

AN EFFICIENCY ANALYSIS OF RAIL PASSENGER SUBSIDIES IN GERMANY COMBINING PERCEIVED AND OBJECTIVE QUALITY OF SERVICE MEASURES AND FINANCIAL INDICATORS

**Paper submitted to the Kuhmo Nectar Conference on Transport Economics, Barcelona
2017**

Heike Link

German Institute for Economic Research (DIW Berlin)
Department Energy, Transport, Environment
Mohrenstr. 58, 10117 Berlin, Germany
e-mail: hlink@diw.de

1. Background

The German approach of franchising regional subsidized rail passenger services, introduced in 1996, has widely been considered as successful in terms of cost savings, service provision and patronage (see Link and Merkert, 2011, Nash et al., 2013, Link, 2016). While at the global scale not only service provision but also traveller satisfaction and punctuality have improved during the first years of franchising, concerns on a decline of quality of service (QoS) have been raised over the last decade. Not only, but also as a response on this, PTAs have started to develop quality reporting schemes within contract controlling. Similar to cost savings, there is a general trend that overall punctuality and traveller satisfaction in regional rail passenger transport have increased but that there are considerable differences between the federal states and PTAs. Cancellations of trains and lack of punctuality in particular in large conurbations such as Berlin, Stuttgart and the Rhine-Ruhr area the most prominent examples of declining service quality, and the crisis of the Berlin S-Bahn system since 2008 has become well-known even outside Germany. These quality concerns challenge a pure technical efficiency analysis without considering quality of service which might lead to different rankings of DMUs, different explanations for efficiency differences between them and might therefore even support misleading conclusions.

Against this background, this paper combines for the first time in rail efficiency analysis both physical output measures such as train-km and passenger-km with quality of service indicators in order to benchmark the sample DMUs simultaneously against conventional and QoS indicators and to explain differences between such combined efficiency scores. For this analysis, the database used in Link (2016) is extended in two directions. First, the analysis

uses more disaggregated data on the level of public transport authorities (PTAs) which are responsible for contract procurement and day-to-day management including quality controlling. Second, the inputs used in Link (2016) - operation subsidies and track access charges - are extended by including also investment subsidies given the presumed impact on quality of service, and the outputs are extended by quality indicators such as punctuality and traveller satisfaction.

Specifically, the following research questions are analysed:

- 1) What is the effect of including quality indicators into efficiency analysis in terms of rankings, efficiency estimates and determinants of efficiency compared with not including quality indicators?
- 2) What is the effect of different options to include quality of service indicators into the analysis?

The remainder of this paper is organised as follows. Section 2 introduces the institutional framework for franchising RRPS with a particular focus on regulating quality of service within the contracts. Section 3 reviews available research on the inclusion of QoS into efficiency analysis. Section 4 and 5 describe the methodology and the database used in this paper. Section 6 reports the estimation results and section 7 concludes.

2. The institutional framework for franchising and quality of service

Since 1996 the federal states have been responsible for procuring subsidised regional rail passenger services (RRPS)¹ from rail companies and for financing them within franchise contracts (so-called regionalisation)². Financially, the franchising system is based on transfers from the federal budget to the federal states at an annual level of about €7 billion³ (so-called regionalisation funds – Regionalisierungsmittel) which can be regarded as a stable and sound public financing for PSOs. The federal states use different institutional approaches to organise this process. While for example the states of Hamburg and Bremen preferred a

¹ The term regional rail services might be misleading because it does not refer necessarily on the length of lines. Regional trains often operate over long distances.

² The aim of this so-called “regionalisation” approach was twofold. Firstly, it aimed to achieve a clear distinction between cost-covering (or even profitable) services (commercial services), and subsidised services operated as Public Service Obligations (PSOs). Secondly, it aimed to achieve a decent level of PSOs at a lowest subsidy possible.

