Article:

Wheat, PE orcid.org/0000-0003-0659-5052 and Smith, A orcid.org/0000-0003-3668-5593 (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. 35-47. ISSN 0022-5258

Do the usual results of railway returns to scale and density hold in the case of heterogeneity in outputs: A hedonic cost function approach

Phill Wheat and Andrew Smith

2014

Abstract

The contribution of this paper is that it highlights the importance of modelling the interaction between returns to scale / density and heterogeneity of services when evaluating the optimal size and structure of passenger rail operations. The implication is that previous estimates of scale and density properties in railways internationally (for both separated and vertically integrated systems) may have been biased – because they did not take heterogeneity of services into account. To overcome this problem we propose and estimate a hedonic cost function which allows us to incorporate measures of train operator heterogeneity, which are central to evaluating the cost effect of merging heterogeneous train operators, and thus informing policy on what is optimal. We illustrate our model via three rail franchise mergers / re-mappings in Britain, and show that the wrong policy conclusion result could be obtained by only considering the scale and density properties, in isolation from heterogeneity.

1.0 Introduction

The conventional result in transportation and in particular rail economics is that increasing the density of utilisation of infrastructure will lower average costs (per train-km) (Hensher and Brewer (2000), Button (2010)). This may certainly be expected when we consider the costs associated with rail infrastructure (e.g. Wheat and Smith (2008), Smith and Wheat (2012a) and Andersson et al (2012)). However scale and/or density effects are also likely to be apparent in situations where industries are structured on an operation only basis, as in the case where passenger rail services are subject to competitive tendering, for example in Europe. For example, Smith and Wheat (2012b) find constant returns to scale (RtS) and increasing returns to density (RtD) with respect to train operation costs only (excluding the cost of infrastructure).

Successive reforms within Europe have seen infrastructure separated from operations to a greater or lesser degree and, though not required yet by legislation, many countries (in particular, Britain, Sweden and Germany) have introduced competitive tendering or franchising of passenger rail services. The further reforms announced in 2012 within the fourth railway package include compulsory tendering of public service contracts (European Commission, 2013). Competitive tendering in rail has also been used outside Europe, for example in Melbourne Australia, Latin America and for some North American commuter services.

Understanding the optimal cost structure of train operations, within separated railway systems, is therefore an important input into policy formulation in railways around the world with respect to determining the optimal size and structure of rail franchises. In the British context, which is the focus of the empirical analysis in this paper, a current policy question is whether to remap existing train operating companies (TOCs) into fewer, larger TOCs.

In this paper we make a new and important contribution to the previous literature as follows. We argue, and show via an empirical example, that appealing to results from previous studies regarding the extent of RtS and RtD in passenger railways could give misleading information regarding the optimal size and structure of passenger rail franchises. This is because the methodology used in previous studies does not adequately consider whether heterogeneity in services provided by train operators affects the estimates of RtS and RtD. In other words, conditional on finding RtS and RtD, there is a question over whether these can still be exploited if the services provided by merging franchises are very different. Thus previous estimates of scale and density properties in railways internationally (for both separated and vertically integrated systems) may have been biased, to the extent that they did not adequately model the interaction between scale /density and heterogeneity of services.

Our proposed methodology, which addresses the above problem, is to adopt a hedonic cost function approach which allows us to incorporate measures of TOC heterogeneity which are central to evaluate the cost effect of merging heterogeneous TOCs, and thus inform policy with regard to what is optimal from a cost perspective.

The structure of this paper is as follows. Following this introduction, Section 2 reviews the literature on the evidence of RtS and RtD in railway operations. Section 3 outlines the methodology. Section 4 outlines the data and the improvements in data relative to previous studies. Section 5 discusses the empirical findings relating to overall scale and density returns and the impacts of influence on costs of heterogeneity in outputs. It also presents, for illustration, predicted cost changes for three re-mappings and discusses the reasons for the each cost change. Section 6 concludes.

2.0 Literature review

There is an extensive literature analysing the cost structure and productivity performance of vertically integrated railways around the world (Oum et al., 1999; Smith, 2006). However, there has been relatively little work looking at the cost structure of passenger train operations sector. To our knowledge all except one are focused on Britain (Affuso, Angeriz, and Pollitt

(2003); Cowie (2002a, 2002b, 2005 and 2009); Smith and Wheat (2012b) and Smith, Nash, and Wheat (2009). Merkert et al (2009) studied railway operations in Britain, Sweden and Germany. Preston (2008) provides a review of, inter alia, previous cost studies of the British rail sector.

An important issue is whether to include an infrastructure input in any analysis of train operating costs. Clearly the infrastructure input may be an important part of the transformation function and so should be considered for inclusion in any analysis. The four papers by Cowie all include some measure of infrastructure input in the analysis (route length or access charge payments). This in turn raises two important and related problems. First, the infrastructure input is hard to measure. Route length is hardly adequate to capture the quality and extent of investment in the infrastructure. On the other hand, access charge payments are essentially transfer payments from Government to the infrastructure manager and are not reflective of the cost of network access for a given TOC (at least in a given year); see also Smith and Wheat (2012b). Second, the inclusion of this input turns the analysis into an assessment of rail industry costs/production, rather than being targeted on the TOCs.

For the above reasons, Smith and Wheat (2012b) argue that, given the measurement problems noted above, infrastructure inputs are best left out of the analysis. The dependent variable in their paper is thus defined as TOC costs, excluding fixed access charges. We follow this approach here (see section 4). Route-km is also included as an explanatory variable in their model, not as a measure of the infrastructure input, but to distinguish between scale and density effects.

Given the focus of our paper it is important to define returns to scale and density in the context of a separated, passenger train operation only service[1]. It should be noted that these two

[1] See Caves et. al. (1981) and Caves et. al. (1984) for use of the terms returns to scale (RtS) and returns to density (RtD) in empirical applications, including vertically integrated railways.

definitions refer to the effect on train operations costs only and not anything to do with infrastructure costs. We distinguish between RtS and RtD since there are two conceptual ways for a train operator to grow. Firstly, a train operator can become geographically larger i.e. operating to and from more points. This is captured by the RtS concept. Secondly, a train operator can grow by running more train hours over a fixed network. This is captured by the RtD concept (see also Cowie (2002b) and Smith and Wheat (2012b)).

The previous findings with regard to scale and density in train operations are as follows. Using a variable returns to scale DEA model, Merkert et al (2009) found that British and Swedish TOCs were below minimum efficient scale, while the large German operators were above. Using parametric methods, Cowie (2002b) finds evidence for increasing RtS and these are increasing with scale, though there is no attempt to differentiate between scale and density returns in the analysis.

Again using parametric methods, Smith and Wheat (2012b) found constant RtS and increasing RtD. One limitation of the Smith and Wheat (2012b) work was the inability to estimate a plausible Translog function. Instead, a restricted variant was estimated selected on the basis of general to specific testing and on whether key elasticities were of the expected sign. This implicitly restricts the variation in RtS and RtD. We remedy this limitation by estimating a Translog simultaneously with the cost share equations and adopt a hedonic representation of the train operations output in order to include characteristics of output in a parsimonious manner. As noted in section 3 we also augment the output specification to get a much better representing of the technology compared to previous study.

3.0 Methodology

A cost function derived from the behavioural assumption of cost minimisation is represented as

$$C_{it} = C(\mathbf{y}_{it}, \mathbf{p}_{it}; \boldsymbol{\beta}) \quad i=1,...,N \quad t=1,...,T \qquad (3.1)$$

where C_{it} is the cost of firm i in year t, $\mathbf{y}_{\mathrm{it}}$ and $\mathbf{p}_{\mathrm{it}}$ are L and M dimension vectors of outputs and prices respectively again for firm i in year t. Firms provide a great deal of different train service outputs, for example TOCs provide train services with different stopping patterns and running speeds. Thus we could consider this an issue of economies of scope. However, we cannot specify the amount of each numerous output for a number of reasons. Firstly, the data does not exist on outputs at such a level of disaggregation. Secondly, if data did exist then the model would have vast numbers of parameters such that partial analysis would be imprecise. Thirdly the Translog cost function cannot accommodate zero levels of outputs very satisfactorily. Instead we adopt the hedonic cost function approach first used by Spady and Friedlaender (1978) which provides a parsimonious method of incorporating output characteristics (termed output quality in their paper) to characterise heterogeneity in outputs. This provides a means of incorporating measures of heterogeneity of output both across and within firms. The former is important for consideration of the cost effect of merging TOCs. As discussed in Jara Diaz (1982), failure to account for output characteristics can result in incorrect policy recommends in relation to optimal firm size.

