Econometric analysis of efficiency gains from on-rail competition

The Competition and Markets Authority commissioned Wheat and Smith (with Rasmussen) from Leeds University's Institute of Transport Studies to undertake research comparing the costs of open access operators with those of franchised train operators after controlling for a number of differences between the two types of operator.

Do open access train operators exhibit inherent cost benefits compared to their franchised counterparts?

Torris Rasmussen, Phill Wheat[1] and Andrew Smith

July 2015

Summary

This paper sets out to inform an emerging policy debate in Britain and Europe on whether to allow open access train operators alongside franchised (tendered) operators. Building on an established cost model (Wheat and Smith, *JTEP*, 2015), the paper utilises UK data to analyse the cost side of the debate. The open access operators studied are found to have much lower costs than expected considering the significant returns to density and scale in the railway industry. Through further analysis this is found to be due to lower input prices and a potential 'open access business model' effect. Decomposition of the cost differences show that this latter effect is non-trivial, but due to the original model having been estimated using data on operators which tended to be larger than open access, there is uncertainty as to how large this effect is. Case study analysis shows that the introduction of more open access services can lower average costs of train provision on key routes in Britain. For some routes it is even possible to lower overall cost if services are switched to open access from the incumbent operator, even though the incumbent loses scale and density benefits from becoming smaller. The paper thus concludes that there are small to no cost disadvantages of allowing open access operators to compete with franchised operators and in some cases there are cost advantages of doing so.

Contents

1. Introduction

The purpose of this research is to provide empirical evidence, looking at input prices and costs more widely of open access operators in comparison with intercity franchised train operating companies (TOCs). We investigate cost differences using an established econometric model for the UK to explore whether existing open access operators are at a cost disadvantage or advantage with respect to incumbent operators. The research also considers two scenarios where open access operators make up a greater proportion of services on some UK intercity routes. The analysis is essentially concerned with understanding which of two cost effects is dominant: cost increases through open access operators being below the minimum efficient scale / density output level; or possible cost reductions through such operators adopting a more agile business model. It therefore gives an indication of this second effect, which we term the open access operator 'business model effect', while controlling for the impact of economies of scale and density. Our paper sits alongside the interim report of a Competition and Markets

[1] Corresponding author: p.e.wheat@its.leeds.ac.uk

Authority policy project to examine the scope for increasing competition in passenger rail services in Great Britain[2]. It does not consider the wider costs or benefits of open access entry.

We demonstrate that open access operators have lower input prices (particularly relating to rolling stock) than franchised intercity operators. This gives them cost advantages, all other things equal, and this impact is quantified. However, all things are not equal as previous research has shown that train companies' costs are affected by a number of factors including economies of scale and density, the number of different kinds of service they are running (heterogeneity) and whether or not they operate stations. We therefore utilise the econometric model in Wheat and Smith (2015) to examine how the costs of open access operators change as they morph to have characteristics approaching franchised intercity TOCs. Importantly this allows us to compare on a like-for-like basis, acknowledging that there will be some error given the underlying model is exactly that, a model (yielding predictions which have error), and that the model itself is estimated only on data from franchised TOCs operating a much larger scale than current open access operators are. This exercise allows us to first compare open access operators with franchised TOCs removed of the cost impact of running stations. Differential access charges paid by OAOs and franchised TOCs do not play a part here as these are stripped out of the costs of both sets of operators in the analysis. It then allows for the remaining cost gap to be allocated to various differences between open access and franchised TOCs (density, scale, heterogeneity and prices). Finally the model allows us to predict the cost impact of a substantial increase in open access market share on some selected intercity franchises, with or without a corresponding reduction in the output of the incumbent operator.

2. Methodology

Our analysis starts with direct descriptives of the data to determine if open access operators face lower input prices than franchised TOCs. We use non-parametric tests to confirm whether there is a difference in the input prices of open access and a subset of franchised TOCs, using Wilcoxon rank sum tests. These tests are able to establish whether any systematic differences in input costs between open access and franchised intercity TOCs are sufficient to be statistically significant and therefore reliable.

The main body of the analysis utilises the parameter estimates from the econometric model in Wheat and Smith (2015) to predict the cost for open access operators in conjunction with observed characteristics in terms of the services they run. These predicted costs can then be compared with the observed costs of the open access operators. Through an auxiliary regression it is then possible to see whether, once factors within the model are controlled for, there is a systematic cost difference between open access and franchised operators. Such a difference could be interpreted as some influence of an 'open access business model' effect (MVA Consultancy and ITS, 2011).

By applying the different parameters from the Wheat and Smith (2015) model to real open access operators' characteristics stage by stage, a decomposition analysis is undertaken. This provides important insights as to the reason for the cost differences between open access operators and franchised TOCs. Importantly, one key difference between the majority of franchised train operators and open access operators is that franchised TOCs tend to operate stations while open access operators do not. The econometric model allows us to 'net off' the cost impact of running stations to facilitate a more like for like comparison. It then allows us to disentangle the remaining cost difference into that which arises from the size, density, input prices and heterogeneity differences between franchised TOCs and open access operators.

This approach is adopted since the limited data points available for open access operators make estimation of a dedicated model for open access infeasible and because it is informative to understand how current open access operators' costs compare to what those of franchised TOCs would be expected to be if they were operating the services provided by open access operators, and what the key drivers for this may be. Further research to update the econometric model of Wheat and Smith for both open access and franchised TOCs could be of value but this is outside of the scope of this research. The

[2] https://www.gov.uk/cma-cases/passenger-rail-services-competition-policy-project

approach adopted instead decomposes the observed cost difference between franchised and open access operators into that which is implied using the scale and density characteristics of (smaller) franchised operators and that which is unique to open access.

2.1. The Wheat and Smith (2015) econometric model

We now provide a brief description of the Wheat and Smith (2015) model. For more information please see Annex 2, or, for full detail, the original paper.[3] The model was estimated using data on franchised train operators in the UK and has been used to analyse the optimal size and density of franchised TOCs. It comprises a hedonic translog cost function to account for service quality and includes a rich set of variables to characterise the output of train operators. A strength of the model is that it is relatively flexible in terms of its ability to model the relationship between costs and cost drivers in a way which is not excessively constrained by the assumed mathematical form.

To take into account that train operators provide a number of outputs, and that evaluating cost against any one of these outputs can misrepresent some services, the model is able incorporate three outputs: route-km, train hours and stations (if applicable). Route-km is included separately from train hours to distinguish between returns to scale and density (see below). In addition, the train hours variable is actually a hedonic function (Spady and Friedlander, 1978), which implicitly equates quality (heterogeneity) factors to the train hours numeraire. What this means is that the train hours for a given train operator are amended to take into account the heterogeneity within its service. This includes the speed that trains run, the average length of trains and the extent to which different services are provided (measured by both the proportion of train-km which are of a given type (intercity, London commuting and regional) and the number of rolling stock types used by the train operator). The full list is given in Table 1.

A key difference of the Wheat and Smith (2015) model vis-à-vis previous studies is that it uses train hours as the primary output variable, instead of train-kms. This reflects the unique data source available to the authors (previously no train hours data was available for econometric studies of this type). Although many costs of running a train are primarily based on distance, time driven costs, such as vehicle leasing and staff costs account for the majority of total costs. Further, the Wheat and Smith (2015) model does allow for the influence on cost of distance run (through use of average speed) and a different number of vehicle hours or kms (through use of average train length).

The rich characterisation of TOC costs in the Smith and Wheat (2015) model is necessary to understand the underlying returns to scale and density properties of the industry. In particular, the Wheat and Smith analysis highlighted that while TOCs were found to generally experience from increasing returns to density, the ability of them to exploit them depends on whether the TOC expands by offering similar services or merges with another type of service operator. For example, an intercity operator merging with a London commuting operator may not be able to realise the returns to density savings as services are heterogeneous.

2.2 Approach to Decomposition

The approach taken in this paper uses an econometric decomposition to quantify the reasons for open access operator costs differing from franchised TOCs, namely arising from:

[3] Wheat, P., and Smith, A. (2015). ‘Do the usual results of railway returns to scale and density hold in the case of heterogeneity in outputs: A hedonic cost function approach’. *Journal of Transport Economics and Policy*, 49(1), 35-57.

1) Returns to Density

The Wheat and Smith (2015) model defines returns to density as how costs change in response to a change in number of train hours operated per route km. Thus it is a measure of how costs change when a train operator increases service levels on a fixed network size (in terms of origin and destinations served).

2) Returns to Scale

Returns to scale measures the variation in costs in relation to changes in geographical size. It therefore indicates whether a train company could be more efficient operating at a larger or smaller size. The model is likely to show that open access operators operate at too small a scale. Therefore the effect of both economies of density and scale are expected to result in open access operators having greater costs than franchised TOCs at their current levels.