³ In 2014, the final year for most of the results presented in this paper, regionalisation funds were €7.3 billion. In 2015 the funds amounted at €7.4 billion and for 2016 around €8 billion were available with an agreed increase of 1.8% p.a. in subsequent years.

central responsibility for procuring RRPSs without a separate PTA, other states (for example Bavaria, Thuringia, Saxony-Anhaltine, Berlin, Brandenburg) have founded one central PTA being responsible for all RRP in the state, and a few countries have delegated responsibility to some type of municipal or regionally defined PTAs (for example Hesse, Baden-Wurttemberg). While the procedures of awarding contracts (directly awarded or tendered), and contract controlling as well as all other day-to-day management tasks are under the responsibility of the PTAs, the federal states decide about the general strategy (intensity and pace of tendering, financial aids for rolling stock), the allocation of funds to the PTAs, and whether to spend additional money from the federal states general budgets for regional rail passenger transport. They also have some freedom to decide how the regionalisation subsidies (Regionalisierungsmittel) from the Federal Government are split between operating subsidies granted to TOCs, and investments or – to some extent – to other public transport.

Currently, there are 27 regional authorities (PTAs) responsible for procuring regional rail passenger services. They differ regarding the area to be served, and there are also differences in the number of PTAs within the federal states. The legal framework gives – within the requirements of EU legislation - significant freedom to PTAs, allowing to award service contracts using i) open tenders; ii) non-open tenders; and iii) negotiations. Service contracts have different contractual forms and different degrees of service specifications, varying contract durations and refer to different network sizes (ranging from single lines up to large regional network bundles). In general, all regional rail services are awarded as public service contracts on a non-exclusive basis. The decentralised institutional set-up implies a lack of a general standard for regional rail service contracts. PTAs rather adapt contract features to regional conditions and experience.

Public service contracts for regional rail passenger services usually contain detailed definitions on quality of service such as type and age of rolling stock, required service staff in the trains, train length, number of carriages etc. Most PTAs use bonus/malus payments and penalty management for both the amount of train-km and patronage and for quality management. The majority of PTAs has been establishing quality reporting systems for the purpose of bonus/malus payments and penalty management. The lack of general standards for rail service contracts implies that quality reporting varies between PTAs. This holds for quality of service definitions, the level of quality of service contractually agreed and hence for the indicators collected. Most PTAs collect as a minimum train cancellations (where some

PTAs distinguish between cancellations due to external factors and due to the operating company's failure) and punctuality; however punctuality definitions vary between 3, 5 and 6 minutes delays of trains. Apart from these commonly collected indicators, some PTAs also compile indicators such as cleanness of trains, functioning of passenger information systems, and compliance with the contractually agreed number of carriages per train and service staff in the trains. Many PTAs also report traveller satisfaction based on own surveys, again with different measurement concepts, level of detail (different sets of dimensions such as satisfaction with trains, stations, punctuality, traveller information, staff etc.) and with different scales (mostly varying between 1-5 and 1-6 scales). Except for passenger satisfaction, quality indicators are collected for each service contract and used for contract management of PTAs. In most cases this information is not publicly available. Apart from this, traveller satisfaction is also subject to nation-wide surveys (see TNS Infratest, Infas) which have the advantage of common definitions over regions and time.

3. Available research

Most of the available studies on rail productivity and efficiency focus on the impact of vertical integration versus vertical separation (see for example Ivaldi and McCullough, 2001, Bitzan, 2003, Jensen and Stelling, 2007, Growitsch and Wetzel, 2009, Cantos *et al.*, 1999, Cantos *et al.*, 2010 and Mituzani and Uranishi, 2010). Another stream of research on rail efficiency deals with the impact of competitive tendering and franchising and on contractual arrangements (see Driessen *et al.*, 2006, Cantos *et al.*, 2010, Nash and Smith, 2006, Affuso and Newbery, 2002, Beck, 2011, Lalive and Schmutzler, 2008, 2011, Hunold and Wolf, 2012 and Link, 2016). The majority of studies, in particular those on vertical separation, but also most of the studies on the impact of competitive tendering are based on international data at country-level, mostly obtained from UIC statistics. A few studies on competitive tendering are based on either full samples of franchising contracts in a single country (see Nash and Smith, 2006, Affuso and Newbery, 2002 for UK franchise contracts and Link, 2016 for Germany) or on subsamples (see Beck, 2011, Lalive and Schmutzler, 2008, 2011, Hunold and Wolf, 2012 for Germany). All of these studies (except Link, 2016) focus on the impact of vertical separation and competitive tendering respectively on rail companies' efficiency rather than on the use of public spending for rail. Link (2016) differs by focusing on efficiency analysis in the use of rail passenger subsidies and regarding the choice of the German decision making units (DMUs) which are responsible for subsidising regional rail passenger services.