Using the notation of Spady and Friedlaender (1978), replace the lth element of $\mathbf{y}_{\mathrm{it}}$, y_{lit}, with ψ_{lit} where

$$\psi_{lit}\left(y_{lit}, \mathbf{q}_{\mathbf{lit}}\right) = y_{lit} \cdot \phi\left(q_{1lit}, \ldots, q_{Blit}\right) \tag{3.2}$$

Where y_{lit} is now the lth "physical output" and q_{blit} is the bth quality characteristic of the lth physical output. ψ_{lit} is assumed homogenous of degree one in the physical output. This implies that a doubling of y_{lit} results in a doubling of ψ_{lit}; this is required for identification of the function within the wider cost function and sets y_{lit} to be the numeriere of ψ_{lit}. We consider

$\phi_l \ \forall l$ to be Cobb Douglas as in Bitzan and Wilson (2008) (as opposed to Translog as in Spady and Friedlaender's formulation) given the large number of quality variables in our formulation.

Spady and Friedlaender (1978) discuss the implicit restrictions associated with adopting the hedonic formulation. They term the function "quality separable" since the impact of the quality variables on the associated primary output is independent of prices (and also of the level of other primary outputs). Ultimately this restriction is the price of adopting the hedonic function, however it makes the model far more manageable in terms of parameters to be estimated (we estimate 34 parameters for the hedonic formulation, but the unrestricted Translog would require estimation of circa 140 parameters; there are only 243 observations) Given the Cobb Douglas form for ϕ_l in (3.2), an eloquent way to describe the implication of the "quality separable" restriction is that the elasticity of cost with respect to the quality variable is proportional to the elasticity of cost with respect to the primary output.

We estimate a Translog cost function in $\boldsymbol{\psi}_{\mathbf{it}}$, $\mathbf{p}_{\text{it}}$ and, given that our model utilises panel data, a non-neutral technology trend

$$\ln(C_{it}) = \begin{cases} \alpha + \sum_{l=1}^{L} \beta_l \ln(\psi_{lit}) + \sum_{m=1}^{M} \delta_m \ln(P_{mit}) + \gamma_T t + \frac{1}{2}\sum_{l=1}^{L}\sum_{b=1}^{L} \beta_{lb} (\ln(\psi_{lit}))(\ln(\psi_{bit})) \\ + \frac{1}{2}\sum_{m=1}^{M}\sum_{c=1}^{M} \delta_{mc} (\ln(P_{mit}))(\ln(P_{cit})) + \sum_{l=1}^{L}\sum_{m=1}^{M} \kappa_{lm} (\ln(\psi_{lit}))(\ln(P_{mit})) \\ + \sum_{l=1}^{L} \lambda_{Tl} t \ln(\psi_{lit}) + \sum_{m=1}^{M} \varphi_{Tm} t \ln(P_{mit}) + \gamma_{TT} t^2 \end{cases} \quad (3.3)$$

Shephards Lemma is applied to (3.3) to yield the cost share equations:

$$\frac{\partial \ln(C_{it})}{\partial \ln(P_{mit})} = S_m = \delta_m + 2\delta_{mm} \ln(P_{mit}) + \sum_{l=1}^{L} \kappa_{lm} \ln(\psi_{lit}) + \varphi_{Tm} t \quad \text{m=1,...,M} \quad (3.4)$$

We estimate the model parameters as a system of the cost function and the factor shares to aid both the precision of estimates and also to ensure that the estimated cost shares are as close as possible to the true cost shares (which by (3.4) is a requirement of economic theory). In addition to the cost shares, economic theory associated with the existence of a dual cost function provides a set of useful restrictions to aid estimation. Firstly, symmetry of input demand with respect to price requires $\delta_{mc} = \delta_{cm}$ and also there is symmetry in the cross derivatives of outputs, $\beta_{lb} = \beta_{bl}$. Secondly, the cost function must be linear homogenous of degree 1 in prices. This requires:

$$\left\{ \begin{array}{lll} \sum_{m=1}^{M} \delta_m = 1 & \sum_{c=1}^{M} \delta_{mc} = 0 & m = 1,\ldots,M \\ \sum_{m=1}^{M} \kappa_{lm} = 0 \quad l = 1,\ldots,L & \sum_{m=1}^{M} \varphi_{Tm} = 0 & \end{array} \right\} \tag{3.5}$$

A convenient way of imposing (3.5) on (3.3) and (3.4) is to divide input prices and cost by one of the input prices.

Given there are parameters implicit in ψ_{lit}, estimation is undertaken using non-linear Seeming Unrelated Regression. To avoid the errors in the cost shares summing to zero for each observation, one of the cost shares has to be dropped. We drop the cost share for the Mth input (i.e. the input whose price is used to divide cost and all other prices by).

Therefore, after imposing symmetry and linear homogeneity of degree one in input prices on (3.3) and (3.4), the system of M equations to be estimated is:

$$\left\{ \begin{array}{l} \ln\left(\frac{C_{it}}{P_{Mit}}\right) = \left\{ \begin{array}{l} \alpha + \sum_{l=1}^{L} \beta_l \ln(\psi_{lit}) + \sum_{m=1}^{M-1} \delta_m \ln\left(\frac{P_{mit}}{P_{Mit}}\right) + \gamma_T t + \frac{1}{2}\sum_{l=1}^{L}\sum_{b=1}^{L} \beta_{lb}(\ln(\psi_{lit}))(\ln(\psi_{bit})) \\ + \frac{1}{2}\sum_{m=1}^{M-1}\sum_{c=1}^{M-1} \delta_{mc}\left(\ln\left(\frac{P_{mit}}{P_{Mit}}\right)\right)\left(\ln\left(\frac{P_{cit}}{P_{Mit}}\right)\right) + \sum_{l=1}^{L}\sum_{m=1}^{M-1} \kappa_{lm}(\ln(\psi_{lit}))\left(\ln\left(\frac{P_{mit}}{P_{Mit}}\right)\right) \\ + \sum_{l=1}^{L} \lambda_{Tl}\, t \ln(\psi_{lit}) + \sum_{m=1}^{M-1} \varphi_{Tm}\, t \ln\left(\frac{P_{mit}}{P_{Mit}}\right) + \gamma_{TT} t^2 \end{array} \right. \\ S_m = \delta_m + 2.\delta_{mm} \ln\left(\frac{P_{mit}}{P_{Mit}}\right) + \sum_{l=1}^{L} \kappa_{lm} \ln(\psi_{lit}) + \varphi_{Tm} t \qquad m = 1,\ldots,(M-1) \end{array} \right\} \tag{3.6}$$

In addition to the symmetry and linear homogeneity in prices, the cost function has to be concave in input prices. This can not easily be imposed on the Translog function form since the restrictions are a function of the data. Instead, we compute the matrix of second derivatives of input prices at each data point to verify if it is negative definite; a necessary and sufficient condition for concavity in prices (see Diewert and Wales (1987) for the expression for a translog function). A further condition that is not imposed, but checked post estimation, is that the factor demand own-price elasticities are negative for all inputs. The Allen-Uzawa own-price elasticities and partial elasticities of substitution are given as:

$$\sigma_{mm} = \left(\delta_{mm} + S_m\left(S_m - 1\right)\right)/S_m^2 \tag{3.7}$$

and

$$\sigma_{mc} = \left(\delta_{mc} + S_c S_m\right)/S_c S_m \tag{3.8}$$

respectively. If σ_{mc} <0, the two inputs are complements, if σ_{mc} >0 then they are substitutes.