3) The impact of heterogeneity of the TOC's service

The Wheat and Smith model explicitly considers the impact of cost from a number of factors in addition to the impact of the primary output, train hours. In particular, the model includes both average characteristics of services, namely average train length, average train speed and average passenger loading, as well as measures of the extent to which a variety of services are offered by a given operator (such as proportion of train-km operated by service type and number of rolling stock types operated). This allows for a rich description of costs. Of particular relevance here is that open access operators tend to run shorter trains (in terms of number of vehicles per train) than franchised TOCs. As discussed in the results section, having open access operators operating shorter lengths of trains compared to franchised TOCs implies that there is a discrepancy between a per vehicle hour and per train hour measure of cost. We present both.

4) Input price differences

Analysis of the raw data presented in the first part of the results section indicates that open access operators have lower input prices than intercity franchised TOCs. We use our econometric model to evaluate the impact that this has on costs. This impact will work to reduce open access costs relative to franchised operators.

5) The 'open access business model effect'

Both MVA Consultancy and ITS (2011) and the Rail Value for Money Study commissioned jointly by the DfT and ORR (2011) hypothesized that open access operators may have a more agile business model which permits them to achieve lower costs than comparable franchise operators and therefore offset density and scale disadvantages. We investigate this by attempting to find evidence that open access operators have systematically lower costs than franchised intercity TOCs once density, scale and input price effects have been controlled for.

To do this we run an auxiliary regression. First we used the Wheat and Smith (2015) model to predict the cost for all train operators (both franchised and open access) given their characteristics. Second, we regress the resulting gap on a constant and an open access dummy variable. The coefficient on the dummy variable indicates the extent to which open access operators are, systematically, cheaper (or more expensive) than their franchised equivalents, all other things equal.

Note that this is the only new element of econometric estimation in this paper relative to the Wheat and Smith (2015) paper; that is the base model (that we use for examining density, scale, input prices and predicting costs for open access) is that estimated by Wheat and Smith. It is out of the scope of this work (both in terms of resource and time) to update the underlying model. The interrogation of the model in this way to produce results to help inform a specific policy question is, however, new.

6) The impact of the number of stations

A key difference between open access operators and franchised TOCs is that the former do not operate stations, while most of the latter do. This affects the total cost of franchised TOCs, and could therefore be reflected in costs per unit output. The Wheat and Smith model explicitly considers costs to be a function of running trains and operating stations. This makes it possible to estimate what effects running stations have on total costs for train operators. Wheat and Smith (2015) did not account for potential heterogeneity in stations, but the major stations are run by Network Rail directly, with franchised TOCs operating the smaller ones along their routes (which they lease from Network Rail). The nature of the intercity TOCs considered in this paper further narrows the heterogeneity of stations in the sample.

Scenario Analysis

In addition to using the econometric model to decompose the cost difference between open access and franchised intercity TOCs, a further analysis uses the model to consider the cost implications of increasing the scale of open access operations in a non-marginal manner to determine how overall route costs would change as a result of an expansion of open access operations. It was decided for this analysis to use an increase in the size of the open access operators such that open access accounts for 15% of the train hours of the market (currently open access account for 5% of train hours on intercity services; these being defined here as East Coast, West Coast and Cross Country). Further, it is assumed that this increase is small enough such that open access operators could maintain their favourable input price levels.

Two scenarios are considered. First, that open access operators grow to take 15% of the market without any reduction in the franchised TOC's output (Scenario 1). This can be thought of as allocating growth in future train hours (e.g. from network capacity expansion) to open access. Second, that open access grow to 15% of the market by substituting a proportion of franchised services (Scenario 2).

To implement these scenarios we modelled growing the existing open access operations to 15% of the market, holding the other characteristics of the open access operator constant. In a similar manner, for scenario 2, when reducing the train hours of the franchised operator, we again hold all other characteristics of the operator constant. In reality this may slightly overestimate any savings because open access operators operate shorter trains (thus total vehicle hours in Scenario 2 fall). However, adjustments to take train length into account would need to make assumptions in themselves, so we adopt this approach to keep the analysis tractable. In reality there will be other characteristics of the operation which may vary (such as increases in route length) which would imply offsetting effects. We report results for each of the three franchised operators separately (evaluated for 2009/10) as each has different unit costs (which makes a difference to the total cost change in Scenario 2) as well as different operation sizes (which affects the extent to which open access has to grow to reflect 15% of the market). Given the current concentration of open access on the East Coast route relative to other routes, we have used the same average open access operator for all routes (computed as explained in Section 4).

Finally, it is important to highlight the elasticity of cost with respect to train hours to used for each operator to make the calculations. For the franchised operators we use the elasticity predicted in the model. For the open access operator, given our discussion in section 5 we do not use the elasticity predicted by the model, as this estimate could be picking up a mix of the open access business model effect and the true elasticity. Instead, we utilise a more conservative value of 0.73 which reflects the lowest of the three intercity TOCs considered in the analysis. This is a worst case assumption – that is, it probably overstates the cost of open access operators as they grow – since, in practice, we would expect a smaller operator to have a lower elasticity.

3. Data

The model by Wheat and Smith (2015) relates TOC costs (excluding access charges which are transfer payments) to three outputs – route-km, train hours and number of stations operated – and two input prices – payroll staff and other cost per rolling stock unit. The data is from 1999/00 to 2009/10.

In addition to the three outputs, the model contains a rich characterisation of train services, including the average speed, train length and passenger load, as well as characterisations of the diversity of services operated by a given train operator, namely the proportion of intercity, London commuting and regional services operated and the number of 'generic rolling stock' types.

To undertake the current analysis, the Wheat and Smith (2015) data has been supplemented by compatible data for open access operators between 2001/02 and 2011/12 (19 observations in total). This paper makes use of the combined dataset to estimate the open access dummy variable factor detailed in the methodology. Table 1 summarises the sources of data for analysis.

Most data types were available for open access operators. However, it was not possible to collect data on passenger loadings due to its sensitive nature. Instead, it has been assumed that all open access operators have an average of 100 passengers per train. Importantly, however, the model is not overly sensitive to this assumption. This is because the cost elasticity associated with passenger load is very small, so variations in the value of this variable do not affect costs in a substantive manner.[4]

The variables that characterise heterogeneity in the Wheat and Smith model included several variables not available to other studies (notably the use of train hours and number of generic rolling stock types which were provided by the DfT). Clearly there are other possible variables (such as age of rolling stock and type of tractive power available in their database). However, Wheat and Smith examined inclusion of these variables and determined through both statistical testing and evaluation of the economic sensibility of coefficients that the collection of variables in the model represent an appropriate specification given the data.

Table 1: Data Descriptions and source

Name	Description	Franchised data source	Open access data source
Outputs			
Route km	The km length of the rail network a TOC uses for its services. The scale measure.	National Rail Trends (Office of Rail Regulation, n.d.)	National Electronic Sectional Appendix (Network Rail, n.d.)
Train hours	Primary driver of train operating costs	National Modelling Framework Timetabling Module	Same source
Average vehicle length of train	Vehicle km/Train km	Network Rail	+ open access operators official information
Average speed	Train km/Train Hours	National Modelling Framework Timetabling Module	Same source

[4] For comparison, the three purely intercity TOCs considered in the decomposition have values of 90 to 240; however, it is not unreasonable to imagine that open access operators will be towards the lower end of this range given the shorter trains they operate.

Name	Description	Franchised data source	Open access data source
Passenger Load Factor	Passenger km/Train km	Passenger-km data from National Rail Trends. Train-km data from Network Rail.	Assumed to be 100 passengers per train service for open access operators due to confidentiality issues
Intercity TOC indicator[5]	Fraction of services that are of intercity nature	National Rail Trends and approximations based on train hours run by pre-dating TOCs[6]	All open access operators considered are judged to be of an intercity nature only
London South Eastern indicator	Fraction of services that are of a commuting nature into and around London		
Number of rolling stock types operated	Heterogeneity in generic rolling stock used	National Modelling Framework Rolling Stock Classifications	From web search (generally one type)
Stations operated	Number of stations operated by the TOC	National Rail Trends	No open access operators operate stations
Open access indicator			
Open access operator indicator	Dummy variable for open access operators	Zero for franchised TOCs	One for open access TOCs
Prices			
Non-payroll cost per unit rolling stock		TOC accounts for costs. Rolling stock number from TAS industry numbers	Train operator accounts and rolling stock numbers from open access operators' official information
Staff costs (on payroll) per number of staff		Both from TOC accounts	Same source

Adapted and updated from Wheat and Smith (2015)

For the descriptives and the decomposition analysis, a subset of train operators are evaluated. The data used is summarised in Table 2 below. Due to open access operators historically being of an exclusive intercity nature only, these should be evaluated against a subset of franchised TOCs running only intercity services. Furthermore, we have chosen not to include the early data points for open access

[5] The use of Intercity, London South Eastern and regional classifications pre-dates privatisation, but is still used in the rail industry today (see National Rail Trends, for example). However, to reflect the growing move to mixed franchises in recent years, Wheat and Smith (2015) amend this classification to be a proportion of services operating in each classification. Importantly for the analysis going forward, we compare open access operators only to those TOCs which operate wholly intercity services to avoid convoluting service differences with other cost differences.