From a methodological perspective, there is no common trend to favour one of the two standard tools for efficiency analysis, e.g. either stochastic frontier analysis or data envelopment analysis.

Research on the incorporation of quality of service into cost function and efficiency analysis is scarce, given the difficulties in defining quality criteria, collecting the necessary indicators and incorporating them into the tools of efficiency analysis. The few available studies are mainly concentrated on the electricity sector, presumably because of the development and refinement of incentive-based regulation schemes where quality of service has started to attract attention. In studies of the electricity sector, quality of service is usually represented by one or two indicators such as electricity outages or interruption time (see for example Korhonen and Syrjänen, 2003, CEPA, 2003, Ajodhia et. Al., 2004, Growitsch et al., 2005, Giannakis et al., 2005). Although the relationship between organisational forms, franchising and competitive tendering in public transport and traveller satisfaction as an indicator of quality of service has been gaining attention over the last years (see for example Mouwen and Rietveld, 2013, Zhang et al., 2016, Paha et al, 2013), to the best of our knowledge quality aspects have so far not been included in any analysis of productivity and efficiency in rail or public transport. Rare examples of benchmarking and efficiency studies in the transport sector which include QoS indicators refer to air transport. Assaf et al. (2014) include airport delays into an efficiency analysis for airports and Merkert and Assaf (2015) measure airport efficiency in a combined quality/profitability indicator which includes traveller satisfaction as a quality of service variable.

The rather small research evidence and application experience implies that any state-of-the-art approach of incorporating quality into efficiency analysis is lacking. Most often QoS is reflected within indicators which measure quality reductions such as outage numbers, interruption times, customer minutes lost or airport delays. Usually they are treated as undesirable or “bad” inputs which have to be minimised against given output levels, while Merkert and Assaf (2015) is the only example of including a QoS measure as output by using customer satisfaction indicators.

In summary, although QoS in rail passenger transport but also in other transport sectors has been perceived as an important dimension of the service, there is an apparent lack of research which combines economic performance with the level of quality of service. The research

presented in this paper aims at contributing to this field by comparing a two-stage DEA model without quality indicators with two model versions which include one objective quality indicator and one measure of perceived quality of service.

4. Methodology

The approach in this paper follows the one taken in Link (2016) in so far as - in contrast to studies on rail companies' efficiency - the efficiency in using rail passenger subsidies are analysed. This different research focus implies a different formulation of inputs than in traditional rail efficiency analysis, e.g. the models do not include the typical inputs such as labour, material (e.g. energy) and capital (e.g. rolling stock, other facilities) as variables, but define monetary subsidies (further disaggregated into operating subsidies, investment subsidies and infrastructure charges) as inputs⁴.

This paper employs a two-stage bootstrapped efficiency analysis with a DEA application in the first stage and a Tobit panel model for explaining the first-stage efficiency scores. The first stage DEA is a deterministic, non-parametric method to specify the efficient production frontier. Based on observed inputs (w) and outputs (y) for a set of N decision-making units (DMUs), relative efficiency scores for each unit are derived by searching for the most efficient DMUs in the sample, that is those units with the lowest input levels for a given level of output⁵. These form the efficiency frontier which envelopes all observed data points in the sample. Efficiency of each DMU is measured by the distance Ω between the single observation and the frontier.

The main reason for choosing DEA as the tool of analysis in this paper was its capability of handling multiple outputs and inputs and of dealing with data which are on different units. This was a major advantage for incorporating QoS variables which are on different measurement units than the conventional variables at physical or monetary units, and which

⁴ Ideally, splitting up subsidies further into spending on labour, material and capital would enable to link the analysis to traditional rail efficiency analysis and could provide further insights into complementarity and substitution properties of inputs, in particular if the data would allow an econometric cost function approach such as stochastic frontier models. However, this type of disaggregating subsidies appears to be empirically impossible.

