0.513
proportion of unemployed	-0.371	0.123	0.003	-0.613	-0.129
operating cost per km	0.068	0.041	0.098	-0.013	0.148
paying passenger boardings	0.004	0.008	0.599	-0.011	0.019
first dummy	0.477	0.090	0.000	0.300	0.654
arriva dummy	-0.273	0.036	0.000	-0.343	-0.202
constant	-0.869	0.314	0.006	-1.489	-0.250
R <sup>2</sup> = 0.25					

Table E.15: Including proportion of population eligible for concessionary fares (replication of Table C.12A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.073	0.043	0.091	-0.012	0.158
count of smaller operators (20 or more buses)	-0.007	0.006	0.249	-0.019	0.005
km per passenger	0.091	0.034	0.008	0.024	0.157
proportion of households with no car	0.335	0.132	0.012	0.075	0.596
proportion of unemployed	-0.462	0.147	0.002	-0.751	-0.174
operating cost per km	0.101	0.043	0.020	0.016	0.186
paying passengers boardings	0.002	0.008	0.772	-0.013	0.017
proportion of population under 15	0.644	0.359	0.075	-0.065	1.352
proportion of population over 60	0.222	0.182	0.224	-0.137	0.581
constant	0.344	0.718	0.632	-1.070	1.759
R <sup>2</sup> = 0.17					

Table E.16: Including proportion of population eligible for concessionary fares and controlling for observations identified by First and Arriva as potentially inaccurate (replication of Table C.13A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.071	0.042	0.093	-0.012	0.153
count of smaller operators (20 or more buses)	-0.004	0.006	0.483	-0.016	0.008
km per passenger	0.107	0.034	0.002	0.040	0.174
proportion of households with no car	0.432	0.128	0.001	0.180	0.684
proportion of unemployed	-0.500	0.142	0.001	-0.780	-0.221
operating cost per km	0.072	0.041	0.079	-0.008	0.153
paying passenger boardings	0.004	0.007	0.612	-0.011	0.019
proportion of population under 15	0.813	0.354	0.023	0.115	1.511
proportion of population over 60	0.471	0.167	0.005	0.142	0.800
first dummy	0.535	0.082	0.000	0.372	0.697
arriva dummy	-0.294	0.038	0.000	-0.369	-0.219
constant	0.979	0.705	0.166	-0.411	2.369
R <sup>2</sup> = 0.28					

Table E.17: Instrumental variable estimation (replication of Table C.5A in The additional note on the price concentration)

Instrumented: single national operator, count of smaller operators (20 or more buses), paying passengers boardings.

Instruments: population, medium/urban LTA, urban/rural LTA, PTE, mostly urban, North East, North West.

e) First-stage regression for 'single national operator'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	0.126	0.032	0.000	0.063	0.189
proportion of households with no car	0.151	0.174	0.386	-0.192	0.495
proportion of unemployed	0.094	0.190	0.622	-0.282	0.469
operating cost km	-0.045	0.066	0.492	-0.175	0.084
proportion of population under 15	0.841	0.525	0.111	-0.194	1.877
proportion of population over 60	1.024	0.255	0.000	0.520	1.528
first dummy	0.017	0.145	0.906	-0.268	0.302
arriva dummy	-0.187	0.338	0.581	-0.854	0.480
population	-0.226	0.035	0.000	-0.296	-0.157
medium/urban LTA	0.158	0.089	0.078	-0.018	0.333
urban/rural LTA	0.147	0.070	0.038	0.008	0.286
PTE	0.116	0.113	0.308	-0.108	0.339
mostly urban	0.034	0.068	0.616	-0.100	0.169
North East	-0.378	0.089	0.000	-0.553	-0.203
North West	0.003	0.091	0.972	-0.177	0.184
constant	6.617	1.113	0.000	4.421	8.813
R <sup>2</sup> = 0.45					

f) First-stage regression for 'count of smaller operators'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.865	0.206	0.000	-1.270	-0.459
proportion of households with no car	1.965	1.126	0.083	-0.256	4.186
proportion of unemployed	-0.583	1.231	0.636	-3.011	1.845
operating cost km	-0.002	0.424	0.996	-0.839	0.835
proportion of population under 15	-0.471	3.394	0.890	-7.165	6.222
proportion of population over 60	0.133	1.651	0.936	-3.124	3.391
first dummy	-0.267	0.935	0.776	-2.111	1.577
arriva dummy	0.988	2.187	0.652	-3.325	5.302
population	0.671	0.228	0.004	0.222	1.121
medium/urban LTA	0.411	0.575	0.476	-0.724	1.545
urban/rural LTA	0.913	0.455	0.046	0.016	1.809
PTE	-0.357	0.733	0.627	-1.802	1.088
mostly urban	0.127	0.440	0.773	-0.741	0.996
North East	0.719	0.573	0.211	-0.411	1.850
North West	0.448	0.592	0.450	-0.719	1.615
constant	5.049	7.198	0.484	-9.148	19.246
	R <sup>2</sup> = 0.21				

c) First-stage regression for 'paying passengers boardings'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.132	0.187	0.482	-0.501	0.237
proportion of households with no car	-1.743	1.024	0.091	-3.763	0.278
proportion of unemployed	2.191	1.120	0.052	-0.017	4.400
operating cost km	-0.667	0.386	0.086	-1.428	0.094
proportion of population under 15	4.244	3.087	0.171	-1.844	10.333
proportion of population over 60	2.866	1.502	0.058	-0.097	5.829
first dummy	-0.383	0.850	0.653	-2.061	1.294
arriva dummy	1.740	1.989	0.383	-2.183	5.664
population	2.570	0.207	0.000	2.161	2.978
medium/urban LTA	-0.130	0.523	0.804	-1.162	0.902
urban/rural LTA	-0.202	0.413	0.625	-1.018	0.613
PTE	-1.639	0.666	0.015	-2.953	-0.324
mostly urban	0.908	0.401	0.025	0.118	1.698
North East	-2.085	0.521	0.000	-3.113	-1.056
North West	0.059	0.538	0.913	-1.002	1.121
constant	-12.703	6.547	0.054	-25.616	0.211
R <sup>2</sup> = 0.61					

d) 2SLS estimates

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.262	0.123	0.035	0.019	0.504
paying passenger boardings	0.022	0.042	0.598	-0.060	0.104
count of small operator (20 or more buses)	0.011	0.012	0.400	-0.014	0.035
km per passenger	0.104	0.041	0.012	0.023	0.185
proportion of households with no car	0.390	0.160	0.016	0.075	0.704
proportion of unemployed	-0.485	0.142	0.001	-0.766	-0.205
operating cost km	0.093	0.051	0.070	-0.008	0.193
proportion of population under 15	0.545	0.411	0.186	-0.265	1.356
proportion of population over 60	0.223	0.243	0.360	-0.256	0.702
first dummy	0.556	0.111	0.000	0.337	0.774
arriva dummy	-0.285	0.258	0.270	-0.793	0.223
constant	-0.207	1.215	0.865	-2.603	2.189
R <sup>2</sup> = 0.18					