[6] The TOCs that operate a mixture of service types tend to have been formed from the merger of previously single service type TOCs. An example is the current Greater Western franchise which was formed from three previous operators (Great Western, Great Western Link and Wessex) operating intercity, commuting and regional services.

operators, as these are affected by Hull Trains, the only operator at the time, being in a ‘ramp-up’ phase. Later entrants ramped up their service levels faster, and start-up years for open access operators are not included in the extended sample. This leaves us with 15 observations for the years ending between March 31st 2005 and March 31st 2012. We consider these adjustments to be reasonable although we are aware that it adds further uncertainty about the exact magnitude of our results.

In the decomposition analysis, the open access data are matched with the nine observations of the franchised TOCs that exclusively ran intercity services in all of the last three years of the Wheat and Smith (2015) sample, namely Cross Country, East Coast and West Coast. We restrict comparators to the last three years to try to align the open access to the franchised data as much as possible given the limits of the Wheat and Smith sample period. Further, we consider it sensible to undertake three years of comparison so as to average out volatility in the cost and output data to avoid making misleading comparisons based on one year only.

Table 2: Summary of Data Used in the Descriptives and Decomposition

Open access operators	Years ending March 31st	Franchised TOCs	Years ending March 31st
Hull Trains	2007 – 2012	East Coast	2008-2010
Grand Central	2009 – 2012	West Coast	2008-2010
Wrexham and Shropshire	2009 - 2011	Cross Country	2008-2010

4. Results 1: Input price analysis

Descriptive analysis in this section finds that there are significant differences between open access operators and franchised, exclusively intercity, TOCs in input and output prices in the dataset using non-parametric tests. As mentioned briefly above, the Wilcoxon rank-sum test is able to establish whether there are any systematic differences in input costs between open access and franchised intercity TOCs which are sufficient to be statistically significant and therefore reliable. The null hypothesis is that there is no difference, so a rejection of the null implies there is evidence for a difference.

Table 3 summarises the findings of the input price exercise. It shows that open access operators have lower labour prices than franchised exclusively intercity TOCs, significant at the 5% level. This difference is 10.3% on average, which indicates that open access operators are able to command lower labour costs in their input markets compared to franchised exclusively intercity TOCs. The data, however, points towards this cost benefit having decreased in the latter years of the sample, as open access operators have experienced increased labour prices.

For the other input price, all other costs divided per vehicle, open access operators experience on average an ‘other’ price that is 33.6% lower than that of franchised exclusively intercity TOCs. The data does not, however, indicate a clear trend on whether these prices are converging or not. The Wilcoxon rank-sum test establishes significance for this result at the 1% level. Due to the limited details in the train operators’ financial reports it is not possible to ascertain exactly what causes this other price difference, but it might include lower cost rolling stock, some of the effect of open access operators not operating stations and other cost effects not related to the size and density of the operators.

An average of the two input prices, weighted by the open access cost share, shows that open access input prices are 29% lower than intercity franchised TOCs.

Table 3: Mean input prices

Measure	Value	P-value of difference	Significance
Labour price in £ per member of staff, 2014 prices			
Franchised intercity TOCs	50 232	0.0111	**
Open access operators	45 035		
Other input price in £ per vehicle, 2014 prices			
Franchised intercity TOCs	953 668	0.006	***
Open access operators	633 350		

***,**,* denotes statistical significant from 0 at the 1%, 5% and 10% levels, respectively

5. Results 2: Model results, auxiliary regression and econometric decomposition

Model results: Returns to scale and density

The estimated returns to scale and density are summarised in Table 4. It can be seen that the model predicts that open access operators are very small both from a density and scale perspective. The mean returns to scale for open access operators (3.699) indicates that these operators face substantial increasing returns to scale. In particular, as output scale increases by 1%, costs only increase by 0.27% (1/3.699). This is compared to the mean franchised exclusively intercity TOC, which is operating under decreasing returns to scale (0.822), which indicates a 1% increase in scale implies a cost increase of 1.22% (1/0.822). Thus, these results would suggest that open access operators are too small in terms of network size.

Table 4: Returns to Scale and Density for train operators, 2005 onwards

Measure	Value
Median result for all train operators (including open access)	
- Returns to scale	0.898
- Returns to density	1.178
Median result for franchised exclusively intercity TOCs	
- Returns to scale	0.822
- Returns to density	1.059
Mean result for open access operators	
- Returns to scale	3.699
- returns to density	3.026

For quantitative results for returns to density the result is broadly the same, as the mean result for returns to density for open access operators is 3.115. This implies that as train hours (or train km) increase 1%, costs only increase 0.32% ((1/3.115)/100). For comparison, the mean franchised exclusively intercity TOC operates with returns to density at 1.058, which indicates that (broadly speaking) franchised intercity operators are close to optimal operational density from a cost perspective. Taken literally, this would suggest that as open access operators grow larger, they would be able to deliver services at substantially lower cost than they do today and that only if they grew to a density (frequency) level close to a current franchise operator, will they approach minimum efficient scale (i.e. lowest unit cost). Therefore, this could indicate that open access operators have high unit costs at their current scale.

The 'Open Access Business Model' dummy variable

Turning to the econometric model, we first investigate whether there is a systematic favourable cost effect for open access operators relative to intercity franchised operators within the Wheat and Smith model. As explained in section 2, we quantify this by first predicting the cost of open access and franchised operators using the Wheat and Smith (2015) model. We then undertake an auxiliary regression of the prediction error (actual cost less predicted cost) against a dummy variable which is one if the train operator is open access and zero otherwise. This tells us whether there is systematic under or over prediction for the cost of open access operators in the model.

The coefficient estimate on the open access dummy is large and negative at -0.984. It indicates that the average open access operator's cost is over-predicted by the model such that, on average, they are 62.6% cheaper. Given that the average prediction error for all franchised intercity TOCs in sample is zero (by construction), it indicates that, taken literally, all other things equal, open access operators are 62.6% cheaper than franchised intercity TOCs.

As the model is based on franchised TOC data only, this could be interpreted as the 'open access business model effect'. However, as mentioned above, it is also possible that there could be errors in the model when applied to small open access operators. This is because franchised TOCs are generally much larger operations than that of open access operators, which introduces the possibility that some of the cost impact captured in the open access dummy is due to the model being imprecise at such low levels of scale and density due to these levels of explanatory variables not being present in the data used in the model.

To investigate this further we first note that the open access operators are predicted to have returns to density of 3. This implies a 1% increase in output implies only a 0.33% increase in cost. This seems very extreme and would indicate that the Wheat and Smith model over-predicts the costs of very small TOCs.

Second, in Figures 1 and 2, below, we include an additional bar showing 'density adjustment based on average franchised elasticity'. This considers what the costs of open access operators (per train or vehicle – it does not matter due to indexing) would be if instead of allowing the econometric model to use the Returns to Density (RtD) of 3 we fixed this at the mean RtD for the intercity franchised TOCs (1.058). Importantly even if we adopt this cautious RtD, we still find that open access costs are systematically lower than this prediction (the difference is 34% of actual open access costs). Thus, there does still seem to be evidence for an 'open access business model effect' even after allowing for error in the model due to out of sample prediction.

Third, we have calculated that the RtD required to fully eliminate the global adjustment (averaged over the nine TOC calculations (i.e. three TOCs over three years)) is 0.58, which indicates that open access operators would have to be operating at decreasing returns to density. This seems at odds with other evidence and standard production economic theory.

Therefore, we conclude that there is evidence for an open access business model effect which more than offsets the returns to density penalty that open access operators face, even though there is some uncertainty as to exactly what the magnitudes of the positive business model effect and negative density effect are in practice. Overall, the data and model indicate that open access operators operate at slightly lower cost (crucially holding train length constant and before changing input prices).

Econometric decomposition

The econometric analysis is able to decompose the determinants of cost differences in more detail than simple descriptives of the raw data. The econometric decomposition seeks to analyse what causes the observed cost differences between open access operators and franchised exclusively intercity TOCs. This is done through progressively making the modelled open access operator morph into having the characteristics (cost driver levels) of a franchised intercity TOC. These modelled results are then compared to that of the actual mean open access and franchised intercity train operators.

We consider an average open access operator and morph it into three wholly intercity TOCs (East Coast, West Coast and Cross Country) each evaluated for three years (the last three years in our data). We then average the indexed costs at each stage to present one decomposition. Thus, the results presented below are the average of nine separate exercises (three intercity TOCs each with three years). We believe this to be more robust than undertaking one decomposition based on morphing the open access operator into a franchised TOC which is actually the average of the values of the nine observations of the franchised costs. We attempted both approaches but found the given the average franchised TOC approach yielded some odd results (which is not too surprising given that averaging the cost drivers over a number of TOCs yields a set of data which is not in our model). Particular issues arose with respect to the number of stations operated where Cross Country operates no stations, but East Coast and West Coast do. To evaluate the model at the average number of stations gave different results than evaluating the model separately at zero stations and 12 and 17 stations (the values for East Coast and West Coast respectively).