4.0 Data

We utilise a panel data set of 28 TOCs over 11 years (2000 to 2010[2]). The panel is unbalanced with a total of 244 observations in total. The unbalanced nature of the panel reflects the re-franchising and importantly, re-mapping of franchises over time.

We define TOC cost as total reported cost less access charge payments to Network Rail (the railway infrastructure manager). This definition follows from Smith and Wheat (2012b). We net off access charge payments as they are (indirectly) merely transfer payments from Government to the infrastructure manager and are not reflective of the cost of network access for a given TOC (at least in a given year). Importantly, TOCs are compensated for changes in the access charge payments over time by the construction of the franchise contracts[3]. It is therefore important to note that netting off access charge transfer payments to Network Rail

[2] Quoted years are for year end to 31st March e.g. 2000 is April 1999 to March 2000.

[3] It should also be noted that since 2001/02 Network Rail received some of its funding directly from central government via the Network Grant. As such the sum of access charges over all TOCs does not reflect the full cost of infrastructure provision for years beyond 2002. This is another reason that access charges do not reflect the opportunity cost of network access.

does not mean that we estimate a variable cost function. We consider that we estimate a total cost function since this cost represents the total cost under the control of the franchisee (for the duration of the franchise).

The cost data is sourced from the TOC's publicly posted accounts, while access charge payments are sourced direct from Network Rail. We believe these to be the best sources of these data given that the TOC accounts do not report access charges in a consistent manner across TOCs.[4]

Regarding the explanatory variables, Table 1 summarises the data. There are three primary outputs; route-km, train-hours and number of stations operated. We consider TOCs producing train services (train hours) and operating stations. In addition, route-km is included to distinguish between geographical size and intensity of operations. Thus it is analogous to the use of route-km in integrated railway studies to distinguish between scale and density effects (Caves et al, 1985). Conceivably route-km could have been included as a characteristic of the primary train hours output. However adopting this approach would have imposed, *a priori*, a more restrictive relation between scale and density effects; the hedonic function adopted imposes proportionality between the cost elasticity with respect to the primary output and the cost elasticity with respect to the quality variable. Given the focus of this study towards optimal size/utilisation of TOCs, it was deemed that the more flexible approach should be adopted.

With respect to other studies, we note a number of improvements in our specification of outputs. Firstly, we include both stations operated and train operations measures. Station operation is an important activity for some TOCs but less so for others and as such should not

[4] In particular it is obvious that some TOCs are itemising in their accounts only variable access charges rather than the sum of variable and (generally the much larger) fixed charge.

be ignored[5]. Only Smith and Wheat (2012b) considered stations within analysis. Secondly, we have train hours available for this study. This, along with distance measures (incorporated via average speed measures) and train length measures are the key drivers of costs since these measures include both time based and distance based cost drivers. We are not aware of any previous railway cost study, either of vertically integrated or separated railways, which has taken account of train hours, length and speed in the model.

A key element of this study is to consider the cost implications of merging TOCs which produce outputs with different characteristics. Therefore in addition to including the average characteristics of TOC output (train length, speed and passenger load factor), we include two further sets of measures to account for diversity in TOC service provision. The first is the proportion of train-km that correspond to each of three service groups (intercity, London and South Eastern (commuting) and the remainder regional). q_{42} and q_{52} pick up systematic cost differences, over and above that captured by the other output characteristics, from TOCs providing intercity and commuting services respectively (we drop the proportion for regional services to prevent perfect collinearity). For example, we can expect that intercity TOCs will, all other things equal, be more expensive due to such factors as the need to provide higher quality rolling stock and better on train services. As well as including these terms, we include interactions between the service group proportions. The majority of TOCs provide only one service group, thus the interaction variables are only non-zero for a select set of TOCs, the majority of which were formed from re-mappings of TOCs that provided a single service type but in the same geographical area, and have subsequently been merged into one. Thus the coefficients on these interaction variables would provide an indication of any cost increasing

[5] Two TOCs do not operate any stations. This is dealt with by modelling those TOCs as a cost function comprising only two outputs and the two input prices. Furthermore, we allow the coefficients with respect to the route-km (and the interactions with other variables) to be different for those TOCs that do operate stations.

(or decreasing) impact of TOCs providing heterogeneous service mixes, over and above any change in other service level characteristics.

Second, we also include the number of generic rolling stock types operated by a TOC. These are taken from the rolling stock classifications within the Department for Transport's Network Modelling Framework model. Essentially they classify rolling stock into speed bands and traction source (electric or diesel) and whether they are multiple units or loco-hauled. The more rolling stock types that are operated, the more likely there is heterogeneity in service provided within a TOC.

It should be noted that when it comes to evaluating franchise re-mappings, it will not just be the rolling stock type and franchise service type proportion heterogeneity that affect the cost change. Instead, the other average heterogeneity characteristic variables will be different. Thus it is difficult to assess the impact of changes in heterogeneity by looking at the signs on the service type and rolling stock type variables in isolation. We return to this in the results section.

We have defined two input prices, relating to payroll staff costs and non-payroll costs. Payroll staff costs include all labour costs from staff which are directly employed by the TOC. Thus a natural price measure is staff cost divided by staff numbers. The divisor for non-payroll staff is less clear. Firstly, once we net off access charge payments, the publically available accounts only do not allow for costs to be consistently broken up any further than staff and non-payroll costs. Non-payroll costs include rolling stock capital lease payments, rolling stock non-capital lease payments, other outsourced maintenance costs and energy costs and other costs. The only divisor that we have available is number of rolling stock units and we adopt this in the price. This is a limitation of the data, however we believe that this is the best solution (because classification issues between rolling stock and other costs mean that it is not possible to compute two separate prices for rolling stock and other; see also Smith and Wheat (2012b)).

We do check for concavity in input prices in our estimated model and this is fulfilled at all data points which gives us some reassurance that our input prices data are not having perverse effects. Perhaps the most important implication of our definition of input prices is that we would expect there to be a reasonable degree of substitutability between the two inputs at the margin since functions such as train maintenance can be outsourced and thus staff activity can be taken off the payroll.

[Table 1 here]

5.0 Results

We divide this section into four sub-sections. First we consider the suitability of the estimated model in terms of being consistent with economic theory and whether the model is suitably parsimonious. Such verification is important since otherwise the scale, density and heterogeneity properties of the model may originate from spurious accuracy rather than legitimate explanatory power. Second we focus on the scale and density properties of the model. Third we consider the impact of heterogeneity of output on costs and scale and density. Finally we show how these three factors (scale, density and heterogeneity) affect the expected cost changes for two specific mergers in our dataset and also for one hypothetical, but currently highly topical, potential merger.

5.1 Consistency with economic theory

The parameter estimates are shown in Table 2. The R^2 measure of fit for the cost function equation and the cost share equation are 0.928 and 0.489 respectively. The higher R^2 for the cost function primarily reflects the fact that the dependent variable is in logarithms while it is in levels in the cost share equation. The fitted cost shares are all between zero and one and we have evaluated the Hessian at each data point and found it to be negative definite for all observations; thus the function is concave in input prices over the relevant range.

[Table 2 here]

We have also computed the Allen-Uzawa own-price elasticities and partial elasticities of substitution (given in (3.8) and (3.9)). The mean estimated own-price elasticities are -0.297 and -1.345 for other expenditures and staff price respectively, which are both negative and so in line with expectations. The own-price elasticities are negative for all observations. The cross elasticity is 0.632 which is positive and thus indicates the two inputs are substitutes and this is the case when the elasticity is evaluated for each observation. This may reflect the degree to which some labour activitycan be taken in-house (therefore appear on payroll costs) versus be out-sourced (appearing under non-payroll costs). This is likely to be the case for non-capital rolling stock expenditure activities where maintenance can be performed in-house or by a third party or ROSCO. More generally, at the margin it is reasonable that there are some substitution possibilities between staff and rolling stock (capital) (choosing rolling stock that requires less staffing costs). Other restrictions such as homogeneity of degree one in input prices and symmetry are guaranteed by imposition.