Table E.18: Impact of monopoly market structure on bus fares – logarithmic form (replication of Table N.1 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.080	0.041	0.054	-0.001	0.162
km per passenger	0.103	0.032	0.001	0.040	0.167
proportion of households with no car	0.435	0.126	0.001	0.186	0.684
proportion of unemployed	-0.493	0.139	0.001	-0.768	-0.218
operating cost per km	0.074	0.040	0.067	-0.005	0.154
proportion of population under 15	0.772	0.342	0.025	0.097	1.447
proportion of population over 60	0.444	0.162	0.007	0.125	0.763
first dummy	0.539	0.081	0.000	0.379	0.699
arriva dummy	-0.298	0.036	0.000	-0.368	-0.227
constant	0.932	0.677	0.170	-0.402	2.267
R <sup>2</sup> = 0.28					

Table E.19: Impact of monopoly market structure on bus fares – semi-log form<sup>136</sup> (replication of Table N.2 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.101	0.043	0.021	0.016	0.186
km per passenger	0.023	0.016	0.149	-0.008	0.054
proportion of households with no car	1.584	0.467	0.001	0.663	2.506
proportion of unemployed	-15.007	3.977	0.000	-22.849	-7.165
operating cost per km	0.024	0.019	0.217	-0.014	0.061
proportion of population under 15	4.631	2.027	0.023	0.635	8.628
proportion of population over 60	2.400	0.872	0.006	0.681	4.118
first dummy	0.490	0.075	0.000	0.342	0.637
arriva dummy	-0.228	0.030	0.000	-0.287	-0.169
constant	-1.294	0.532	0.016	-2.342	-0.245
R <sup>2</sup> = 0.23					

<sup>136</sup> Only dependent variable transformed logarithmically.

Table E.20: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – logarithmic form (replication of Table N.3 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.091	0.044	0.043	0.003	0.178
km per passenger	0.104	0.031	0.001	0.044	0.165
proportion of households with no car	0.114	0.133	0.394	-0.149	0.376
proportion of unemployed	-0.332	0.131	0.012	-0.590	-0.074
operating cost km	0.051	0.039	0.188	-0.025	0.128
proportion of population under 15	0.628	0.364	0.086	-0.090	1.347
proportion of population over 60	0.664	0.210	0.002	0.249	1.079
first dummy	0.512	0.092	0.000	0.331	0.693
arriva dummy	-0.185	0.051	0.000	-0.285	-0.085
large urban	-0.109	0.086	0.208	-0.278	0.061
medium urban	-0.168	0.081	0.040	-0.328	-0.008
mixed urban/rural	-0.128	0.085	0.134	-0.296	0.040
mostly rural	-0.316	0.098	0.001	-0.508	-0.123
mostly urban	-0.193	0.078	0.014	-0.346	-0.040
other large conurbations	-0.003	0.084	0.972	-0.169	0.164
constant	1.291	0.752	0.088	-0.193	2.775
R <sup>2</sup> = 0.33					

Table E.21: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – semi-log form (replication of Table N.4 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.110	0.047	0.019	0.018	0.202
km per passenger	0.024	0.015	0.111	-0.005	0.053
proportion of households with no car	0.377	0.466	0.419	-0.542	1.297
proportion of unemployed	-10.453	3.656	0.005	-17.664	-3.243
operating cost km	0.015	0.018	0.413	-0.021	0.050
proportion of population under 15	3.565	2.021	0.079	-0.422	7.551
proportion of population over 60	3.261	1.043	0.002	1.203	5.318
first dummy	0.473	0.080	0.000	0.314	0.631
arriva dummy	-0.121	0.049	0.015	-0.218	-0.024
large urban	-0.123	0.087	0.159	-0.295	0.049
medium urban	-0.198	0.084	0.019	-0.363	-0.033
mixed urban/rural	-0.151	0.083	0.070	-0.315	0.013
mostly rural	-0.339	0.098	0.001	-0.533	-0.145
mostly urban	-0.219	0.080	0.007	-0.378	-0.060
other large conurbations	-0.013	0.089	0.882	-0.188	0.162
constant	-0.938	0.554	0.092	-2.031	0.155
R <sup>2</sup> = 0.30					

**Data used in Frontier report and additional data provided  
by one of the operators**

Table E.22: Impact of market structure on bus fares (replication of Table C.3A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.103	0.046	0.024	0.014	0.193
count of smaller operators (20 or more buses)	-0.005	0.006	0.434	-0.017	0.007
km per passenger	0.090	0.033	0.008	0.024	0.156
proportion of households with no car	0.250	0.111	0.025	0.031	0.468
proportion of unemployed	-0.378	0.127	0.003	-0.629	-0.127
operating cost per km	0.092	0.042	0.028	0.010	0.175
paying passenger boardings	0.002	0.008	0.786	-0.013	0.017
constant	-0.915	0.324	0.005	-1.554	-0.276
R <sup>2</sup> = 0.16					

Table E.23: Impact of market structure on bus fares (levels regression) (replication of Table C.6A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.166	0.072	0.022	0.024	0.308
count of smaller operators (20 or more buses)	-0.003	0.007	0.665	-0.017	0.011
km per passenger	0.029	0.025	0.248	-0.021	0.080
proportion of households with no car	1.406	0.539	0.010	0.344	2.468
proportion of unemployed	-15.572	4.911	0.002	-25.256	-5.889
operating cost per km	0.038	0.026	0.155	-0.014	0.090
paying passenger boardings	0.000	0.000	0.056	0.000	0.000
constant	1.202	0.100	0.000	1.004	1.400
R <sup>2</sup> = 0.16					

Table E.24: Impact of smaller operators on fares charged by national operators (replication of Table C.7A in The additional note on the price concentration)

a) Small operators with 40 or more buses

explanatory variable	parameter	standard	p-value	95% confidence	
	estimate	error		intervals	
single national operator	0.096	0.045	0.036	0.006	0.185
count of smaller operators (40 or more buses)	-0.021	0.013	0.099	-0.046	0.004
km per passenger	0.091	0.033	0.007	0.026	0.156
proportion of households with no car	0.246	0.111	0.028	0.027	0.466
proportion of unemployed	-0.373	0.127	0.004	-0.624	-0.123
operating cost km	0.090	0.041	0.031	0.008	0.171
paying passenger boardings	0.002	0.008	0.798	-0.013	0.017
constant	-0.890	0.323	0.006	-1.527	-0.253
R <sup>2</sup> = 0.18					

b) Small operator with 100 or more buses

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.110	0.046	0.017	0.020	0.201
count of smaller operators (100 or more buses)	-0.021	0.029	0.460	-0.078	0.035
km per passenger	0.088	0.034	0.010	0.021	0.154
proportion of households with no car	0.255	0.111	0.023	0.035	0.474
proportion of unemployed	-0.376	0.128	0.004	-0.628	-0.124
operating cost km	0.091	0.042	0.031	0.008	0.174
paying passenger boardings	0.001	0.008	0.914	-0.014	0.016
constant	-0.894	0.324	0.006	-1.533	-0.254
R <sup>2</sup> = 0.17					