In addition, because the scale of operation of open access operators and franchised TOCs are so different, it is not useful to present results in absolute cost terms. Instead, a scale measure has to be chosen as a unit divisor. However, this is to some extent arbitrary. As such, we present two divisors, train-hours and vehicle-hours.

Table 5 presents the numerical results. Figures 1 and 2 present visualisations of the results in terms of the cost index at each stage of the decomposition (Figure 1) and the percentage change in costs from one stage to another in the decomposition (Figure 2). It is important to note that each line of the table represents an adjustment to open access costs over and above all the previous adjustments. For example, 'Adjustment to Franchise Train Density Level' includes the ''OA Business model' adjustment removed' to open access costs as well.

The remainder of the discussion in this section addresses the key comparisons of greatest policy relevance.

Table 5: Decomposition of open access operator costs to intercity franchised TOC costs[7]

Stage of Decomposition		Cost per Train Hour		Cost per Vehicle Hour	
		Growth per step	Index	Growth per step	Index
1	Open access operator actual		100		100
2	'OA business model' adjustment removed	166%	265	166%	265
3	Adjustment to franchised train density level	-57%	114	-57%	114
4	Adjustment to franchise service heterogeneity (crucially train length)	18%	134	-14%	98
5	Adjustment to franchised scale of operation	-32%	91	-35%	64
6	Adjustment to franchised input prices	43%	129	43%	91
7	Addition of franchised stations operated (and reconciliation to actual franchised costs)	24%	160	20%	109

[7] This table reports the predicted cost change from an open access operator to a franchised TOC. As such the index is relative to open access (100) and the growth percentages from one step to another refer to the incremental process of achieving an open access to franchised TOC. This is highlighted, since in parts of the text in Section 5, the percentage refers to the reverse relationship. For example, in section 5.3.1 open access operators are said to be 37.7% cheaper than franchised operators (line 7) per train hour when the index is 160 for franchised operators. 160 indicates franchised TOCs are 60% more expensive than open access operators. The discrepancy is simply to do with the percentages relating to different cost bases; both are correct.

Figure 1: Cost index resulting from progressively morphing an open access TOC into an intercity franchised TOC

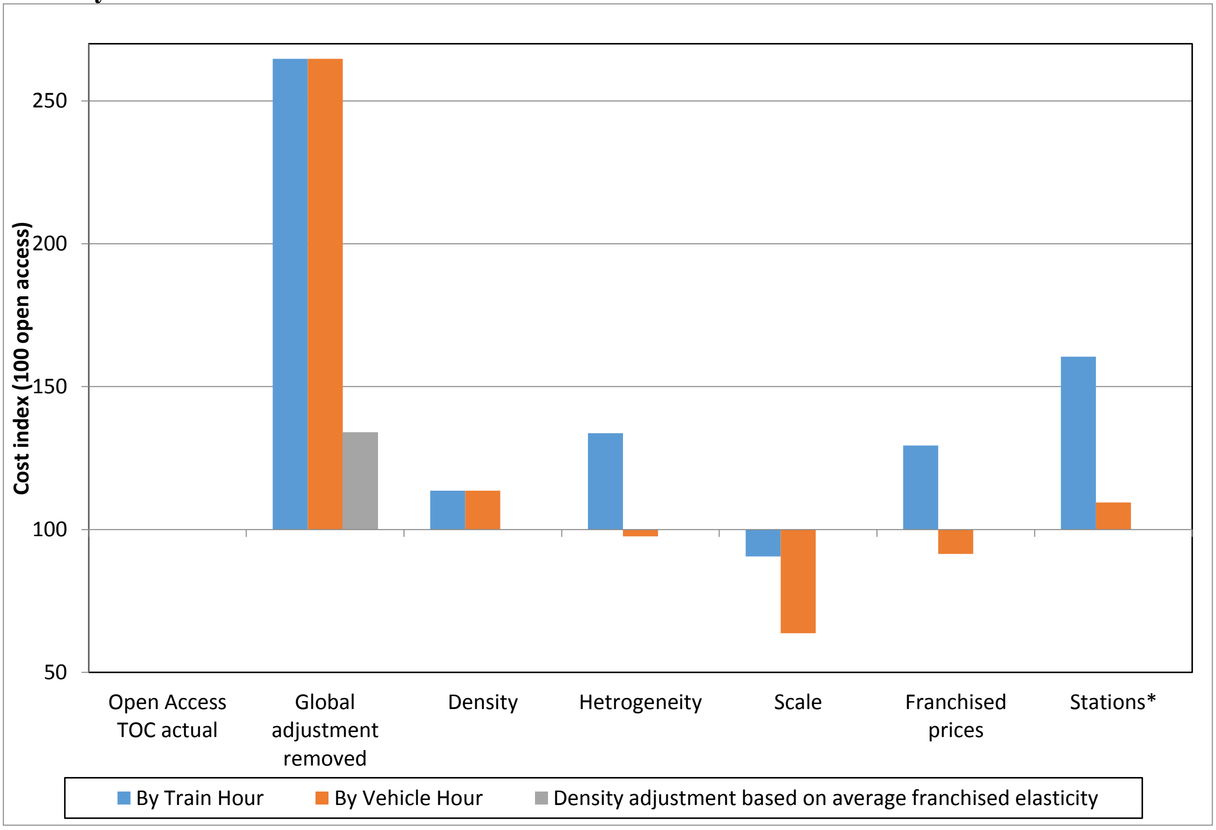
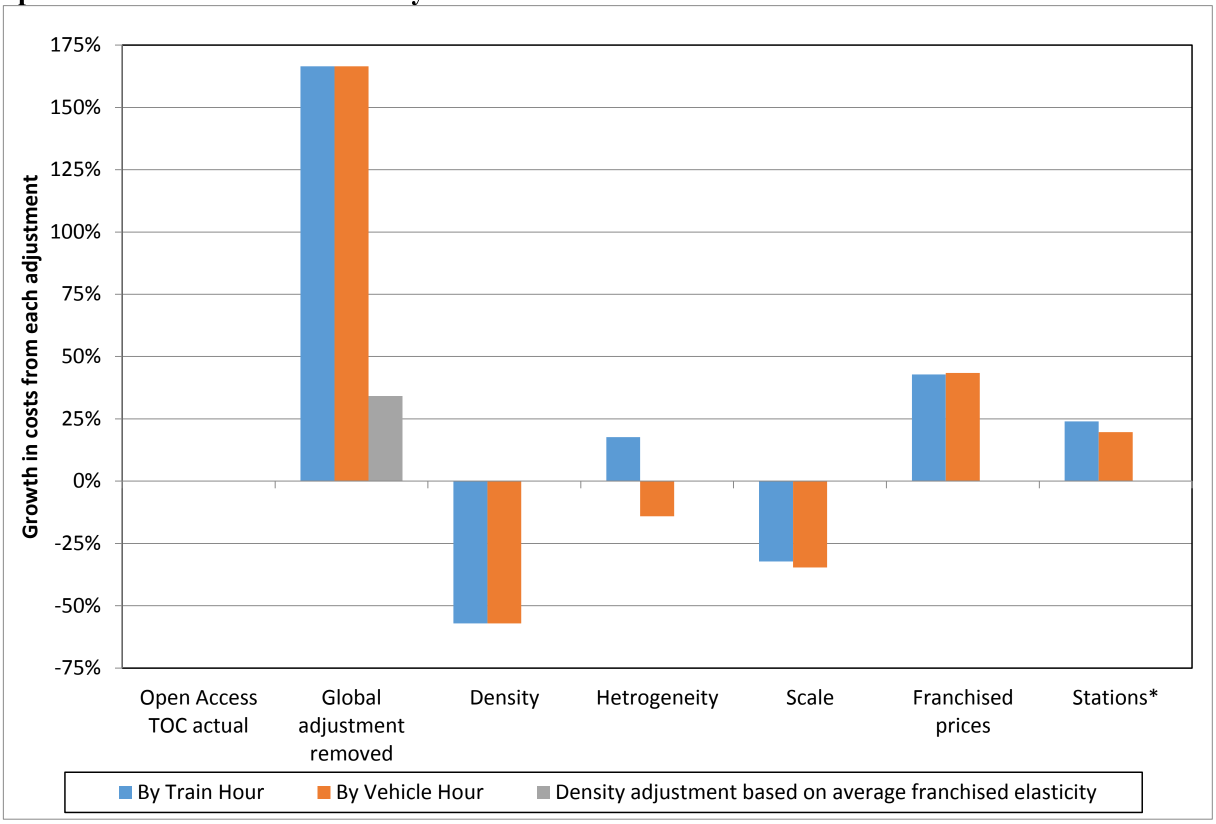


Figure 2: Percentage growth in costs at each stage resulting from progressively morphing an open access TOC into an intercity franchised TOC



Actual cost comparison

The first and last lines in Table 5 are derived directly from the cost (and output) data. They are not the predicted costs from the model. This data shows that, when open access operators are compared to the three franchised TOCs which operate only intercity services, then both per train hour or per vehicle hour, open access operators are cheaper. Per vehicle hour they are 8.6% cheaper (100/109-1)[8], while per train hour they are substantially cheaper – 37.7% (100/160-1).[9] The difference reflects that open access operators run shorter trains (less vehicles per train). Of course, operating different lengths of trains does not imply that such lengths are inappropriate to specific circumstances.