⁵ The performance of a DMU is defined as the ratio between the weighted sums of outputs and inputs whereby the weights are allocated within the DEA model.

are also amongst themselves on different units. Furthermore, DEA does not require any assumptions about cost minimisation or profit maximisation or about the functional form of the frontier and does not require price information. Caveats of the method include the potential danger of deeming too many observations as inefficient because the concept fails to recognise nonconvexities of the envelope (see for example De Borger and Kerstens, 1996). A major problem is the deterministic nature of the method which allocates variations in firm's performance exclusively to inefficiency, and which lacks statistical inference. This, however, can be solved by bootstrapping procedures to correct for the potential bias in the estimated distance function from the true frontier and to calculate confidence intervals (see for example Simar and Wilson, 2000).

Depending on the assumption imposed regarding the ability of firms to influence either inputs or outputs the distance function can be applied as input-oriented, that is reflecting a situation where the output set is fixed by exogenous factors, or as output-oriented which refers to the opposite situation. In this paper, an input oriented distance function is chosen. This is justified by the fact that the federal states and their PTAs can rather decide on the level of subsidies than on output levels such as train-km and passenger-km which are substantially influenced by long-term planning and contracts as well as by macro-economic and demographic factors. It should be noted that according to Coelli and Perelman (1999) output and input-oriented models estimate exactly the same frontier surface and therefore, identify the same set of DMUs as efficient, however, the scores themselves differ. They also show that for railways the orientation does not matter as much as it does for other industries.

The distance function can be estimated under alternative assumptions on the type of returns to scale; that is constant (CRS) versus variable (VRS). The input-oriented DEA model with constant returns to scale is (see Coelli *et al.*, 2005)

$$\begin{array}{ll}
 \text{Min wrt} & \Omega_i, \lambda: \Omega_i \\
 \text{st} & \sum_k \lambda_k y_k - y_i \geq 0, \quad k = 1, \dots, N \\
 & \Omega_i w_i - \sum_k \lambda_k w_k \geq 0, k = 1, \dots, N \\
 & \lambda_k \geq 0
 \end{array} \tag{1}$$

where λ_k represent the optimal weights for inputs and outputs which are searched for each DMU by the optimisation program. The value of Ω_i is the input-oriented technical efficiency

score for the i th DMU⁶. A VRS model is obtained by imposing the additional convexity constraint

$$\sum_k \lambda_k = 1. \quad (2)$$

The German rail sector is, as most rail industries in other countries, characterised by imperfect competition, budget restrictions and regulatory deficits. Therefore, the CRS assumption appears to be not appropriate. Furthermore, it should be borne in mind that this paper does not analyse a technical relationship between outputs (passenger-km and train-km) and inputs such as labour, material and capital but between outputs and monetary public subsidies as discussed at the beginning of this section. Therefore, this paper presents efficiency scores under variable returns to scale (VRS). A bootstrapping approach based on Simar and Wilsson (1998) is applied in order to obtain confidence intervals and to overcome the potential problem of biased results in the second-stage regressions. The bias-corrected results for the efficiency scores were obtained from 1000 iterations.

Table 1: DEA models in this paper

	V0	V1	V2
Inputs			
Operating subsidy	X	X	X
Investment subsidy	X	X	X
Infrastructure charges	X	X	X
% age of unpunctual trains			X
Outputs			
Train-km	X	X	X
Passenger-km	X	X	X
Punctuality		X	
Traveller satisfaction		X	X

For the analysis presented in this paper, four output variables and three input variables were defined. Based on common practice in studies of rail efficiency, this paper uses train-km and passenger-km as output variables. For QoS, the indicators punctuality of trains and traveller

⁶ Note that most DEA studies use the Farrell (1957) measure which is obtained by the reciprocal of the Shephard distance function.

satisfaction were considered, whereby the first represents an objective measure of QoS while the second is a measure of perceived QoS. The definition of inputs contains three monetary variables: operating subsidies, investment subsidies (both directly to be influenced by the DMUs) and infrastructure charges to be paid (exogenously given by the infrastructure provider). The paper presents results for three DEA models. All of them use train-km and passenger-km as output variables but differ regarding the specification and inclusion of QoS indicators (see table 1).