Table E.25: Fares and fleet quality (replication of Table C.8A in The additional note on the price concentration)

a) Proportion of buses with disability access certificate

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.089	0.046	0.055	-0.002	0.180
count of smaller operators (20 or more buses)	-0.005	0.006	0.444	-0.017	0.007
km per passenger	0.101	0.033	0.003	0.035	0.166
proportion of households with no car	0.224	0.112	0.047	0.003	0.445
proportion of unemployed	-0.332	0.130	0.012	-0.588	-0.075
operating cost km	0.121	0.044	0.007	0.034	0.208
paying passenger boardings	0.003	0.008	0.718	-0.013	0.018
proportion of buses with disability access certificate	-0.347	0.126	0.006	-0.595	-0.100
constant	-0.726	0.338	0.033	-1.392	-0.060
R <sup>2</sup> = 0.20					

b) Proportion of buses with low floor access

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.117	0.046	0.011	0.027	0.207
count of smaller operators (20 or more buses)	-0.004	0.006	0.502	-0.016	0.008
km per passenger	0.083	0.034	0.016	0.016	0.151
proportion of households with no car	0.321	0.110	0.004	0.105	0.537
proportion of unemployed	-0.451	0.129	0.001	-0.705	-0.197
operating cost per km	0.071	0.041	0.082	-0.009	0.151
paying passenger boardings	0.000	0.008	0.986	-0.015	0.015
proportion of buses with low floors access	-0.361	0.123	0.004	-0.604	-0.119
constant	-0.962	0.321	0.003	-1.596	-0.329
R <sup>2</sup> = 0.20					

c) Proportion of buses with GPS

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.100	0.044	0.025	0.013	0.186
count of smaller operators (20 or more buses)	-0.002	0.006	0.658	-0.014	0.009
km per passenger	0.097	0.030	0.002	0.037	0.156
proportion of households with no car	0.236	0.104	0.025	0.031	0.441
proportion of unemployed	-0.389	0.122	0.002	-0.629	-0.148
operating cost per km	0.150	0.045	0.001	0.062	0.238
paying passenger boardings	0.006	0.008	0.431	-0.009	0.021
proportion of buses with GPS	-0.274	0.081	0.001	-0.433	-0.115
constant	-0.955	0.318	0.003	-1.583	-0.327
R <sup>2</sup> = 0.22					

d) Proportion of buses with CCTV

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.097	0.045	0.033	0.008	0.186
count of smaller operators (20 or more buses)	-0.003	0.006	0.581	-0.015	0.008
km per passenger	0.086	0.033	0.010	0.021	0.152
proportion of households with no car	0.235	0.109	0.033	0.019	0.451
proportion of unemployed	-0.328	0.125	0.009	-0.574	-0.083
operating cost per km	0.094	0.042	0.026	0.012	0.177
paying passenger boardings	0.002	0.008	0.773	-0.013	0.018
proportion of buses with CCTV	-0.334	0.095	0.001	-0.521	-0.148
constant	-0.672	0.320	0.037	-1.304	-0.040
R <sup>2</sup> = 0.21					

e) Average fleet age

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.105	0.045	0.020	0.017	0.193
count of smaller operators (20 or more buses)	-0.005	0.006	0.432	-0.017	0.007
km per passenger	0.090	0.034	0.008	0.024	0.156
proportion of households with no car	0.248	0.112	0.028	0.027	0.468
proportion of unemployed	-0.373	0.131	0.005	-0.631	-0.116
operating cost per km	0.092	0.042	0.029	0.009	0.174
paying passenger boardings	0.002	0.008	0.788	-0.013	0.017
average fleet age	-0.022	0.078	0.780	-0.176	0.132
constant	-0.854	0.411	0.039	-1.663	-0.044
$R^2 = 0.17$					

Table E.26: Impact of journey distance (replication of Table C.9A in The additional note on the price concentration)

a) Population density

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.109	0.046	0.019	0.018	0.199
count of smaller operators (20 or more buses)	-0.005	0.006	0.447	-0.017	0.007
km per passenger	0.088	0.035	0.012	0.020	0.157
proportion of households with no car	0.204	0.121	0.094	-0.035	0.442
proportion of unemployed	-0.358	0.132	0.007	-0.618	-0.099
operating cost km	0.076	0.044	0.083	-0.010	0.161
paying passenger boardings	0.001	0.008	0.942	-0.014	0.016
population density	0.020	0.014	0.147	-0.007	0.048
constant	-0.917	0.326	0.005	-1.559	-0.275
$R^2 = 0.18$					

b) Area size

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.098	0.046	0.035	0.007	0.189
count of smaller operators (20 or more buses)	-0.002	0.007	0.794	-0.015	0.011
km per passenger	0.091	0.034	0.008	0.024	0.158
proportion of households with no car	0.230	0.115	0.046	0.004	0.456
proportion of unemployed	-0.378	0.128	0.004	-0.631	-0.125
operating cost km	0.081	0.044	0.066	-0.005	0.167
paying passenger boardings	0.002	0.008	0.812	-0.013	0.017
area size	-0.016	0.014	0.254	-0.043	0.011
constant	-0.767	0.359	0.034	-1.475	-0.058
R <sup>2</sup> = 0.17					

c) Type of LTA

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.138	0.051	0.008	0.037	0.239
count of smaller operators (20 or more buses)	-0.005	0.007	0.469	-0.018	0.008
km per passenger	0.086	0.033	0.010	0.021	0.150
proportion of households with no car	0.005	0.124	0.969	-0.239	0.249
proportion of unemployed	-0.256	0.121	0.035	-0.494	-0.018
operating cost km	0.080	0.043	0.065	-0.005	0.165
paying passenger boardings	-0.002	0.007	0.740	-0.017	0.012
large urban	-0.150	0.091	0.104	-0.330	0.031
medium urban	-0.259	0.085	0.003	-0.426	-0.091
mixed urban/rural	-0.152	0.084	0.071	-0.318	0.013
mostly rural	-0.318	0.089	0.000	-0.493	-0.143
mostly urban	-0.201	0.075	0.008	-0.349	-0.054
other large conurbations	-0.085	0.078	0.273	-0.239	0.068
constant	-0.599	0.313	0.057	-1.216	0.018
R <sup>2</sup> = 0.24					