Importantly a key additional reason for the difference in costs is that franchised TOCs operate stations whilst open access operators do not and this difference is not addressed via the station access charges because of the way costs and recharges are reported. We address this problem in the next sub-section.

Identifying a better franchised comparator: Removing the impact of operating stations

The model predicts that operating stations results, on average for the three franchises, a 24% increase in operating costs on a per train hour basis and 20% on a per vehicle basis.[10] This broadly indicates that station operation accounts for 20% of operations cost for these TOCs.

Importantly, as noted above, open access operators do not operate stations. Thus a useful exercise is to compare open access costs to the predicted cost for franchised operators excluding the cost impact of operating stations.

This more like-for-like comparison indicates that, per train hour, open access operators are 22% cheaper than franchised intercity TOCs (see line 6; 100/129-1). However, per vehicle hour open access operators are 10% (see line 6; 100/91-1) more expensive than franchised intercity TOCs. Again, the discrepancy between the two measures reflects the shorter trains operated by open access operators, but now we have stripped out the distorting effects of station costs.

Explaining the remaining difference

The remaining elements of Table 5 and Figures 1 and 2, provide a decomposition of the difference between the costs of these now like-for-like operators. We provide a narrative of the remaining elements of the table:

- The second line of Table 5 simply allows the model to predict the costs of the open access operator given its output levels and input prices without any adjustment by the dummy variable discussed in 5.2. For now we maintain the interpretation that this treats the open access operator as a franchised TOC rather than allowing it to be affected by some special open access effect. This simply states that costs on both a per vehicle and train hour increase by 166% if this effect is removed. We return in section 5.4 to discuss whether such a large magnitude is plausible.

- The third line then allows for the density of operation to be increased such that the open access operator now conceptually operates a service frequency (train hours/route-km) the same as franchised operators. This reduces unit costs. Importantly, the net impact of the two factors (lines 2 and 3) is that the adjusted open access operator is 14% more expensive than the actual open access operators.

- The next adjustment (line 4) quantifies the fact that open access operators operate a service which has shorter trains and operates less variety of rolling stock. Crucially this is where the numeraire of either train hours or vehicle hours matters; running shorter trains implies costs per vehicle hour will be greater for open access operators than for franchised operators (all other

[8] The indices in the table are rounded, hence the percentage calculation here of 8.6% differs slightly from the calculation of 100/109-1.

[9] The indices in the table are rounded, hence the percentage calculation here of 37.7% differs slightly from the calculation of 100/160-1.

[10] Of course, Cross Country does not operate stations; however, it is still included as part of the data to model how much operating a station (or not) changes costs.

things equal). As such the adjustment in train length is sufficient to mean that per vehicle hour the morphed open access operator is now cheaper than the actual open access operators.

- Line 5 adjusts for different network sizes between franchised and open access operators. When the size of the network and train hours are increased to franchised levels (implying train density does not change however), the unit costs of the morphed TOC fall to reflect the exploitation of increasing returns to scale. On both a cost per vehicle hour and cost per train hour basis the morphed TOC is now 36% and 9% cheaper than the actual open access operator.

- The balancing item is to reconcile line 5 with line 6 (the 'better franchised comparator'). The only difference is that line 6 includes input prices at the level of franchised TOCs while line 5 maintains the lower open access input prices.

The two key policy implications of the above are firstly that **the cost penalty from low density operation is outweighed by the 'open access business model'**. Thus the open access model does not seem to be at a cost disadvantage from operating a less frequent service. Or, more precisely, on a like-for-like basis (line 6 in Table 5), open access operators are cheaper per train-hour (by 22%) and marginally more expensive per vehicle hour (10%) than franchised intercity TOCs.

Secondly, if open access operators can be expanded to the scale, density and service characteristics of franchised operators while maintaining their lower input prices, **then there will be unit cost savings over comparable provision of services by franchised train operators on both a per train hour and per vehicle hour basis**.

A quantification of this can be achieved by combining a (conservative) estimate of the 'business model effect' discussed above (34%) and the input price benefit of open access operators quantified in line 6 of Table 5, then open access operators are found to be 48% cheaper than franchised TOCs[11] if they did operate at the same scale and density and heterogeneity of the franchised TOCs. However, it is unlikely that all of the 'business model' and input price benefits could be retained if open access operators moved to the same density, scale, etc. of franchised TOCs, so this figure should be treated with caution.

6. Results 3: Scenario analysis – the cost implications of growing OAO market share

The above analysis considers how open access operators currently compare to franchised operators, where open access are providing a small number of services at the margin. This section considers the cost implications of increasing the scale of open access operations in a non-marginal manner as described in the methodology section (section 2).

The results are summarised in Table 6. This table shows the results for the two scenarios as explained below (see also the notes to Table 6).

Scenario 1 – This considers the cost change from expanding open access operations to 15% of the market holding the franchised operator at the current level.[12] The relevant measure is the unit cost index (per train hour) for open access. This shows that increasing open access to 15% of the market decreases unit costs for open access operators by between 22.8% (77.2/100 -1) and 27.6% (72.4/100 -1) depending on the route considered. Thus, as expected, the model shows substantial unit cost savings for open access operators from increasing their size. The unit costs of franchised operators are unchanged and thus this approach also reduces overall unit costs for traffic on the East Coast and West Coast routes where the open access unit costs are less than the franchised operators. This is quantified in the second to last column of Table 6 where overall route costs per train hour are predicted to fall 5.2% and 4.6% for the East Coast and West Coast routes, respectively, while for Cross Country they are predicted to increase

[11] Calculated as (1/(1.34x1.43))-1, where 1.34 is the cost growth resulting from the business model effect and 1.43 is the cost growth from moving from open access to franchised input prices.

[12] We note that network infrastructure upgrades and new technology (such as on-board electronic signalling) coming on-stream over the next two decades may make this a realistic scenario.

by 1.9% given that the Cross Country franchise has lower costs per train hour than the open access operator.

Scenario 2 – This scenario considers increasing the output of open access operators through substituting franchise output, keeping total train hours the same. The cost implication is less clear cut, for two reasons. Firstly, the measure of interest is how total cost (the sum of open access and franchised operators) changes as opposed to how unit cost for each operator changes. This is because total output (measured by train hours) does not change – only the allocation of that output to different operators does – and so it is important to understand whether this results in lower or greater overall costs for delivering this fixed output. As such the base costs of operation are relevant as well as considering how unit costs change for each operator as their respective sizes are changed. East Coast and West Coast have average costs greater than the open access operators, while Cross Country has lower average costs. Second, whilst open access operators make unit cost savings as they increase output, because franchised operators reduce output, they forgo some density economies and so their unit costs increase.[13]

The result is that for the East Coast and West Coast routes, total costs fall by a non-trivial 3.8% and 2.7% respectively whilst for the Cross Country route, total costs rise by 5.5%. Given that total train-hours have remained the same before and after, these figures represent unit cost (per train-hour) changes as well.

Thus the conclusion from these two scenarios is, firstly, that expanding open access without reducing the size of the franchised operator reduces unit cost of the open access operator. Secondly, however, it is less clear what happens to total cost when the market is spliced differently between open access and franchised operators, and this should be considered on a case by case basis taking into account the exact desired specification of the services.

Table 6: Cost implication of substituting open access services for franchised services up to 15% of the market

	Unit cost index (per train hour)		**Operations Cost (£m 2014 prices)**			**Change in cost for the route**	
	Open Access	**Franchised**	**Open Access Cost**	**Franchised Cost**	**Total Cost**	**Per train hour cost Scenario 1**	**Total and per train hour cost Scenario 2**
East Coast route							
Base[5]	100[1]	100[1]	17.44	553.29	570.73		
15%/85% train hrs	77.2[2]	102[3]	34.43	514.68	549.11	-5.2%[4]	-3.8%
West Coast route							
Base[5]	100	100	17.44	526.79	544.22		
15%/85% train hrs	74.2[2]	103[3]	46.42	483.00	529.42	-4.6%[4]	-2.7%
Cross Country route							
Base[5]	100	100	17.44	292.97	310.41		
15%/85% train hrs	72.4[2]	104[3]	59.69	267.64	327.33	1.9%[4]	5.5%

[13] Note that we hold route-km constant when changing the market share for the franchised TOCs.