On the basis of the above it thus appears that the estimated function does represent a cost function consistent with economic theory. As such we can have confidence that the estimated cost function can be used to infer the properties of the underlying technology.

We test several restrictions on the Translog both with a view of obtaining a more parsimonious function and to test economic hypotheses about the underlying technology. Of interest are:

- Homotheticity – the cost function is homothetic if it can be written as the product of a function in outputs and a function in input prices (and since we have panel data, time) i.e. $C(\boldsymbol{\psi}, \mathbf{P}, t) = f(\boldsymbol{\psi}).g(\mathbf{P}).h(t)$. Thus it requires that $\kappa_{1l} = 0$, $\lambda_{Tl} = 0$ l=1,2,3, $\kappa'_{12} = 0$, $\lambda'_{T1} = 0$ and $\varphi_{T1} = 0$ - 9 restrictions.

- Homogeneity – This refers to homogeneity in outputs. It is a special case of homotheticity in the sense that it implies unchanging returns to scale i.e. constant output elasticity i.e. $f(\boldsymbol{\psi}) = \psi_1^{\beta_1}\psi_2^{\beta_2}\psi_3^{\beta_3}$. It requires $\kappa_{1l} = 0$, $\lambda_{Tl} = 0$, $\beta_{lb} = 0$ l=1,2,3 b=1,2,3, $\kappa'_{12} = 0$, $\lambda'_{T1} = 0$, $\varphi_{T1} = 0$ and $\beta'_{l2} = 0$ l=1,2 - 17 restrictions.
- Unitary Elasticity of Substitution – This implies that $\sigma_{12} = 1$ in (3.8). This requires $\delta_{12} = 0$ which given the restrictions imposed by linear homogeneity of degree one in input prices implies $\delta_{11} = 0$ - 1 restriction
- Homogeneity and Unitary Elasticity of Substitution – This is the Cobb-Douglas restrictions (if we additional impose Homogeneity in the time trend) – 19 restrictions (additional $\lambda_{TT} = 0$)
- No hedonic characteristics – This requires $\phi_i = 0$ i=1,..,9. If this is supported the model reduces to one which is linear in parameters – 9 restrictions.

All hypotheses are rejected as reported in Table 3. This shows that the flexible specification is required to describe the underlying technology. Thus we retain the model in Table 2 as our preferred model and now discuss the findings on returns to scale and density.

[Table 3 here]

5.2 Returns to Scale and Density

As described in section 2 we have defined returns to scale (RtS) and returns to of density (RtD) specifically for train operations. RtS measures how costs change when a TOC grows in terms of geographical size. RtD measures how costs change when a TOC grows by running more services (measured by train-hours) on a fixed network. When we apply these definitions to the model in (4.1) then the expressions are:

$$RtS_{it} = \frac{1}{\left(\frac{\partial \ln C_{it}}{\partial \ln \psi_{1it}} + \frac{\partial \ln C_{it}}{\partial \ln \psi_{2it}} + \frac{\partial \ln C_{it}}{\partial \ln \psi_{3it}} \right)} \tag{5.1}$$

and

$$RtD_{it} = \frac{1}{\left(\frac{\partial \ln C_{it}}{\partial \ln \psi_{2it}} \right)} \tag{5.2}$$

The definition of RtD and RtS adopted is in relation to the hedonic output. Given the normalisation of train hours within the hedonic function, our findings on RtD and RtS with respect to ψ_2 can interchangeably be described in terms of variation in train hours (holding stations operated and network length and other things, including output characteristics, equal).

The rejection of the null hypothesis of homogeneity in outputs indicates that RtS and RtD will be non-constant and vary with the levels of the hedonic outputs, time and the level of prices. Figure 1 plots of RtS and RtD for all observations against train hours. 27% and 100% of observations exhibit increasing RtS and RtD respectively. The definitions of RtS and RtD are that there are increasing returns if the estimate is greater than unity, constant returns if the estimate is unity and decreasing returns if the estimate is less than unity. RtS and RtD evaluated at the sample mean of the data are 0.891 and 1.209 respectively. Constant RtS is rejected in favour of decreasing RtS at the 1% level (p-val=0.0055) and RtD is rejected in favour of increasing RtD at any plausible significance level (p-val<0.0000). Thus from these statistics it does seem that British TOCs exhibit increasing RtD but decreasing RtS.

This is an economically plausible finding. TOCs are likely to be able to lower unit costs by running more services on a fixed network i.e. increasing RtD. For example by better diagramming of rolling stock and staff they can reduce wasted time. Thus it is likely rolling stock can be used more intensively in a given time period which ultimately spreads any fixed lease charges over more units of output (train hours). Ultimately inputs into the production process suffer from indivisibilities and these can be more productively combined at higher usage levels[6].

However, TOCs may struggle to make unit cost savings or even prevent unit costs increasing when the size of the network served increases, holding utilisation (train hours per route-km) constant. This can arise since (to some extent) indivisibilities in inputs are route specific rather than network specific. For example, it can be envisaged that the utilisation benefits of running more trains between point A and B will be greater than utilisation benefits from running a set of services from A to B and then adding a new service from two unrelated points C and D. The latter scenario (for the same total train hours) is likely to require more rolling stock units and more staff hours than the former since there are two rather than one operational routes. To provide a less abstract (but extreme) example, the addition of a branch line to an existing network would not be expected to exploit higher utilisation of rolling stock since it is (almost) an independent operation to the rest of the network.

RtS is actually be found to be decreasing for some observations i.e. unit costs increasing as scale increases. To explain this we appeal to the common theory of the firm which considers that there is an optimal scale of a firm and that at some output level it gets very difficult to coordinate inputs, and thus unit costs start to rise (the firm is larger than the minimum efficient scale point). Note that the same pattern of variation in RtD is found, that is there exists a minimum efficient density level, but no TOC (yet) operates at a high enough density to attain it.

We now breakdown the RtS and RtD findings by TOC types – intercity, commuting (into London - LSE), regional and mixed TOCs. Figure 2 provides a plot which considers RtD

[6] Importantly indivisibility of inputs is a RtD issue rather than a cost efficiency issue since the explanation relates to the characteristic of the production technology rather than the extent to which minimum cost conditional on a level of output is achieved.

against train density for different TOC types holding all other characteristics[7] at the TOC type sample mean. We only plot over the density range of the central 80% of the distribution observed for each TOC type. This avoids showing RtD estimates from the model which are clearly out of sample and not realistic e.g. intercity TOC services always operate at low densities due to the long distance nature of the services and so are only plotted over this range.

Overall, holding characteristics at the sample mean and over the middle 80% of the distribution, Figure 2 shows that all TOC types exhibit increasing RtD and that this does fall with density, although RtD are never exhausted within the middle 80% of the sample. At any given train hours per route km level, intercity TOCs exhibit the lowest RtD, while LSE exhibit the strongest (and indeed even at the 90th percentile density in sample the RtD estimate is in excess of 1.2). Intuitively, the curve for mixed TOCs is somewhere in-between the curves for intercity and regional.