d) UK regions

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.094	0.050	0.062	-0.005	0.194
count of smaller operators (20 or more buses)	-0.001	0.006	0.894	-0.013	0.011
km per passenger	0.085	0.034	0.012	0.019	0.151
proportion of households with no car	0.188	0.109	0.085	-0.026	0.403
proportion of unemployed	-0.225	0.147	0.128	-0.515	0.065
operating cost km	0.063	0.045	0.162	-0.026	0.152
paying passenger boardings	0.001	0.008	0.883	-0.015	0.017
East of England	0.187	0.066	0.006	0.056	0.318
North East	0.032	0.094	0.736	-0.153	0.217
North West	0.108	0.084	0.202	-0.058	0.274
Scotland	0.168	0.076	0.028	0.018	0.318
South East	0.211	0.077	0.007	0.059	0.362
South West	0.224	0.076	0.004	0.074	0.373
Wales	0.112	0.066	0.093	-0.019	0.242
Yorkshire and the Humber	0.134	0.069	0.053	-0.002	0.270
constant	-0.602	0.403	0.137	-1.398	0.194
$R^2 = 0.21$					

Table E.27: Using total km per passenger instead of commercial km per passenger as proxy for service levels (replication of Table C.10A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.138	0.053	0.010	0.033	0.243
count of smaller operators (20 or more buses)	-0.004	0.006	0.555	-0.016	0.009
total km per passenger	0.019	0.031	0.533	-0.042	0.081
proportion of households with no car	0.314	0.123	0.012	0.071	0.557
proportion of unemployed	-0.458	0.143	0.002	-0.740	-0.177
operating cost per km	0.047	0.045	0.292	-0.041	0.135
paying passenger boardings	-0.005	0.008	0.574	-0.021	0.012
constant	-0.996	0.347	0.005	-1.680	-0.312
R <sup>2</sup> = 0.11					

Table E.28: Controlling for observations identified by First and Arriva as potentially inaccurate (replication of Table C.11A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.116	0.045	0.011	0.026	0.205
count of smaller operators (20 or more buses)	0.000	0.006	0.971	-0.012	0.012
km per passenger	0.105	0.035	0.003	0.037	0.173
proportion of households with no car	0.310	0.108	0.005	0.096	0.523
proportion of unemployed	-0.384	0.124	0.002	-0.628	-0.139
operating cost per km	0.066	0.041	0.106	-0.014	0.146
paying passenger boardings	0.004	0.007	0.613	-0.011	0.019
first dummy	0.477	0.090	0.000	0.301	0.654
arriva dummy	-0.272	0.036	0.000	-0.343	-0.202
constant	-0.900	0.316	0.005	-1.522	-0.277
R <sup>2</sup> = 0.26					

Table E.29: Including proportion of population eligible for concessionary fares (replication of Table C.12A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.093	0.046	0.043	0.003	0.183
count of smaller operators (20 or more buses)	-0.007	0.006	0.274	-0.019	0.005
km per passenger	0.089	0.033	0.008	0.024	0.155
proportion of households with no car	0.338	0.132	0.011	0.079	0.598
proportion of unemployed	-0.469	0.146	0.002	-0.758	-0.180
operating cost per km	0.100	0.043	0.022	0.015	0.185
paying passengers boardings	0.002	0.008	0.787	-0.013	0.017
proportion of population under 15	0.613	0.353	0.084	-0.084	1.310
proportion of population over 60	0.204	0.179	0.258	-0.150	0.557
constant	0.245	0.708	0.729	-1.152	1.643
R <sup>2</sup> = 0.18					

Table E.30: Including proportion of population eligible for concessionary fares and controlling for observations identified by First and Arriva as potentially inaccurate (replication of Table C.13A in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.089	0.044	0.046	0.002	0.176
count of smaller operators (20 or more buses)	-0.004	0.006	0.517	-0.016	0.008
km per passenger	0.106	0.034	0.002	0.039	0.172
proportion of households with no car	0.434	0.127	0.001	0.183	0.686
proportion of unemployed	-0.506	0.142	0.000	-0.786	-0.227
operating cost per km	0.071	0.041	0.084	-0.010	0.151
paying passenger boardings	0.004	0.007	0.625	-0.011	0.018
proportion of population under 15	0.785	0.349	0.026	0.096	1.473
proportion of population over 60	0.454	0.164	0.006	0.130	0.778
first dummy	0.533	0.083	0.000	0.370	0.696
arriva dummy	-0.293	0.038	0.000	-0.367	-0.219
constant	0.885	0.697	0.205	-0.489	2.259
R <sup>2</sup> = 0.28					

Table E.31: Instrumental variable estimation (replication of Table C.5A in The additional note on the price concentration)

Instrumented: single national operator, count of smaller operators (20 or more buses), paying passengers boardings.

Instruments: population, medium/urban LTA, urban/rural LTA, PTE, mostly urban, North East, North West.

a) First-stage regression for 'single national operator'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	0.114	0.032	0.000	0.052	0.177
proportion of households with no car	0.135	0.174	0.441	-0.209	0.478
proportion of unemployed	0.109	0.190	0.567	-0.266	0.485
operating cost km	-0.023	0.066	0.726	-0.152	0.106
proportion of population under 15	1.025	0.525	0.052	-0.011	2.060
proportion of population over 60	1.063	0.255	0.000	0.559	1.566
first dummy	0.043	0.145	0.766	-0.242	0.328
arriva dummy	-0.182	0.338	0.591	-0.849	0.485
population	-0.188	0.035	0.000	-0.258	-0.119
medium/urban LTA	0.190	0.089	0.034	0.015	0.366
urban/rural LTA	0.097	0.070	0.168	-0.041	0.236
PTE	0.063	0.113	0.580	-0.161	0.286
mostly urban	0.011	0.068	0.867	-0.123	0.146
North East	-0.331	0.089	0.000	-0.506	-0.156
North West	0.014	0.092	0.882	-0.167	0.194
constant	6.504	1.113	0.000	4.308	8.700
R <sup>2</sup> = 0.40					

b) First-stage regression for 'count of smaller operators'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.865	0.206	0.000	-1.270	-0.459
proportion of households with no car	1.965	1.126	0.083	-0.256	4.186
proportion of unemployed	-0.583	1.231	0.636	-3.011	1.845
operating cost km	-0.002	0.424	0.996	-0.839	0.835
proportion of population under 15	-0.471	3.394	0.890	-7.165	6.222
proportion of population over 60	0.133	1.651	0.936	-3.124	3.391
first dummy	-0.267	0.935	0.776	-2.111	1.577
arriva dummy	0.988	2.187	0.652	-3.325	5.302
population	0.671	0.228	0.004	0.222	1.121
medium/urban LTA	0.411	0.575	0.476	-0.724	1.545
urban/rural LTA	0.913	0.455	0.046	0.016	1.809
PTE	-0.357	0.733	0.627	-1.802	1.088
mostly urban	0.127	0.440	0.773	-0.741	0.996
North East	0.719	0.573	0.211	-0.411	1.850
North West	0.448	0.592	0.450	-0.719	1.615
constant	5.049	7.198	0.484	-9.148	19.246
R <sup>2</sup> = 0.21					

c) First-stage regression for 'paying passengers boardings'