Notes:

1. The open access and franchised operators index numbers of 100 are used to compare the changes in open access and franchised TOCs respectively over time. The absolute numbers of 100 do not mean that they start with the same unit costs.
2. For the open access operators scenario 1 and scenario 2 both imply growth to 15% of the current total market. Thus only one set of unit costs is shown for open access operators as they are the same for scenarios 1 and 2
3. The change in unit costs for franchised TOCs only applies to scenario 2, as in scenario 1 their output levels and unit costs are unchanged.
4. This is calculated by first calculating the per train hour costs for the route as a whole in the Base (including costs from both the franchised and open access operator) and then for the same size of franchise operation (as in the Base case) but larger open access operator.
5. Given that the open access operator we consider has the average of the characteristics of all the open access operators we have in the database, the total costs and per train hour costs that are calculated for the route in the Base case includes this stylised operator. In reality not all routes currently have the same level of open access provision, so our analysis is to some extent abstract. However, we maintain such an approach because it avoids having to predict costs for very small open access operations.

7. Conclusion

In this paper we have considered how open access operators' costs compare to franchised TOCs' costs. We have looked: at comparisons of input prices of current open access operators and franchised TOCs; at the results of applying the parameter estimates from the Wheat and Smith (2015) model to open access operators and intercity TOCs characteristics to compare whether open access operators appear to have lower costs than a franchised intercity TOC would be predicted to have while operating similar services; at what the model suggests are the key drivers of this, through decomposition analysis; and at scenarios to assess the implications of modelled open access operators running a greater volume of service on some intercity routes.

We have demonstrated that open access operators have lower input prices (particularly relating to rolling stock) than franchised intercity operators. This gives them cost advantages, all other things equal, and this impact has been quantified. An average of the two input prices, weighted by the open access cost share, shows that open access input prices are 29% lower than franchised intercity operators, and for each of the two input prices, the difference is statistically significant. However, all things are not equal, as we have highlighted, and it is important to take into account a number of factors when comparing train companies' costs, including the economies of scale and density at which they operate. We proceed to use the econometric model in Wheat and Smith (2015) to examine how the costs of open access operators change as they morph to have characteristics approaching franchised intercity TOCs. Importantly this allows us to compare on a like-for-like basis, acknowledging that there will be some error given the underlying model is exactly that, a model (yielding predictions which have error), and that it is based on parameters estimated on franchised TOC data. It allows us to compare open access operators with franchised TOCs removed of the cost impact of running stations. It then allows for the remaining cost gap to be allocated to various differences between open access and franchised TOCs (density, scale, heterogeneity and prices).

We find that the cost disadvantages of sub-optimal scale and density appear to be more than offset by the efficiency advantages offered by open access operators being able to adopt a different business model. There is evidence that this business model gives tangible effects both in the input markets, where open access operators are able to command significantly lower prices than franchised TOCs, and in the efficiency of the operation (quantified by the open access dummy variable). Of course, open access operators also do not run stations, which makes them appear cheaper. However, even if franchised TOCs were not to operate stations, open access operators would still be cheaper on a pure cost perspective, at least when measured per train hour (by 22%); though open access operators would be marginally more expensive per vehicle hour (by 10%). Therefore, despite operating at small scale and density, open access operators' costs are broadly in line with those of franchised TOCs when the cost effect of running stations is stripped out of the latter to enable a like-for-like comparison.

The effects of lower input prices have been predicted, with open access operators facing 43% higher costs if they were to adopt the input prices experienced by franchised TOCs.[14] Thus, open access operators benefit greatly from the ability to command lower input prices. The difficulty with this is that as they grow larger in the UK, where the input markets are limited due to the physical dimensions of the infrastructure, these benefits may well disappear. Currently open-access operators are able to obtain small amounts of rolling stock, on a marginal basis, and there does not exist a large pool of rolling stock that could be utilised for larger scale operations. Unlike the open access operators on the continent, UK operators are not able to use existing rolling stock running on continental European networks (due to differences in loading gauge and platform heights for example).

There is also evidence that there is a pure 'open access business model' effect, which relates to the efficiency of open access operators. However, the magnitude of this is uncertain, even if there is strong evidence that a positive effect exists. Further research into this is necessary to establish the exact magnitude, as currently there are few similar sized comparators to open access TOCs and too small a sample to estimate a dedicated econometric model. This limits the decomposition of an open access business model effect from the returns to density and scale properties of a model. Indeed, as more data (on both franchised and open access operators) becomes available, an update to the econometric model in Wheat and Smith (2015) would seem desirable, as such an exercise may shed more light on the exact returns to density and scale faced by small operators. However cautious sensitivity analysis suggests that the difference with and without the business model effect is in the region of 34% of actual open access costs.

One focus of our paper has been to ask whether the current limited scale of open access operators has resulted in such operators having higher costs than franchised intercity TOCs. As noted, we find that this is not the case. A second focus has been to assess what would be the implications if open access operations were expanded considerably.

To address this second question we use our model to assess two scenarios. In both cases we assume the open access operators retain their current business model and unit cost advantages. For the first scenario we find that if open access operators were to grow substantially – to a level where their volume is equivalent to 15% of the current train hours of each of the three intercity routes (East Coast, West Coast, Cross Country), leaving the franchised operators running the same level of services, then the unit costs for open access operators would fall by between 23% and 28% depending on which route is being considered. Average total costs would fall 5.2% and 4.6% for the East Coast and West Coast routes respectively, while for Cross Country they are predicted to increase by 1.9%. This could be interpreted as a scenario in which future volume growth is picked up by open access operators, whilst franchised TOCs continue to run the same volume of services as currently.

The second scenario assumes that open access operators take a 15% market share from franchised operators (leaving a 15% / 85% open access / franchised TOC split). In that case, as in the first scenario, open access unit costs would again fall by between 23% and 28%, depending on which route we are considering, as these operators capture the cost benefits of economies of density. However, our model finds that, in this second scenario, franchised TOC units costs would rise as they run fewer trains down the fixed route and thus lose economies of density. Overall though, train operating costs on the East Coast and West Coast routes would fall by between 3% and 4% according to our model; whilst on Cross Country, unit costs would rise by around 5.5% (the latter finding results from the franchised TOC on the Cross Country route having quite low unit costs at the outset and a different operation from the East and West Coast franchisees).

[14] This reconciles to the 29% cheaper weighted average of input prices given that this 43% refers to the lower open access cost base while the 29% is relative to the franchised operator (input price) base. Any remaining differences are due to rounding, different approaches to averaging and some input substitution (in the case of the 43% figure from the econometric model).

The conclusions from the scenario analysis is that expanding open access without reducing the size of the franchised operator reduces the unit costs of the open access operator. However, it is less clear what happens to total cost when the market is spliced differently between open access and franchised operators, and this should be considered on a case by case basis taking into account the exact desired specification of the services. We found, for example, that introducing a greater proportion of open access operators lowered overall costs on the East Coast and West Coast mainlines, but not on the routes operated by Cross Country.

Overall, purely from a cost perspective, the analysis in this paper suggests that there are small to no cost disadvantages from permitting further marginal open access competition, and some evidence for cost advantages, which thus weighs in favour of such a policy going forward in order to gain the wider competition benefits. The findings are more complex when considering more extensive open access competition, where open access operators are assumed to take up to 15% of the market, though again our scenario analysis suggests that there could be cost advantages for two of the three routes analysed. Therefore, a case-by-case analysis is needed to assess the precise effects on any given route. It should be noted that we have assumed that open access operators are able to maintain their business model and input price advantages even as they grow to a size equivalent to 15% of the market. If this is not the case, the full extent of the savings may not be realised.

Annex 1: Literature Review

The literature on open access passenger services has been limited by the small size of the market. Up until the late 2000s only the UK had any true open access operators of significance, and due to concerns about government finances even this market was limited.[15] However, this paper is as much a part of the stronger, but still emerging field studying the cost structure and productive performance of tendered passenger train operators, which has developed following the liberalisation of British and European railways. Therefore this literature review is split into two sections, the first reviewing the literature on open access passenger services, while the second considers the development of econometric research on the cost of operating passenger rail services.

The literature on open access operations

Reviewing the existing literature on open access operators, most papers assume no other cost (dis)benefits for open access TOCs (Nash and Preston, 1992; Preston et al, 1999; 2002; Preston 2010). These papers sought to expand the literature in the wake of rail liberalisation, both in the UK and Sweden, to inform a growing debate on the potential cost and benefits of open access operations. Due to there not being any econometric literature on train operator costs – including open access operators – the assumption of no other cost (dis)benefits has the effect of automatically making these operators more expensive (due to the general findings of economies of scale and density in the rail industry), thus less beneficial from a welfare perspective.