The policy conclusion from the analysis of RtD is that most TOCs should be able to reduce unit costs if there is further growth in train hours in response to future increases in passenger demand. This is important given the strong upward trend in passenger demand since rail privatisation in Britain and also noting that the trend seems to be continuing, even during the recession at the end of the sample period (Office of Rail Regulation, 2012). It is also relevant for recent policy in Britain following Sir Roy McNulty's Rail Value for Money study, since unit cost reductions of around 25% are targeted for the TOCs, and according to the results of our paper (increasing RtD), part of this unit cost reduction will occur naturally as train hours increase on a fixed network (though other savings will also be needed and the ability to grow volumes will be constrained to some extent by capacity and also by demand). In the wider EU context, the European Commission has aggressive targets for rail passenger usage and market

[7] In this sub-section 'characteristics' refers to all other variables in the cost function and not just the output characteristic variables in ψ_{2it}.

share which will increase passenger train density and therefore should reduce unit costs (assuming that train-km can be expanded without the need for investment in infrastructure). Our results show that the LSE service type has substantial scope for unit cost savings from increasing usage and this also holds for many regional TOCs given the large spread of usage levels across this group. However there is less scope for unit cost savings (and possibly a risk of decreasing RtD from large increases in usage) for intercity TOCs and regional TOCs at the high usage end of the spectrum.

[Figure 1 here] [Figure 2 here] [Figure 3 here]

Figure 3 provides a similar plot for RtS. This shows that for all of the central 80th percent of the train hours distribution, intercity (and mixed TOCs) exhibit decreasing RtS. LSE TOCs exhibit increasing returns to scale only for the very smallest in sample, whilst regional TOCs are the only TOC type to have an appreciable range of scale exhibiting increasing returns to scale. Thus our results are consistent with a u-shaped average cost curve, although it would appear that most TOCs are operating at or beyond the minimum unit cost point.

This finding has important implications for examining the optimal size of TOCs and is relevant to the recent franchise policy change that has resulted in substantial franchise re-mapping. The chief aim of many of these mergers was to capture the benefits of sharing of staff and rolling stock between services and to reduce the number of operators running out of London stations. This has tended to result in larger franchises e.g. Great Western re-mapping, which implies an increase in the size of TOCs which, given our findings on RtS, is likely to increase rather than reduce unit costs. However, there are a number of other factors that change through re-mapping TOCs relevant to our model, notably possible reduction in overlap of franchises (which increases the density of operation) and a move to a mixture of the type of services provided. We have demonstrated that TOCs tend to have increasing RtD which acts to reduce unit costs following TOC mergers. As discussed above, there are also important heterogeneity factors to take into account. Which effect will dominate in a given situation is an interesting research question. Once we have described our findings regarding heterogeneity we return to the cost implications for mergers, via a set of real world examples.

Finally in considering the policy implications of our findings on RtD and RtS, it must be remembered that our analysis concerns the costs of passenger train operations only. Just because unit costs can be reduced by running more train hours or by franchise remapping does not mean that this is the best course of action; best from the perspective of either minimising whole system cost or maximising welfare. There may be demand side constraints such that running extra train services may not yield a sufficiently large increase in passenger usage to justify the extra cost. There may also be a reduction in competition between franchises if franchise overlap is reduced, which may result in a net disbenefit. Finally running extra train services may have negative externalities to other services due to infrastructure congestion and other infrastructure costs. Thus this analysis should be used alongside analyses of other aspects of the railway system to evaluate the merits or demerits of specific interventions. Note that when we consider merging/remapping TOCs in 5.4, then these issues of congestion and demand side constraints are less important given we are simply rearranging the provision of existing services.

5.3 Implications of heterogeneity

We now turn to the impact of TOC heterogeneity on costs; the other variables populating the hedonic cost function i.e. the ϕ_{j2} j=1,..,9 variables and related coefficients in Table 1. Firstly, the elasticity of cost with respect to average train length, train speed and passenger load factor are proportional to the elasticity with respect to train hours, with the coefficient on the characteristic acting as the proportionality constant:

$$\frac{\partial \ln C_{it}}{\partial \ln q_{j2it}} = \phi_{j2} \frac{\partial \ln C_{it}}{\partial \ln \psi_{2it}} \text{ j=1,2,3} \tag{5.3}$$

All ϕ_{j2} j=1,2,3 coefficients are less than unity indicating the cost elasticities with respect to these characteristics are lower than for train hours. This is intuitive. Generally from an operations perspective, it is cheaper to add vehicles to existing trains (q_{12}) rather than run more train services (e.g. there is still only one driver). Likewise the passenger load factor coefficient (q_{32}) is very low which indicates the very low marginal cost of carrying extra passengers once the number of train hours and train length are controlled for. Finally the train speed coefficient (q_{22}) implies that running trains a greater distance, holding train hours constant, increases costs less than increasing train hours and distance together. This result will primarily be due to staff costs being time based rather than distance based, all other things equal.

In terms of implications for RtD and RtS, given the findings of decreasing RtD and RtS with the size of ψ_{2it}, a TOC operating the same train hours can be expected to have greater RtD and RtS if it operates shorter trains, slower trains and/or a has a lower passenger load factor. This follows from the fact that the level of the hedonic output, ψ_{2it} is found to be an increasing function of q_{12}, q_{22} and q_{32}. Furthermore, these findings are intuitive.

Turning to the findings specifically on the effect of TOCs providing a mixture of service types, which is given by the coefficients on the interaction proportion variables and number of generic rolling stock types operated i.e. q_{j2it} j=4,...,9. To explain the findings it is useful to consider some stylised examples. Table 4 presents the growth in the hedonic output Ψ_2 from the base case of a wholly regional TOC. Table 4 firstly considers the impact of mixing service types and then considers the additional impact of a TOC operating more rolling stock types which is likely when TOCs provide more service types (highlighted grey). Importantly it shows that while mixed TOCs are more expensive than regional TOCs, they are not more expensive than exclusively intercity or LSE TOCs, all other things equal. Adding in the effect of increasing rolling stock types increases the growth rate in the hedonic output further relative to a wholly regional TOC, however mixed TOCs still are less costly than pure intercity and LSE TOCs.

Thus Table 4 would indicate that allowing TOCs to produce mixed services is beneficial. However, it should be noted that heterogeneity and changes in heterogeneity are captured in our model via a complex set of variables (including train speed, train length and passenger load factor) as well as the TOC type dummies/number of rolling stocks etc. All these characteristics will change following a franchise re-mapping (and not just the TOC type dummies/ rolling stock variable). Thus the overall effect is a complex interaction of all heterogeneity characteristics, density, scale and input prices. As such when we actually consider specific re-mappings which result in mixed TOCs, the overall heterogeneity effect may actually be cost increasing (as is indeed the case in the Greater Western example consider in the next sub-section).

[Table 4 here]

5.4 The impact of franchise re-mapping

In this sub-section we consider how the estimated model predicts the cost change from re-mapping franchises[8]. The franchise re-mapping in recent years has, in most cases, the following implications:

---

[8] Note that we can not simply compare the sum of costs for the pre-re-mapped TOCs with those from the post-re-mapped TOCs because there is output, input price and technical change growth between the time periods that they are observed in our dataset. Further the last year and first year of data are often cost data with the most measurement error given the required adjustments to align costs to match a standard financial years (when in fact re-mappings occur within years). Thus we use the model to predict the cost change.

- In general there has been a rationalisation to larger franchises. Thus there will be scale effects, which given the finding of decreasing RtS for large TOCs could increase unit costs.
- Irrespective of whether the re-mapped TOC(s) are larger, the move to integrating TOCs of various service types results in a removal of franchise overlap which implies that the sum of the route-km for all the re-mapped TOC(s) will be less than the sum of the route-km for the previous TOCs. This implies that for a given usage level (train hours), density of usage increases. Thus there will be density effects, which given the finding of increasing returns to density, implies a decrease in unit costs.
- The re-mapped franchises now provide more than one service type, as opposed to the previous TOCs which, in most part, operated only one service type. Thus the TOCs formed from re-mapping will have TOC heterogeneity measures (length of train, average speed etc.) which are weighted averages of the previous TOCs. This will not necessarily be cost neutral given the flexible form that the quality variables enter into the model (there are non-constant elasticity effects in the model). The new TOCs will also have non-zero values for some of the TOC service type heterogeneity interaction terms i.e. there will be effects from the TOC providing a mixed service. Furthermore, they may be operating different numbers of rolling stock types (see Table 4).