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
km per passenger	-0.132	0.187	0.482	-0.501	0.237
proportion of households with no car	-1.743	1.024	0.091	-3.763	0.278
proportion of unemployed	2.191	1.120	0.052	-0.017	4.400
operating cost km	-0.667	0.386	0.086	-1.428	0.094
proportion of population under 15	4.244	3.087	0.171	-1.844	10.333
proportion of population over 60	2.866	1.502	0.058	-0.097	5.829
first dummy	-0.383	0.850	0.653	-2.061	1.294
arriva dummy	1.740	1.989	0.383	-2.183	5.664
population	2.570	0.207	0.000	2.161	2.978
medium/urban LTA	-0.130	0.523	0.804	-1.162	0.902
urban/rural LTA	-0.202	0.413	0.625	-1.018	0.613
PTE	-1.639	0.666	0.015	-2.953	-0.324
mostly urban	0.908	0.401	0.025	0.118	1.698
North East	-2.085	0.521	0.000	-3.113	-1.056
North West	0.059	0.538	0.913	-1.002	1.121
constant	-12.703	6.547	0.054	-25.616	0.211
R <sup>2</sup> = 0.61					

d) 2SLS estimates

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.251	0.135	0.066	-0.016	0.517
paying passenger boardings	0.018	0.041	0.658	-0.063	0.100
count of small operator (20 or more buses)	0.008	0.012	0.525	-0.016	0.031
km per passenger	0.104	0.040	0.010	0.025	0.183
proportion of households with no car	0.401	0.156	0.011	0.094	0.709
proportion of unemployed	-0.498	0.139	0.000	-0.772	-0.225
operating cost km	0.084	0.049	0.090	-0.013	0.180
proportion of population under 15	0.540	0.413	0.192	-0.274	1.354
proportion of population over 60	0.247	0.245	0.314	-0.236	0.731
first dummy	0.548	0.108	0.000	0.335	0.761
arriva dummy	-0.288	0.252	0.253	-0.785	0.208
constant	-0.135	1.254	0.915	-2.607	2.338
R <sup>2</sup> = 0.22					

Table E.32: Impact of monopoly market structure on bus fares – logarithmic form (replication of Table N.1 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.098	0.044	0.028	0.010	0.185
km per passenger	0.102	0.032	0.001	0.040	0.165
proportion of households with no car	0.438	0.126	0.001	0.190	0.685
proportion of unemployed	-0.500	0.139	0.000	-0.775	-0.225
operating cost per km	0.073	0.040	0.073	-0.007	0.152
proportion of population under 15	0.745	0.337	0.028	0.080	1.410
proportion of population over 60	0.429	0.160	0.008	0.114	0.744
first dummy	0.537	0.081	0.000	0.376	0.697
arriva dummy	-0.296	0.035	0.000	-0.366	-0.227
constant	0.844	0.671	0.210	-0.480	2.168
R <sup>2</sup> = 0.28					

Table E.33: Impact of monopoly market structure on bus fares  
 – semi-log form<sup>137</sup> (replication of Table N.2 in The additional  
 note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.116	0.046	0.012	0.025	0.206
km per passenger	0.022	0.016	0.158	-0.009	0.053
proportion of households with no car	1.588	0.467	0.001	0.667	2.509
proportion of unemployed	-15.199	3.997	0.000	-23.081	-7.318
operating cost per km	0.022	0.019	0.241	-0.015	0.060
proportion of population under 15	4.469	2.006	0.027	0.514	8.424
proportion of population over 60	2.322	0.862	0.008	0.623	4.022
first dummy	0.487	0.075	0.000	0.339	0.635
arriva dummy	-0.228	0.029	0.000	-0.286	-0.170
constant	-1.239	0.525	0.019	-2.274	-0.204
R <sup>2</sup> = 0.24					

<sup>137</sup> Only dependent variable transformed logarithmically.

Table E.34: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – logarithmic form (replication of Table N.3 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.108	0.045	0.018	0.019	0.197
km per passenger	0.103	0.031	0.001	0.043	0.164
proportion of households with no car	0.118	0.132	0.372	-0.142	0.379
proportion of unemployed	-0.337	0.131	0.011	-0.594	-0.079
operating cost km	0.051	0.039	0.194	-0.026	0.127
proportion of population under 15	0.592	0.351	0.093	-0.099	1.284
proportion of population over 60	0.635	0.205	0.002	0.230	1.040
first dummy	0.509	0.091	0.000	0.330	0.689
arriva dummy	-0.184	0.050	0.000	-0.283	-0.085
large urban	-0.111	0.086	0.196	-0.280	0.058
medium urban	-0.176	0.081	0.031	-0.336	-0.016
mixed urban/rural	-0.126	0.084	0.133	-0.291	0.039
mostly rural	-0.315	0.096	0.001	-0.504	-0.125
mostly urban	-0.192	0.077	0.013	-0.344	-0.041
other large conurbations	-0.009	0.084	0.917	-0.174	0.156
constant	1.176	0.730	0.109	-0.264	2.616
R <sup>2</sup> = 0.34					

Table E.35: Impact of monopoly market structure on bus fares with LTAs as explanatory variables – semi-log form (replication of Table N.4 in The additional note on the price concentration)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.126	0.047	0.009	0.032	0.219
km per passenger	0.023	0.015	0.120	-0.006	0.052
proportion of households with no car	0.393	0.466	0.399	-0.525	1.312
proportion of unemployed	-10.614	3.669	0.004	-17.852	-3.377
operating cost km	0.014	0.018	0.441	-0.022	0.050
proportion of population under 15	3.388	1.966	0.086	-0.489	7.266
proportion of population over 60	3.129	1.023	0.003	1.112	5.146
first dummy	0.469	0.080	0.000	0.312	0.627
arriva dummy	-0.122	0.048	0.013	-0.217	-0.026
large urban	-0.123	0.087	0.159	-0.295	0.048
medium urban	-0.205	0.084	0.015	-0.370	-0.040
mixed urban/rural	-0.147	0.082	0.076	-0.309	0.015
mostly rural	-0.336	0.097	0.001	-0.527	-0.145
mostly urban	-0.218	0.080	0.007	-0.375	-0.060
other large conurbations	-0.018	0.088	0.843	-0.192	0.157
constant	-0.874	0.538	0.106	-1.935	0.188
R <sup>2</sup> = 0.31					