Following previous studies such as Merkert *et al.*, 2010, Cantos *et al.*, 2010, this paper regresses the efficiency scores obtained from the first-stage DEA approach against a set of independent explanatory variables in order to explain differences in the efficiency between the federal states. The explanatory variables contain two types of data. First, a set of policy variables which reflect the decisions of the PTAs how to organise rail franchising. To these variables belong the share of tendered train-km, the share of train-km under net contracts, the average size of contracts (expressed in train-km) and the average contract duration. Second, environmental variables which cannot be influenced by the DMUs such as the population density, car density and the rate of unemployment.

The efficiency scores are bound in the (0,1) interval and represent a censored variable which requires an appropriate econometric treatment by means of corner solution, that is Tobit models (named after Tobin, 1958). The general formulation of a one-limit Tobit model with 0 as limit is usually given as index function (or latent variable model)

$$\begin{aligned} y_i^* &= \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2) \quad , \\ y_i &= \max(0, y_i^*) \end{aligned} \quad (3)$$

where \mathbf{x}_i' represent the explanatory variables, $\boldsymbol{\beta}$ is the vector of parameters to be estimated, y_i is the observed dependent variable and y_i^* is the unobserved, latent variable. The log-likelihood for this model is a mixture of discrete and continuous distributions as follows

$$\log L = \sum_{y_i > 0} -\frac{1}{2} \left[\log(2\pi) + \log \sigma^2 + \frac{(y_i - \mathbf{x}_i' \boldsymbol{\beta})^2}{\sigma^2} \right] + \sum_{y_i = 0} \log \left[1 - \Phi \left(\frac{\mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right] \quad (4)$$

where the first term refers to the classical regression for the non-limit observations and the second term gives the relevant probabilities for the limit observations (see Greene, 2002). The Tobit model can be estimated by MLE procedures. Marginal effects are obtained by

$$\frac{\partial E(y_i|x_i)}{\partial x_i} = \Phi\left(\frac{x_i'\beta}{\sigma}\right)\beta \quad (5)$$

and indicate that a change in x_i' affects the conditional mean of y_i^* in the positive part of the distribution and in addition also the probability that the observation falls in that part of the distribution.

For the type of data used in this paper, the basic Tobit model has to be extended to a panel context. As common, the two standard approaches for incorporating individual heterogeneity are available, that is the fixed and random effects model. For incorporating individual heterogeneity, the basic Tobit model is modified to

$$\begin{aligned} y_{it}^* &= \alpha_i + x_{it}'\beta + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2) \\ y_{it} &= \max(0, y_{it}^*) \end{aligned} \quad (6)$$

where α_i represents unobserved individual heterogeneity. While the fixed effects model allows correlation between α_i and x_{it}' , the random effects model assumes that heterogeneity is uncorrelated with the regressors. The fixed effects model is estimated by including a set of dummy variables d_i , ($i = 1, \dots, N$) which indicate membership in group i :

$$\begin{aligned} y_{it}^* &= \sum_{i=1}^N d_{it} \alpha_i + x_{it}'\beta + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2) \\ y_{it} &= \max(0, y_{it}^*) \end{aligned} \quad (7)$$

The corresponding log likelihood function is given by

$$\begin{aligned} \ln L &= \sum_{i=1}^N \sum_{t=1}^T (1 - c_{it}) \ln \Phi(-\eta_i - x_{it}'\gamma) + c_{it} \ln[\theta \varphi y_{it} - \eta_i - x_{it}'\gamma] \\ c_{it} &= 1, \quad \text{if } y_{it} > 0 \\ c_{it} &= 0 \text{ otherwise} \end{aligned} \quad (8)$$

with $\eta_i = \alpha_i / \sigma$. Following Olsen's (1978) transformation, the parameters γ and θ are

$$[\gamma, \theta] = \left[\frac{\beta}{\gamma}, \frac{1}{\sigma} \right]. \quad (9)$$