The extent to which mergers can deliver cost savings through exploiting increasing RtD depends on the relative heterogeneity characteristics before and after re-mapping. We quantify this effect by providing the evaluated ψ_2 divided by route-km for the TOC, which is termed the 'heterogeneity adjusted (HA) density' measure. It is this that determines the extent to which a TOC can exploit any increasing RtD since RtD is defined with respect to the hedonic output. It should be noted that it is the proportional change in this measure from the before to after re-mapping situation which gives the extent to which density is changing; the absolute number is meaningless (it is a function of the units of the data). If the proportional change in HA density is greater than the proportional change in train hours density then we say heterogeneity is reinforcing the returns to density (and scale) effects. This is because the density measure that is actually driving RtD/RtS is increasing more than the naive measure of density (train hours density). Similarly, if the reverse is true we say heterogeneity is dampening the RtD (and RtS) effects.

[Table 5 here]

Clearly, a priori for a given merger, there are conflicting effects; with increasing density generally reducing costs, increasing scale of operations increases costs and the impact of changes in heterogeneity being ambiguous. We consider two real world mergers and also a hypothetical merger, which is quite topical at present, due to the policy aspiration of several northern English regions to expand and become franchisor of the enlarged Northern franchise. The characteristics of each merger are described in Table 5, alongside the predicted cost changes. We can make the following observations:

- Greater Western merger – This is found to increase costs. This is for two reasons. Firstly, there is an exhaustion of RtS i.e. the new franchise is simply too large. Secondly there is a large fall in the impact of heterogeneity on ψ_2. The result is that while train hours density increases by 57%, heterogeneity adjusted train density increases by only 12%. This implies that the Greater Western TOC is unable to exploit increasing RtD as much as we would expect based on the large increase in train density, thus there is only a weak off-setting cost reduction effect from density relative the cost increasing scale effect (the impact of heterogeneity is to dampen any density effect).

- London Eastern re-mapping – This is found to decrease costs. Importantly both the new franchises have substantial increasing RtD and one TOC still has large increasing RtS (the other has constant returns to scale). Thus we conclude that these TOCs are not past the minimum efficient scale points.
- New Northern franchise – This results in a small increase in costs. This seems to be due to the decreasing RtS faced by both the Northern and New Northern TOCs. Furthermore, it is predicted by the model that the New Northern franchise will have exhausted RtD. Overall the effect of heterogeneity changes is approximately neutral from one mapping to the other.

6.0 Conclusion

The contributions of this paper are as follows.

(1) It has been argued from a theoretical perspective, and demonstrated via an empirical example, that econometric estimation of economies of scale and density in passenger train operations requires careful attention to the modelling of heterogeneity between train operators. In particular, the power of a hedonic translog cost function containing train hours (in place of train-km) – a data innovation in itself – together with a number of TOC characteristics within the hedonic function, is demonstrated. Based on this approach, it is possible to distinguish between different scale and density effects depending on the output characteristics of the TOC and not just the usual overall output level and input price level as in a simple (non-hedonic) translog cost function.

(2) In the British policy context we use our model to study the cost implications of the cost implications of three actual (or proposed) TOC mergers. This analysis demonstrates the importance of modelling the intricate relationship between cost and scale, density and heterogeneity explicitly. In particular, changes in heterogeneity characteristics played a substantial role in the Great Western re-mapping since these changes prevented

exploitation of the returns to density. Since franchise mergers also reduce on rail competition which maybe undesirable (Jones, 2000), the supposed cost savings from exploiting RtD are important in supporting the case for mergers. It is therefore illuminating that our study suggests that these returns may not be realised in all cases.

(3) Though our empirical example is focused on the British TOCs it also has wider implications. Our findings suggest that previous estimates of scale and density properties in railways internationally may have been biased, to the extent that they did not adequately model the interaction between scale /density and heterogeneity of services. In terms of regulatory policy, in interpreting evidence on scale and density returns in railways, our model suggests that policy makers need to take service heterogeneity into account. Failure to do so may mean that policy decisions are made on the basis of supposed scale/density returns that cannot be realised in practice. Modelling railway operations is complex and thus to address specific policy questions (such as the cost implications of mergers) a rich model, such as that developed in this paper, is required.

7.0 References

Affuso, L., A. Angeriz, and M.G. Pollitt (2002): 'Measuring the Efficiency of Britain's Privatised Train Operating Companies', Regulation Initiative Discussion Paper Series, no: 48, London Business School.

Affuso, L., A. Angeriz, and M.G. Pollitt (2003): 'Measuring the Efficiency of Britain's Privatised Train Operating Companies', mimeo (unpublished version provided by the authors).

Andersson, M, Smith, A., Wikberg, A. and Wheat, P. (2012): 'Estimating the marginal cost of railway track renewals using corner solution models', *Transportation Research Part A,* 46 (6), 954-964.

Bitzan, J. D. And Wilson , W.W. (2008). ‘A hedonic Cost Function Approach to Estimating Railroad Costs’. *Research in Transportation Economics*, 20. 69-95.

Button, K. (2010). *Transport Economics*. 3rd Edition, Edward Elgar Publishing Ltd.

Caves, D.W., Christensen, L.R. and Tretheway, M.W. (1984), ‘Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ’, *The RAND Journal of Economics,* 15 (4) 471-489.

Caves, D.W., Christensen, L.R. and Swanson, J.A. (1981), ‘Productivity Growth, Scale Economies, and Capacity Utilisation in U.S. Railroads, 1955-74, *American Economic Review*, vol. 71, issue 5, pp. 994-1002.

Caves, D.W., Christensen, L.R., Tretheway, M.W. and Windle, R.J. (1985), ‘Network Effects and the Measurement of Returns to Scale and Density for U.S. Railroads’, in Daughety, A.F., ed., *Analytical Studies in Transport Economics*, Cambridge, Cambridge University Press, pp. 97-120.

Cowie, J. (2002a): ‘Subsidy and Productivity in the Privatised British Passenger Railway’, *Economic Issues* 7 (1), 25-37 38

Cowie, J. (2002b): ‘The Production Economics of a Vertically Separated Railway – The Case of the British Train Operating Companies’. *Trasporti Europei*, August 2002, 96-103.

Cowie, J. (2005): ‘Technical Efficiency versus Technical Change – The British Passenger Train Operators’, in Hensher, D. A. (ed.) *Competition and ownership in land passenger transport: selected refereed papers from the 8th International Conference (Thredbo 8), Rio de Janeiro, September 2003*, London, Elsevier, 2005.

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), 85-104.

Diewert, W. E and Wales, T. J. (1987): 'Flexible Functional Forms and Global Curvature Conditions' *Econometrica* 55 (1), 43-68.

European Commission (2013). The Fourth Railway Package – Completing the Single European Railway Area to Foster European Competitiveness and Growth. COM (2013) 25 Final.

Hensher, D. and Brewer, A (2000). *Transport: An Economics and Management Perspective.* Oxford University Press.

Jara-Diaz, S. R (1982). 'The estimation of transport cost functions: a methodological review'. *Transport Reviews,* 2 (3), 257-278.

Jones, I. (2000). 'Developments in Transport Policy. The Evolution of Policy Towards On-Rail Competition in Great Britain'. *Journal of Transport Economics and Policy,* 34 (3), 371-384.

Kumbhakar, S.C. (1997) 'Modelling allocative inefficiency in a Translog cost function and cost share equations: an exact relationship'. *Journal of Econometrics* 76 351–356

Kumbhakar, S.C. and Lovell, C.A.K (2000). *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge UK.

Merkert, R. smith, A.S.J. and Nash, C.A. (2009). 'Benchmarking of train operating firms - A transaction cost efficiency analysis', *Journal of Transportation Planning and Technology*.

Office of Rail Regulation (2012). *National Rail Trends*. Available at http://www.rail-reg.gov.uk/server/show/nav.2026 accessed 29/11/2012.

Oum, T.H., Waters, W.G. (II) and Yu, C. (1999), 'A Survey of Productivity and Efficiency Measurement in Rail Transport', *Journal of Transport Economics and Policy*, vol. 33 (I), pp. 9-42.