### Control for service quality

Table E.36: Regression of the average kilometers per passenger on LTA types (dependent variable in logs)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
large urban	0.006	0.268	0.981	-0.522	0.535
medium urban	-0.208	0.250	0.406	-0.700	0.285
mixed urban/rural	0.036	0.220	0.870	-0.397	0.470
mostly rural	0.280	0.226	0.218	-0.166	0.727
mostly urban	-0.204	0.210	0.332	-0.619	0.210
other large conurbations	0.020	0.236	0.933	-0.445	0.485
Constant	0.157	0.179	0.381	-0.196	0.509
R <sup>2</sup> = 0.04					

Table E.37: Regression of kilometres per passenger on the area size  
(both variables in logs)

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
area size	0.039	0.036	0.274	-0.031	0.110
Constant	-0.303	0.407	0.458	-1.105	0.499
R <sup>2</sup> = 0.06					

### Consistency of approach to control variables

Table E.38: Replication of table N2 from the Additional Note on Price Concentration Analysis with the proportion of ABs included as an explanatory variable

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
single national operator	0.082	0.042	0.053	-0.001	0.165
count of smaller operators (20 or more buses)	-0.002	0.006	0.765	-0.015	0.011
km per passenger	0.021	0.015	0.169	-0.009	0.051
proportion of households with no car	1.512	0.465	0.001	0.596	2.429
proportion of unemployed	-18.415	4.329	0.000	-26.953	-9.877
proportion of Abs	-1.225	0.612	0.047	-2.433	-0.018
operating cost per km	0.000	0.000	0.113	0.000	0.000
paying passenger boardings	0.029	0.019	0.138	-0.009	0.067
proportion of population under 15	4.039	2.115	0.058	-0.132	8.209
proportion of population over 60	1.736	0.952	0.070	-0.141	3.612
first dummy	0.491	0.078	0.000	0.336	0.645
arriva dummy	-0.231	0.033	0.000	-0.296	-0.166
Constant	-0.637	0.636	0.317	-1.891	0.616
R <sup>2</sup> = 0.25					

## **F MARKET STRUCTURE AND FLEET QUALITY**

- F.1 This annexe examines the relationship between market structure and fleet quality. Using regression analysis, we find some evidence that the degree of competitive pressure faced by subsidiaries of the four national operators is one of the factors which have a positive impact on the quality of their bus fleets. In particular, our findings suggest that those operators which are exposed to a stronger competitive pressure tend to have younger fleets and proportionately more buses suitable for access by disabled passengers, as well as more buses fitted with satellite navigation devices and CCTV cameras.
- F.2 The analysis is conducted using the same dataset which we use for our price concentration study and which is described in detail in Annexe E. The geographical unit used in this analysis is an operating division of a national operator, for example, Arriva in the Shires and First Chester & Wirral rather than LTA (as in the case of price concentration analysis in Annexe E). This is because we do not know the composition of the fleet employed by each operator at the level of LTAs. This means that some potentially useful information about the relationship between fleet quality and market structure will be lost through data aggregation. For example, an operating division may have a fleet of vehicles of varying quality which may be deployed in different LTAs depending on the level of competition faced in each LTA. We were not able to capture this information in our analysis.
- F.3 However, we expect that each operator's overall fleet composition will reflect the competitive pressure faced across the individual LTAs in which they operate, and that this will be captured using the methodology which is explained at paragraphs F.7 to F.8 below.
- F.4 For each bus operator we consider the following aspects of fleet quality:
- average fleet age

- proportion of buses fitted with GPS or similar devices that are, intended to be used to monitor punctuality
- proportion of buses with Disability Access Certificate or low floor access, and
- proportion of buses with CCTV cameras.

F.5 We relate these four measures using regression analysis to our estimate of the degree of competitive pressure from both other national bus operators and small operators, and to the following explanatory factors which we expect to influence their fleet investment decision:

- total population
- proportion of AB socio demographic group in the population<sup>138</sup>
- proportion of unemployed in the population, and
- proportion of households with no car in the population.

F.6 We expect fleet quality to increase with the intensity of competition from the car and an increase in the affluence of the area, as passengers are likely to demand higher quality of service as their incomes rise. We also control for the population size to allow for the possibility that bigger markets may attract more investment.

F.7 We measured the competitive pressure faced by each operating subsidiary across all the LTAs where it runs services using a method that is best illustrated with an example. Suppose an operator provides services in two LTAs: A and B. We measure the overall competitive pressure faced by this operator by averaging the number of national operators providing commercial or tendered services, or both, across the two LTAs. In addition to a simple average, we also consider the population weighted average to give more weight to larger LTAs as the competitive situation in those

---

<sup>138</sup> ABs is a term used in consumer research for the purpose of socio-economic profiling. It includes persons in senior managerial, administrative or professional roles (group A) and those in a middle managerial role or equivalent (group B).

LTAs may have proportionately bigger effect on the overall fleet composition.

- F.8 We also use the average number of smaller competitors as an additional variable in order to examine the extent to which their presence may impact on the quality of fleet deployed by national operators.
- F.9 Our analysis is based on operators' data for 2006/07. There are 60 operating divisions in our sample. Of that, we excluded three (Go West Midlands, Blue Bus and J W Coaches) where no data was available other than the total number of buses.
- F.10 For each aspect of fleet quality (for example, GPS) we run the following regression

$$p_i = \alpha + f(\beta; MS_i) + \beta X_j + \mu w_{mi} + \epsilon_i$$

where  $p_i$  measures the proportion of buses employed by operator  $i$  which are fitted with GPS,  $f(\cdot)$  is the function which captures the effect of the market structure along the lines explained at paragraphs F.7 to F.8 and  $X_j$  is a vector containing control variables listed at paragraph F.5. We also include a dummy variable  $w_{mi}$  to denote operators which provide services in the West Midlands as we have no information on the presence of National Express which is a major competitor in this area.

- F.11 We use the two measures of national competitors (average and the population weighted average) interchangeably and estimate our regressions in both levels and logarithms.<sup>139</sup> This means that we estimate four regressions for each aspect of quality listed at paragraph F.4 above.

## Results

- F.12 The regression results are summarised in Table D.1 below. For each aspect of fleet quality we present the parameter estimates and their p-values (in brackets) associated with our two measures

---

<sup>139</sup> We have also tried the log-odds transformation in cases where our dependent variable is measured as a proportion. This produced broadly similar results though there were fewer significant coefficients.

of national competitors. The complete results are presented in the tables at the end of this annexe.