Nash and Preston (1992) concluded that open access services competing with a vertically integrated incumbent on the latter's track would provide either the open access operator or the incumbent a cost advantage, depending on whether the access charge was based on or above marginal infrastructure costs. Preston et al (1999) expanded the theoretical analysis to pure on-track competition in a vertically separated sector. Using a bottom up approach for costs, including access charges but not variable operator efficiency, they found that from a new entrant perspective there was significant scope for on-track competition. This would in most cases lead to overall welfare losses, mostly attributed to a considerably lower producer surplus through lower revenue and less benefits of scale on average, while the consumer surplus in most cases was modelled to increase. Preston et al (2002) expanded the application of the method further, looking at theoretical on-track competition with new entrants in both the UK and Sweden. This paper found that the form of competition that would be expected from open access operators would depend on the level of access charges. With the lower Swedish infrastructure charges, head-on competition would be more feasible and increase welfare from status quo.[16] Although it would not reach the theoretical welfare optimum through to overprovision of train services, and making currently profitable parallel routes unprofitable. In the UK, with more expensive access charges, fringe competition would increase passenger welfare, but lead to over provision of services at peak times and increase in the subsidy requirement of the railways (due to cream skimming). Neither of these papers sought to establish whether open access operators had other cost (dis)benefits over and above that of franchised TOCs, or integrated operators in the case of Nash and Preston (1992).

A different strand of research, commissioned by the Office of Rail Regulation in the UK, is based on real world cost data for open access operators. In studying the impacts of concrete service proposals from existing open access operators, MVA Consultancy (2009) used actual cost data from the operators themselves and construct unit cost comparisons. These techniques are limited in that they assume constant returns to scale (or density). MVA Consultancy and ITS (2011) considers both the effects of economies of scale and density in a top-down cost model, where costs were pivoted off that of franchised

[15] For most of the first decade of open access operations, the UK had only one operator running regular open access passenger services – Hull Trains. There are now two open access operators, while a third one was set up but is now defunct. This is not to say that there has not been interest from the industry to start further services, but it is only recently the infrastructure manager and the regulator has come around to the idea (Preston, 2008; Office of Rail Regulation, 2009; Alliance Rail Holdings Ltd., 2014).

[16] Interestingly Sweden became a real world laboratory for this type of competition when MTR Express began servicing the Stockholm – Gothenburg route in March 2015. Railway Gazette, March 23 2015: MTR launches open access inter-city service

TOCs, as well as the possibility of open access operators having other cost benefits that could make the rail industry more efficient. It is expected that the loss of economies of density may be outweighed by other gains, such as lower staff costs. Although no econometric analysis including the open access operators was undertaken in this work, the report finds that the "[literature] suggested that [open access operators] could achieve efficiencies not yet realised by FOs [franchised operators], and could possibly drive out the kind of savings referred to in the McNulty review" [17] (MVA Consultancy and ITS, 2011, p. 86). A drawback of MVA Consultancy and ITS (2011) is that it based its density estimates on the sample mean from Smith and Wheat (2012), only varying it with the number of vehicles an open access operator needed to run different sets of services. In the report's modelling, all other cost faced by open access operators were assumed to experience the same constant returns to density as the modelled franchised TOCs. Nor was there any attempt at estimating economies of scale at the output levels of open access operators: an assumption that they were broadly constant was based on Smith and Wheat (2012). The present paper seeks to investigate the relationship between the economies of density and scale of open access operators and those of franchised intercity TOCs. This will provide better evidence to be used in studies like MVA Consultancy and ITS (2011).

The econometric literature on passenger rail cost structure

There is an emerging literature of econometric cost analysis of train operations, following the vertical separation of railways in the UK and Europe, to which this paper is a new addition Papers in this area are a branch of the considerable work that has been put into developing measures for productivity and cost structure of railways around the world. Most of this has for historical reasons been focused on vertically integrated railways. Oum et al (1999) and Smith (2006) provide overviews.

Econometric literature specifically relating to passenger train operations is less developed. However, several papers based on British data have been published. Notable studies include Affuso et al (2002), Cowie (2002a; 2002b; 2009), Smith and Wheat (2007; 2012) and most recently by Wheat and Smith (2015). To date, work has not explicitly included open access operators, due to limited data availability.

Although none of these studies take into account open access operators, their findings on returns to scale and density are of importance to this paper. Studies that do not consider economies of density have tended to find increasing returns to scale for British TOCs (Cowie, 2002a). Smith and Wheat (2007) found near constant returns to scale, and increasing returns to density. Using a restricted trans-logarithmic function, Smith and Wheat (2012, p. 39) find 'broadly constant returns to scale and increasing returns to density'. By distinguishing between intercity, commuter and regional operations, the hedonic approach adopted by Wheat and Smith (2015) finds that most UK TOCs are above efficient scale, with a sample mean that indicates decreasing returns to scale. The smaller regional and commuter TOCs do experience some positive returns to scale. All operators are, however, experiencing increasing returns to density. Affuso et al (2002; 2009) and Cowie (2009) all impose constant returns to scale on their models, and do not consider returns to density.

This indicates that smaller operators experience strong returns to scale and density. Thus it could be expected that open access operators (being small) operated well below minimum efficient scale and hence could be expected to have higher unit costs. However there may be other factors, such as lower input prices or a more 'agile business model', which may offset this. Our analysis aims to quantify such a trade-off explicitly.

[17] Rail Value for Money Study, 2011.

Annex 2: The Wheat and Smith (2015) econometric model

The Wheat and Smith model was estimated using data on franchised train operators in the UK and has been used to analyse the optimal size and density of franchised TOCs. It comprises a hedonic translog cost function to account for service quality and includes a rich set of variables to characterise the output of train operators. The model is relatively flexible in terms of its ability to model the relationship between costs and cost drivers in a way which is not excessively constrained by the assumed mathematical form. [18]

The full model formulation and explanation is detailed in Wheat and Smith (2015). It can be represented as a system of equations (cost function and input cost shares) and, to accommodate some TOCs having zero stations to manage[19], we have two pairs of equations:

$$\left\{\begin{array}{l} \ln\left(\frac{C_{it}}{P_{2it}}\right)=\left\{\begin{array}{l}\alpha+\sum_{l=1}^{3}\beta_l\ln(\psi_{lit})+\delta_1\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\gamma_T t+\frac{1}{2}\sum_{l=1}^{3}\sum_{b=1}^{3}\beta_{lb}(\ln(\psi_{lit}))(\ln(\psi_{bit}))\\ +\delta_{11}\left(\ln\left(\frac{P_{1it}}{P_{2it}}\right)\right)^2+\sum_{l=1}^{3}\kappa_{l1}(\ln(\psi_{lit}))\left(\ln\left(\frac{P_{1it}}{P_{2it}}\right)\right)\\ +\sum_{l=1}^{3}\lambda_{Tl}t\ln(\psi_{lit})+\varphi_{T1}t\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\gamma_{TT}t^2\end{array}\right.\\ S_1=\delta_1+2.\delta_{11}\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\sum_{l=1}^{3}\kappa_{lm}\ln(\psi_{lit})+\varphi_{Tm}t\end{array}\right\}$$

for those TOCs that operate stations and:

$$\left\{\begin{array}{l} \ln\left(\frac{C_{it}}{P_{2it}}\right)=\left\{\begin{array}{l}\alpha+\beta'_1\ln(\psi_{1it})+\beta_2\ln(\psi_{2it})+\delta_1\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\gamma_T t\\ +\frac{1}{2}\sum_{l=1}^{2}\sum_{b=1}^{2}\beta'_{lb}(\ln(\psi_{lit}))(\ln(\psi_{bit}))+\delta_{11}\left(\ln\left(\frac{P_{1it}}{P_{2it}}\right)\right)^2\\ +\kappa'_{11}(\ln(\psi_{1it}))\left(\ln\left(\frac{P_{1it}}{P_{2it}}\right)\right)+\kappa_{21}(\ln(\psi_{2it}))\left(\ln\left(\frac{P_{1it}}{P_{2it}}\right)\right)\\ +\lambda'_{T1}t\ln(\psi_{1it})+\lambda_{T2}t\ln(\psi_{2it})+\varphi_{T1}t\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\gamma_{TT}t^2\end{array}\right.\\ S_1=\delta_1+2.\delta_{11}\ln\left(\frac{P_{1it}}{P_{2it}}\right)+\kappa'_{11}(\ln(\psi_{1it}))+\kappa_{21}(\ln(\psi_{2it}))+\varphi_{Tm}t\end{array}\right\}$$

for those that do not operate stations. The two sets of equations are linked by common parameters, however each set also has unique parameters on the scale outputs to facilitate sufficient flexibility to recognise that whether the TOC operates stations impacts of the cost elasticities of the remaining outputs. The definitions of the variables are given in Table 1 below.

[18] The Translog function is a second order Taylor series expansion to an arbitrary cost function in logarithms. See Wheat and Smith (2015) for details on the implicit restrictions of the hedonic function.