Preston, J. (2008b), “A Review of Passenger Rail Franchising in Britain: 1996/7 – 2006/7”, *Research in Transportation Economics,* 22, pp. 71-77.

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.

Smith A.S.J and Wheat P. (2012a). ‘Estimation of Cost Inefficiency in Panel Data Models with Firm Specific and Sub-Company Specific Effects’ *Journal of Productivity Analysis.* 37 (1) 27-40.

Smith A.S.J and Wheat P. (2012b). ‘Evaluating alternative policy responses to franchise failure: Evidence from the passenger rail sector in Britain’ *Journal of Transport Economics and Policy*, 46 (1), 25-49.

Smith, A.S.J., Wheat, P. and Nash, C.A. (2010), ‘Exploring the Effects of Passenger Rail Franchising in Britain: Evidence from the First Two Rounds of Franchising (1997-2008)’, *Research in Transportation Economics,* 29 (1) 72-79.

Smith, A.S.J. (2006): ‘Are Britain’s Railways Costing Too Much? Perspectives Based on TFP Comparisons with British Rail; 1963-2002’, *Journal of Transport Economics and Policy*, 40 (1), 1-45.

Wheat, P. and Smith, A.S.J. (2008): ‘Assessing the Marginal Infrastructure Maintenance Wear and Tear Costs for Britain’s Railway Network’ *Journal of Transport Economics and Policy*, 42 (2), 189-224.

Table 1 variables used

Symbol	Name	Description	Data Source
Generic Outputs (ψ)			
$\psi_1 = y_1$			
y_1	Route - km	Length of the line-km operated by the TOC. A measure of the geographical coverage of the TOC	National Rail Trends
$\psi_2 = y_2 q_{12}^{\phi_{12}} q_{22}^{\phi_{22}} q_{32}^{\phi_{32}} e^{\phi_{42} q_{42}} e^{\phi_{52} q_{52}} e^{\phi_{62} q_{62}} e^{\phi_{72} q_{72}} e^{\phi_{82} q_{82}} e^{\phi_{92} q_{92}}$			
y_2	Train Hours	Primary driver of train operating cost	National Modelling Framework Timetabling Module
q_{12}	Average vehicle length of trains	Vehicle-km / Train-km	Network Rail
q_{22}	Average speed	Train-km / Train Hours	National Modelling Framework Timetabling Module
q_{32}	Passenger Load Factor	Passenger-km / Train km	Passenger-km data from National Rail Trends. Train-km data from Network Rail.
q_{42}	Intercity TOC	Proportion of train services intercity in nature	National Rail Trends for the categorisation of TOCs into intercity, LSE and regional. Where TOCs have merged across sectors a proportion allocation is made on an approximate basis with reference to the relative size of train-km by each pre-merged TOC
q_{52}	London South Eastern indicator	Proportion of train services into and around London (in general commuting services)	
q_{62}	$q_{42} q_{52}$	Interaction between Intercity and LSE proportions	
q_{72}	$q_{42}(1 - q_{42} - q_{52})$	Interaction between intercity and regional (non-intercity and non-LSE services) proportions	
q_{82}	$q_{52}(1 - q_{42} - q_{52})$	Interaction between LSE and regional proportions	
q_{92}	Number of rolling stock types operated	Number of "generic" rolling stock types operated	National Modelling Framework Rolling Stock Classifications
$\psi_3 = y_3$			
y_3	Stations operated	Number of stations that the TOC operates	National Rail Trends
Prices			
P_1	Non-payroll cost per unit rolling stock		TOC accounts for cost, Platform 5 and TAS Rail Industry Monitor for rolling stock numbers
P_2	Staff costs (on payroll) per number of staff		TOC accounts (both costs and staff numbers)

Table 2 Parameter Estimates

Parameter	Estimate	P-val		Parameter	Estimate	P-val	
Main parameters				*Hedonic output (ψ_2) parameters*			
α	7.729	0.001	***	ϕ_1	0.701	0.000	***
β_1	-1.831	0.000	***	ϕ_2	0.856	0.000	***
β_2	-0.464	0.256		ϕ_3	0.059	0.609	
β_3	0.592	0.076	*	ϕ_4	0.425	0.031	**
δ_1	1.048	0.000	***	ϕ_5	0.309	0.005	***
γ_T	0.039	0.420		ϕ_6	-1.520	0.002	***
β_{11}	0.100	0.003	***	ϕ_7	-0.157	0.763	
β_{22}	0.048	0.048	**	ϕ_8	-0.463	0.631	
β_{33}	0.109	0.000	***	ϕ_9	0.021	0.139	
β_{12}	0.078	0.045	**				
β_{13}	-0.189	0.000	***	*No-stations model free parameters*			
β_{23}	0.010	0.819					
δ_{11}	0.080	0.000	***	β_1'	-1.170	0.011	**
κ_{11}	-0.058	0.000	***	β_{11}'	0.035	0.323	
κ_{12}	0.067	0.000	***	β_{13}''	0.050	0.335	
κ_{13}	0.004	0.545		κ_{11}'	-0.046	0.000	***
λ_{T1}	0.002	0.663		λ_{T1}'	0.005	0.278	
λ_{T2}	-0.008	0.119					
λ_{T3}	0.002	0.545		R^2			
φ_{T1}	-0.006	0.000	***	Cost Function		0.928	
γ_{TT}	-0.001	0.539		Share Equation		0.489	

***, **, * Statistically significant from zero at the 1%, 5% and 10% levels respectively

Table 3 Results of specification tests

	Homotheticity	Homogeneity	Unitary Elasticity	Cobb-Douglas	Hedonic
Number of Restrictions	9	17	1	19	9
Test statistic - Chi-sq	142.24	371.11	360.63	660.79	114.48
p - val	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4 Heterogeneity findings – Growth in hedonic output (ψ_2) relative to a regional only TOC

TOC Type Composition			Increase in rolling stock types	Growth rate	p-val	
Regional	**LSE**	**Intercity**				
100%	0%	0%	0	0.0%	0.000	***
0%	100%	0%	0	36.2%	0.000	***
0%	0%	100%	0	52.9%	0.000	***
33%	33%	33%	0	0.7%	0.588	
50%	50%	0%	0	3.9%	0.563	
0%	50%	50%	0	-1.3%	1.603	
50%	0%	50%	0	18.9%	0.000	***
33%	33%	33%	6	14.5%	0.000	***
50%	50%	0%	3	10.8%	0.157	
0%	50%	50%	3	5.2%	0.002	***
50%	0%	50%	3	26.8%	0.000	***

Notes: a) The growth rate is constructed as the percentage increase in ψ_2 resulting from a change in the composition of the TOC relative to the base case (a 100% regional TOC). Formally Growth rate = $\left(e^{\phi_{42}q_{42}}e^{\phi_{52}q_{52}}e^{\phi_{62}q_{62}}e^{\phi_{72}q_{72}}e^{\phi_{82}q_{82}}e^{\phi_{92}q_{92}}\right)-1$.

b) The computation is indifferent to the number of rolling stock types in the base case

c) We illustrate the impact of combining rolling stock types by implicitly assuming each TOC type operates three unique rolling stock types.