**Table F.1: Impact of competition on fleet composition**

	levels		logs	
	Average number of national competitors	Weighted average number of national competitors	Average number of national competitors	Weighted average number of national competitors
Average fleet age	-1.341 **	-0.832	-0.435 **	-0.289 *
<i>p value</i>	0.021	0.120	0.008	0.067
Proportion of buses with GPS	0.240 **	0.178 *	0.802	0.724
<i>p value</i>	0.022	0.060	0.357	0.263
Proportion of buses with disability access certificate or low floors	0.179 **	0.094 *	0.609 **	0.379 *
<i>p value</i>	0.006	0.082	0.014	0.091
Proportion of buses with CCTV	0.110	0.035	2.134 **	1.293 *
<i>p value</i>	0.268	0.683	0.022	0.068

\*significant at 10 per cent, \*\*significant at five per cent

F.13 The parameter estimates shown in Table D.2 suggest that operators who face more competition from national operators tend to have younger fleets. The results also suggest that in areas with more competition between national operators proportionately more buses are suitable for access by disabled passengers and proportionately more buses are fitted with GPS and CCTV.

F.14 Not all the results are statistically significant at the conventional significance levels. In particular, for fleet age, three out of four parameter estimates are statistically significant at five or 10 per cent significance level whereas for other aspects of fleet quality two out of four parameter estimates are significant. We also note that the control variables in our models are in general not statistically significant which may indicate problems with model specification though we do not believe that this undermines the main thrust of our results.

**Table F.2: Dependent variable: average fleet age**

a) Data in levels. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	-1.341	0.562	0.021	-2.469	-0.212
average number of small operators (20 or more buses)	-0.041	0.133	0.762	-0.309	0.228
proportion of unemployed	-15.489	77.096	0.842	-170.418	139.441
proportion of AB's	-17.212	10.908	0.121	-39.132	4.708
proportion of households with no car	-0.153	6.329	0.981	-12.872	12.565
population	1.260E-07	0.000	0.235	0.000	0.000
w. midlands	-1.107	0.611	0.076	-2.335	0.120
Constant	15.646	3.484	0.000	8.645	22.647
R <sup>2</sup> = 0.30					

b) Data in levels. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	-0.832	0.525	0.120	-1.888	0.223
average number of small operators (20 or more buses)	-0.168	0.104	0.114	-0.378	0.042
proportion of unemployed	16.919	87.675	0.848	-159.271	193.109
proportion of AB's	-17.547	12.915	0.180	-43.502	8.407
proportion of households with no car	-3.604	6.904	0.604	-17.478	10.269
population	0.000	0.000	0.203	0.000	0.000
w. midlands	-0.899	0.591	0.135	-2.086	0.288
Constant	14.856	4.008	0.001	6.801	22.910
R <sup>2</sup> = 0.26					

c) Data in logs. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	-0.435	0.157	0.008	-0.751	-0.119
average number of small operators (20 or more buses)	-0.032	0.066	0.623	-0.164	0.100
proportion of unemployed	-0.321	0.456	0.486	-1.238	0.597
proportion of AB's	-0.567	0.428	0.192	-1.428	0.294
proportion of households with no car	0.215	0.337	0.527	-0.462	0.892
population	0.077	0.035	0.032	0.007	0.147
w. midlands	-0.102	0.095	0.287	-0.292	0.088
Constant	-0.327	1.925	0.866	-4.198	3.544
R <sup>2</sup> = 0.27					

d) Data in logs. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	-0.289	0.154	0.067	-0.598	0.021
average number of small operators (20 or more buses)	-0.091	0.053	0.091	-0.197	0.015
proportion of unemployed	-0.080	0.491	0.871	-1.067	0.907
proportion of AB's	-0.564	0.480	0.246	-1.529	0.402
proportion of households with no car	0.000	0.366	0.999	-0.735	0.736
population	0.086	0.036	0.022	0.013	0.158
w. midlands	-0.076	0.094	0.425	-0.265	0.114
Constant	0.049	2.092	0.981	-4.158	4.256
R <sup>2</sup> = 0.24					

**Table D.3: Dependent variable: Proportion of buses with GPS**

a) Data in levels. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.240	0.101	0.022	0.037	0.444
average number of small operators (20 or more buses)	0.012	0.025	0.630	-0.038	0.063
proportion of unemployed	-39.986	20.124	0.053	-80.425	0.454
proportion of AB's	-3.460	2.671	0.201	-8.828	1.908
proportion of households with no car	1.881	1.593	0.243	-1.321	5.084
population	0.000	0.000	0.771	0.000	0.000
w. midlands	0.025	0.133	0.851	-0.243	0.293
Constant	1.363	0.879	0.127	-0.403	3.130
R <sup>2</sup> = 0.25					

b) Data in levels. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.178	0.092	0.060	-0.008	0.364
average number of small operators (20 or more buses)	0.033	0.023	0.165	-0.014	0.079
proportion of unemployed	-46.516	21.033	0.032	-88.783	-4.249
proportion of AB's	-3.640	2.998	0.230	-9.664	2.384
proportion of households with no car	2.451	1.597	0.131	-0.759	5.661
population	0.000	0.000	0.604	0.000	0.000
w. midlands	-0.010	0.131	0.938	-0.274	0.254
Constant	1.535	0.945	0.111	-0.364	3.434
R <sup>2</sup> = 0.23					

c) Data in logs. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.802	0.859	0.357	-0.944	2.548
average number of small operators (20 or more buses)	0.350	0.233	0.141	-0.123	0.823
proportion of unemployed	2.310	2.091	0.277	-1.938	6.559
proportion of AB's	2.034	3.031	0.507	-4.125	8.192
proportion of households with no car	-1.813	1.929	0.354	-5.733	2.106
population	-0.124	0.183	0.504	-0.496	0.249
w. midlands	-0.338	0.495	0.499	-1.343	0.667
Constant	8.560	11.875	0.476	-15.574	32.694
$R^2 = 0.23$					

d) Model in logs. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.724	0.637	0.263	-0.570	2.019
average number of small operators (20 or more buses)	0.415	0.243	0.096	-0.078	0.909
proportion of unemployed	2.050	2.081	0.331	-2.178	6.279
proportion of AB's	2.122	2.963	0.479	-3.900	8.144
proportion of households with no car	-1.566	1.709	0.366	-5.040	1.908
population	-0.167	0.196	0.400	-0.565	0.231
w. midlands	-0.399	0.467	0.399	-1.347	0.549
Constant	8.739	11.704	0.460	-15.046	32.525
$R^2 = 0.23$					