[19] Station costs refer to the cost of a TOC operating a station as opposed to paying for access to stations. The latter is paid for via the access charge system which the cost measure in the Wheat and Smith (2015) excludes. Further we believe that the revenue a TOC receives from other TOCs using the stations that it operates enters the TOCs statutory accounts as revenue rather than as a negative cost. As such the cost measure in Wheat and Smith should only include the cost of running stations and train services. However given the detail available in individual accounts there may be some variation in reporting amongst TOCs.

The ψ variables represent the three outputs: route-km, train hours and stations (if applicable). Route-km is included separate from train hours to distinguish between returns to scale and density (see below). In addition, the train hours variable (ψ_2) is actually a hedonic function, which implicitly equates quality (heterogeneity) factors to the train hours numeraire. What this means is that the train hours for a given TOC is amended to take into account the heterogeneity within its service. This includes the speed that trains run, the average length of trains and the extent to which different services are provided (measured by both the proportion of train km which are of a given type (intercity, London commuting and regional) and the number of rolling stock types used by the TOC). The full list is given in Table 1.

A key difference of the Wheat and Smith (2015) vis-à-vis previous studies is that it uses train hours as the primary output variable, instead of train kms. This reflects the unique data source available to the authors (previously no train hours data was available for these type of econometric studies). Although many costs of running a train are primarily based on distance, time driven costs, such as vehicle leasing and staff costs are the majority of total costs. Further, the Wheat and Smith (2015) model does allow for the influence on cost of distance run (through use of average speed) and different number of vehicle hours or kms (through use of average train length). As detailed in Wheat and Smith (2015), while the hedonic functional form does (necessarily) restrict the relation between the cost of one train hour versus one train km, the relation of train km with total cost is still modelled as a translog (second order) relationship. Therefore, while the choice of train hours over other measures is justifiable as the primary output, the Wheat and Smith specification is still flexible enough such that the impact of differing speed and train lengths on costs. Thus cost elasticities can be derived with respect to distance (train-km) as well as train-hours and these are given in Wheat and Smith (2015).

We also include in this section a more fully defined matching of variables from the Wheat and Smith (2015) model and corresponding data sources. This is similar in essence to table 1 in the body of the report above, but may be more informative for a technical audience.

Symbol	Name	Description	Franchised data source	Open access data source
Generic Outputs (ψ)				
$\psi_1 = y_1$				
y_1	Route km	The km length of the rail network a TOC uses for its services. The scale measure.	National Rail Trends (Office of Rail Regulation, n.d.)	National Electronic Sectional Appendix (Network Rail, n.d.)
$\psi_2 = y_2 q_{12}^{\phi_{12}} q_{22}^{\phi_{22}} q_{32}^{\phi_{32}} e^{\phi_{42}q_{42}} e^{\phi_{52}q_{52}} e^{\phi_{62}q_{62}} e^{\phi_{72}q_{72}} e^{\phi_{82}q_{82}} e^{\phi_{92}q_{92}}$				
y_2	Train hours	Primary driver of train operating costs	National Modelling Framework Timetabling Module	Same source
q_{12}	Average vehicle length of train	Vehicle km/Train km	Network Rail	+ open access operators' official information
q_{22}	Average speed	Train km/Train Hours	National Modelling Framework Timetabling Module	Same source
q_{32}	Passenger Load Factor	Passenger km/Train km	Passenger-km data from National Rail Trends. Train-km data from Network Rail.	Assumed to be 100 passengers per train service for open access operators due to confidentiality issues

Symbol	Name	Description	Franchised data source	Open access data source
q_{42}	Intercity TOC indicator[20]	Fraction of services that are of intercity nature	National Rail Trends and approximations based on train hours run by pre-dating TOCs[21]	All open access operators considered are judged to be of an intercity nature only
q_{52}	London South Eastern indicator	Fraction of services that are of a commuting nature into and around London		
q_{62}	$q_{42}q_{52}$	Interaction of Intercity and LSE variables		
q_{72}	$q_{42}(1-q_{42}-q_{52})$	Interaction of Intercity and Regional variables		
q_{82}	$q_{52}(1-q_{42}-q_{52})$	Interaction of LSE and Regional variables		
q_{92}	Number of rolling stock types operated	Heterogeneity in generic rolling stock used	National Modelling Framework Rolling Stock Classifications	From web search (generally one type)
$\psi_3 = y_3$				
y_3	Stations operated	Number of stations operated by the TOC	National Rail Trends	No open access operators operate stations
Open access indicator				
q_{99}	Open access TOC indicator	Dummy variable for open access operators	Zero for franchised TOCs	One for open access operators
Prices				
P_1	Non-payroll cost per unit rolling stock		TOC accounts for costs. Rolling stock number from TAS industry numbers	Open access operator accounts and rolling stock numbers from open access operators' official information
P_2	Staff costs (on payroll) per number of staff		Both from TOC accounts	Same source

Adapted and updated from Wheat and Smith (2015)

[20] The use of Intercity, LSE and regional classifications pre-dates privatisation, however is still used in the rail industry today (see National Rail Trends for example). However to reflect the growing move to mixed franchises in recent years, Wheat and Smith (2015) amend this classification to be a proportion of service operating in each classification. Importantly for the analysis going forward we compare open access operators only to those franchised TOCs which operate wholly intercity services to avoid convoluting service differences with other cost differences.

[21] The TOCs that operate a mixture of service types tend to have been formed from the merger of previously single service type TOCs. An example is the current Greater Western franchise which was formed from three previous operators (Great Western, Great Western Link and Wessex) operating intercity, commuting and regional services.

References

Affuso, L., Angeriz, A. & Politt, M., 2002. *Number 48 - Measuring the Efficiency of Britain's Privatised Train Operating Companies.* London: Regulation Initiative Discussion Paper Series.

Alexandersson, G., 2009. Rail Privatization and Competitive Tendering in Europe. Built Environment, 35(1), pp. 43-58.

Cowie, J., 2002a. Subsidy and Productivity in the Privatised British Passensgr Railway. *Economic Issues,* 7(1), pp. 25-37.

Cowie, J., 2002b. The Production Economics of a Vertically Separated Railway – The case of the British Train Operating Companies. *Trasporti Europei*, Aug, pp. 96-103.

Cowie, J., 2009. The British Passenger Rail Privatisation - Conclusions on Subsidy and Efficiency from the First Round of Franchises. *Journal of Transport Economics and Policy,* 43(1), pp. 85-104.

European Commission, 2013. The Fourth Railway Package - Completing the Single European Railway Area to Foster European Competitiveness and Growth, Brussels: European Commission.

MVA Consultancy, 2009. *Making Better Decisions - Assessment of Alternative Access Applications on the East Coast Mainline,* London: The Office of Rail Regulation.

MVA Consultancy and The Institute for Transport Studies (ITS), University of Leeds, 2011. Modelling the Impacts of Increased On-Rail Competition Through Open Access Operation, London: The Office of Rail Regulation.

Nash, C. A. & Preston, P. M., 1992. Working Paper 354 - Barriers to Entry in the Railway Industry. Leeds: Institute for Transport Studies.

Network Rail, n.d. National Electronic Sectional Appendix. [Online] Available at: http://www.networkrail.co.uk/aspx/10563.aspx [Accessed 23 July 2014].

Preston, J., 2010. Competition for Long-Distance Passenger Rail Services: The Emerging Evidence. In: OECD, ed. The Future for Interurban Passenger Transport: Bringing Citizens Closer Together. Paris: OECD Publishing, pp. 311-332.

Preston, J., Holvad, T. & Raje, F., 2002. Track Access Charges and Rail Competition: a Comparative Analysis of Britain and Sweden, European Transport Conference. Cambridge, s.n.

Preston, J., Whelan, G. & Wardman, M., 1999. An Analysis of the Potential for On-Track Competition in the British Passenger Rail Industry. Journal of Transport Economics and Policy, 33(1), pp. 77-94.

Rail Value for Money Study, 2011. Realising the Potential of GB Rail - Final Independent Report, London: Department for Transport.

Smith, A. & Wheat, P., 2007. *A quantitative study of train operating companies cost and efficiency trends 1996 to 2006: lessons for future franchising policy.* Leiden, The Netherlands, The European Transport Conference.

Smith, A. S. J. & Wheat, P., 2012. Evaluating Alternative Policy Responses to Franchise Failure. *Journal of Transport Economics and Policy,* 46(1), pp. 25-49.

Spady, R. H. and Friedlaender, A. F. (1978). 'Hedonic cost functions for the regulated trucking industry', *The Bell Journal of Economics*, 9 (1), 159–179.

Wheat, P. & Smith, A., 2015. Do the usual results of railway returns to scale and density hold in the case of heterogeneity in outputs: A hedonic cost function approach. *Journal of Transport Economics and Policy*, 49(1), pp. 35-57.