Table 5 The predicted cost impacts of franchise re-mapping

Year of remapping	Name	TOC Type	Route-km	Train-hours	Train Hours Density	Hetrogeniety Adjusted Density	EoS	EoD	Predicted Cost
	Pre-remapping TOCs								
2006/07	Great Western	Intercity	1368	598	0.437	165.1	1.519	1.573	278
	Great Western Link	LSE	581	550	0.947	108.1	1.578	1.729	138
	Wessex	Regional	1394	529	0.380	26.2	1.183	1.421	92
	Total		*3343*	*1677*	*0.502*	*97.3*			*508*
2004/05	Anglia	Regional	669	312	0.467	69.2	1.416	1.614	47
	Great Eastern	LSE	235	555	2.362	404.8	1.492	1.808	95
	WAGN	LSE	414	886	2.139	300.8	1.290	1.620	167
	Total		*1318*	*1753*	*1.330*	*201.8*			*308*
2010/11 - hypothetical	Northern	Regional	2746	2597	0.946	48.5	0.807	1.108	389
	Transpennine Express	Regional	1251	633	0.506	40.4	2.069	1.790	137
	Total		*3996*	*3230*	*0.808*	*46.0*			*527*
	Post-remapping TOCs								
Year of remapping	**Name**	**TOC Type**	**Route-km**	**Train-hours**	**Train Hours Density**	**Hetrogeniety Adjusted Density**	**EoS**	**EoD**	**Predicted Cost**
2006/07	Greater Western	Mixed	2129	1677	0.788	109	0.892	1.188	554
	Total		*2129*	*1677*	*0.788*	*109*			*554*
2004/05	ONE	Mixed	1001	1028	1.027	142	0.996	1.339	170
	Great Northern	LSE	275	725	2.637	383	1.735	1.915	120
	Total		*1276*	*1753*	*1.374*	*194*			*290*
2010/11 - hypothetical	New Northern	Regional	3019	3230	1.070	62	0.724	0.990	579
			3019	*3230*	*1.070*	*62*			*579*

Year of remapping	Percentage change in Characteristics (+ indicates increase)					Cost Change	
	Name	Route-km	Train-hours	Train Hours Density	Hetrogeniety Adjusted Density	£'000	Percent
2006/07	Greater Western	-36%	0%	57%	12%	**45.6**	**9%**
2004/05	ONE/Great Northern	-3%	0%	3%	-4%	**-17.9**	**-6%**
2010/11 - hypothetical	New Northern	-24%	0%	32%	34%	**52.6**	**10%**

Notes: 1) Method for calculating metrics for Post-remapping TOCs: Route-km: taken from actual values in subsequent years; Train-hours: sum of pre-remapping TOCs allocated to post-remapping TOCs through proportion split between post-remapping TOCs in a subsequent year; Predicted cost - in addition to the aforementioned variables, assumptions needed to be made regarding the level of other variables in the function i) input prices - averages of input prices for pre-mapping TOCs ii) levels of other variables in the hedonic output function - taken from actual data for post-remapping TOCs in the subsequent year iii) number of stations operated is taken from subsequent year data for post-remapping TOCs.
2) The New Northern TOC is hypothetical: Measures are calculated as in 2) with the exception: i) route-km this is given as the Northern route-km plus the additional route length of Transpennine Express of the North West route to Glasgow ii) number of stations operated is the sum of the stations operated by the two merging TOCs.

Figure 1 Estimated Returns to Scale and Density against Train Hours for the sample

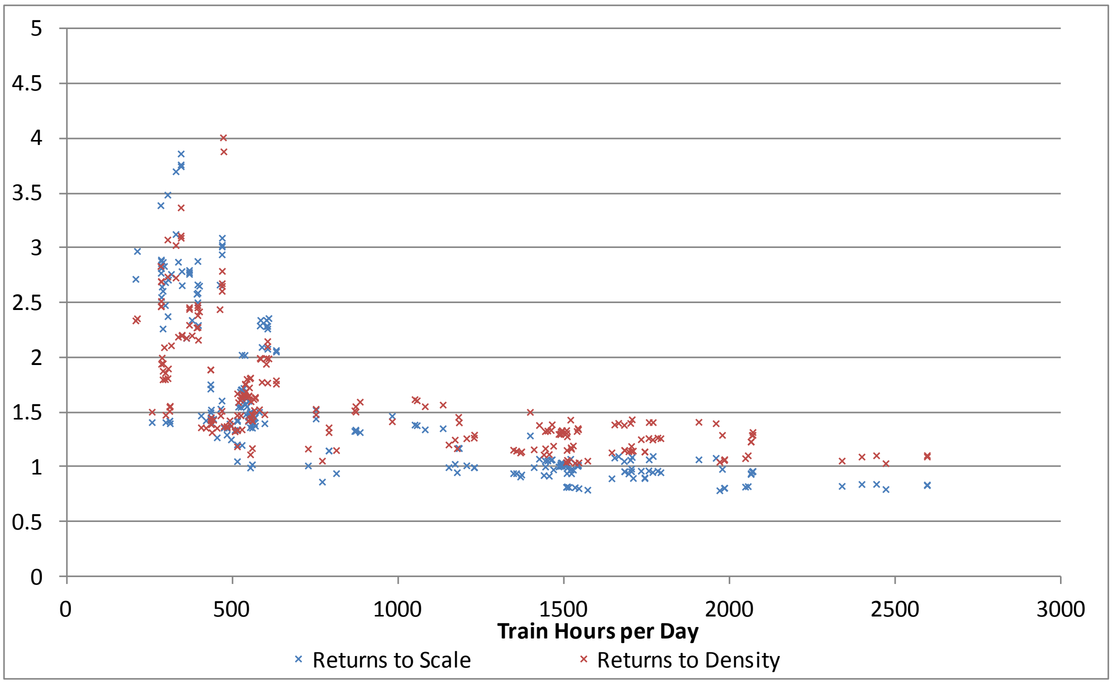

Figure 2 Returns to density for different TOC types holding other variables constant

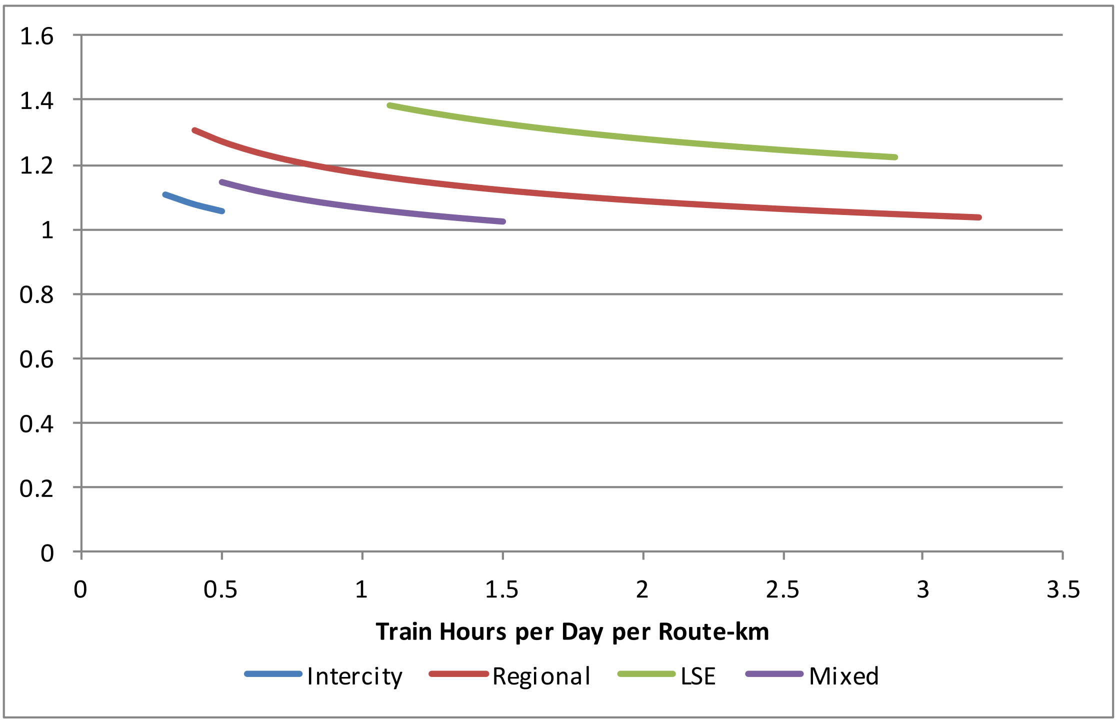

Figure 3 Returns to scale for different TOC types holding other variables constant