**Table D.4: Dependent variable: Proportion of buses with disability access certificate**

a) Data in levels. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.179	0.063	0.006	0.053	0.305
average number of small operators (20 or more buses)	0.002	0.015	0.910	-0.029	0.033
proportion of unemployed	-13.971	12.195	0.258	-38.479	10.536
proportion of AB's	-0.547	2.165	0.802	-4.897	3.803
proportion of households with no car	1.544	0.828	0.068	-0.119	3.207
population	0.000	0.000	0.037	0.000	0.000
w. midlands	0.255	0.050	0.000	0.155	0.355
Constant	0.283	0.685	0.681	-1.093	1.659
R <sup>2</sup> = 0.32					

b) Data in levels. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.094	0.053	0.082	-0.012	0.201
average number of small operators (20 or more buses)	0.020	0.013	0.122	-0.006	0.046
proportion of unemployed	-17.874	12.970	0.174	-43.938	8.191
proportion of AB's	-0.358	2.359	0.880	-5.098	4.382
proportion of households with no car	2.035	0.851	0.021	0.324	3.746
population	0.000	0.000	0.067	0.000	0.000
w. midlands	0.226	0.049	0.000	0.128	0.324
Constant	0.371	0.747	0.622	-1.131	1.873
R <sup>2</sup> = 0.24					

c) Data in logs. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.609	0.238	0.014	0.130	1.088
average number of small operators (20 or more buses)	-0.017	0.097	0.863	-0.213	0.179
proportion of unemployed	-0.612	0.573	0.291	-1.767	0.542
proportion of AB's	-0.008	0.572	0.988	-1.160	1.144
proportion of households with no car	0.496	0.388	0.208	-0.285	1.277
population	-0.092	0.046	0.049	-0.184	0.000
w. midlands	0.383	0.095	0.000	0.192	0.575
Constant	-1.282	2.610	0.626	-6.540	3.976
$R^2 = 0.30$					

d) Model in logs. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.379	0.219	0.091	-0.063	0.820
average number of small operators (20 or more buses)	0.067	0.077	0.387	-0.088	0.221
proportion of unemployed	-0.914	0.590	0.128	-2.102	0.274
proportion of AB's	0.043	0.579	0.941	-1.123	1.210
proportion of households with no car	0.779	0.434	0.079	-0.095	1.652
population	-0.103	0.052	0.053	-0.207	0.001
w. midlands	0.340	0.096	0.001	0.147	0.533
Constant	-1.616	2.650	0.545	-6.955	3.722
$R^2 = 0.32$					

**Table D.5: Dependent variable: Proportion of buses with CCTV**

a) Data in levels. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.110	0.098	0.268	-0.087	0.308
average number of small operators (20 or more buses)	0.032	0.024	0.194	-0.017	0.080
proportion of unemployed	17.508	15.520	0.265	-13.680	48.696
proportion of AB's	2.401	2.228	0.286	-2.076	6.879
proportion of households with no car	-0.456	1.318	0.731	-3.105	2.192
population	0.000	0.000	0.601	0.000	0.000
w. midlands	-0.080	0.108	0.464	-0.297	0.137
Constant	-0.944	0.660	0.159	-2.271	0.383
R <sup>2</sup> = 0.24					

b) Data in levels. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	0.035	0.086	0.683	-0.138	0.209
average number of small operators (20 or more buses)	0.045	0.020	0.030	0.005	0.086
proportion of unemployed	15.674	15.932	0.330	-16.343	47.691
proportion of AB's	2.707	2.349	0.255	-2.013	7.426
proportion of households with no car	-0.116	1.281	0.928	-2.690	2.459
population	0.000	0.000	0.640	0.000	0.000
w. midlands	-0.099	0.102	0.338	-0.305	0.107
Constant	-0.914	0.716	0.208	-2.352	0.524
R <sup>2</sup> = 0.22					

c) Data in logs. Average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	2.134	0.895	0.022	0.329	3.940
average number of small operators (20 or more buses)	-0.167	0.270	0.539	-0.711	0.377
proportion of unemployed	5.290	2.142	0.018	0.968	9.613
proportion of AB's	2.803	2.330	0.236	-1.899	7.505
proportion of households with no car	-3.829	2.251	0.096	-8.372	0.714
population	-0.294	0.144	0.048	-0.585	-0.003
w. midlands	0.296	0.303	0.334	-0.315	0.907
Constant	18.937	9.161	0.045	0.448	37.425
$R^2 = 0.30$					

d) Model in logs. Weighted average number of competitors as measure of competitive pressure.

explanatory variable	parameter estimate	standard error	p-value	95% confidence intervals	
average number of national operators	1.293	0.691	0.068	-0.101	2.686
average number of small operators (20 or more buses)	0.129	0.234	0.584	-0.343	0.601
proportion of unemployed	4.090	2.208	0.071	-0.365	8.545
proportion of AB's	3.018	2.554	0.244	-2.136	8.173
proportion of households with no car	-2.654	2.149	0.224	-6.991	1.683
population	-0.310	0.145	0.038	-0.602	-0.018
w. midlands	0.105	0.271	0.701	-0.443	0.653
Constant	17.283	10.307	0.101	-3.517	38.084
$R^2 = 0.30$					

## **G ANALYSIS OF HEAD TO HEAD COMPETITION FOR SUPPORTED SERVICES**

- G.1 For each LTA and covering all the major companies, we examined the proportion of cases in which there was direct competition between these companies for tendered contracts. If in an area one company bids for a proportion  $p$  of all contracts, and another bids for a proportion  $q$ , under the assumption that their bidding strategies are entirely independent we would expect the proportion of contracts in which they bid jointly to be  $p \times q$ .
- G.2 We constructed contingency tables and applied statistical tests to see if the observed proportion of joint bids was consistent with this expectation. In many cases bidding patterns were consistent with this expectation, but sometimes there were departures that were large enough to be statistically significant - that is providing evidence that something other than chance was operating.
- G.3 However, not all these cases were significant for having a shortfall in joint bids there were common occurrences in which joint bidding occurred significantly more frequently than would be expected by chance.
- G.4 Overall we conclude there is no evidence of major companies showing any systematic avoidance of direct competition with one another when bidding for supported services. Although an apparent shortfall in direct competition can be seen in some situations, this could reasonably occur for operational business reasons such as depot locations.

**Table G.1: Examination of bidding for supported services in PTE areas: Direct competition between major companies**

<b>Results of analysis</b>	<b>Cases</b>	<b>Percentage</b>
Total cases examined	27	
Cases where contracts attracted bids from major companies consistent with independent business operations	10	37%
Cases where contracts attracted bids from major companies <b>more frequently</b> than consistent with independent business operations	8	30%
Cases where contracts attracted bids from major companies <b>less frequently</b> than consistent with independent business operations	9	33%