The random effects model is

$$\begin{aligned} y_{it}^* &= \alpha + x'_{it}\beta + u_{it} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2) \\ y_{it} &= \max(0, y_{it}^*) \end{aligned} \quad (10)$$

Since in the random effects model the regressors are assumed to be uncorrelated with heterogeneity, a different approach to estimation is necessary. The conditional log likelihood in the presence of random effects is given by

$$\ln L_{cond} = \sum_{i=1}^N \ln \prod_{t=1}^T [\Phi(-\tau w_i - x'_{it}\gamma)]^{1-c_{it}} [\theta \varphi(\theta y_{it} - \tau w_i - x'_{it}\gamma)]^{c_{it}} \quad (11)$$

where $w_i = \sigma_\alpha / \sigma$, $w_i \sim [0, 1]$ and γ , θ and τ are unknown parameters to be estimated. Equation (11) includes the unobserved random effect and can thus not serve as the basis for estimation. The unconditional log likelihood is

$$\ln L = \sum_{i=1}^N \int_{-\infty}^{\infty} \prod_{t=1}^T [\Phi(-\tau w_i - x'_{it}\gamma)]^{1-c_{it}} [\theta \varphi(\theta y_{it} - \tau w_i - x'_{it}\gamma)]^{c_{it}} \varphi(w_i) dw_i \quad (12)$$

Finally it has to be noted that, following Greene (1993) the efficiency scores Ω_i were censored at zero by transforming

$$y_i = 1 - \Omega_i \quad (14)$$

in order to avoid concentrating of variables at unity. Both the DEA models and the Tobit regressions were performed in NLOGIT.

5. The data

Although the regionalisation subsidies are paid from tax money, in Germany centrally held and publicly accessible data at a desirable level of disaggregation for an econometric analysis (e.g. contract level, PTA level, level of federal states) is not available. Link (2016) has used a

unique database with the 16 German federal states as observation units which contained variables for train-km and passenger-km of regional rail services, subsidies and fare revenues, and data on the characteristics of contracts (net versus gross contracts, tendered versus non-tendered contracts, network size, contract duration). The data collection procedure, data sources and the necessary own calculations are described in Link (2016). For the purpose of this paper, this database was filled with new data on investment subsidies and quality indicators, was disaggregated to cover the 27 PTAs and extended to cover the period until 2014. All calculation procedures described in Link (2016) had to be applied to this new disaggregated dataset (in particular infrastructure charges, fare revenues). All monetary data was deflated to constant prices of 2010 by using the price deflator for regional rail services.

Data collection and data treatment for quality indicators such as train cancellations, punctuality and traveller satisfaction required assumptions for transforming available indicators to common definitions and scales. Punctuality data (in % of trains arriving on time) were provided by most PTAs, however, based on different definitions of delays (the minutes of delay accepted range from 3 to 6 minutes). While these different definitions might be considered as a problem, it can be argued that PTAs set QoS levels in the service contracts in correspondence to the regional situation where in one region a delay of 6 minutes might be acceptable (since there might be no risk of losing connections) and in other regions the acceptable delay might only be 3 minutes at maximum. The lack of comparability of traveller satisfaction indicators collected by PTAs was a more serious problem due to different dimensions of satisfaction considered (for example: satisfaction with trains, with stations, with staff, with information etc.) and different scales (ranging from 1-5, 1-6, 1-100). Therefore, an alternative source - the overall index of traveller satisfaction with regional rail services collected by the survey institute INFAS - was used and transformed to a 1-5 scale. Since in some cases this data covers broader regions than the PTAs, the index had to be further disaggregated by using the ratios between PTAs satisfaction index and the INFAS satisfaction index. Data on train cancellations were available for some PTAs too, however, the time periods covered were too different and did not match with the available time periods for the other QoS indicators.

Table 2: Descriptive statistics – for 22 PTAs and 28 PTAs 2003-2014

	Variable	Unit	Mean	Standard deviation	Minimum	Maximum
Population density	P_SKM	inhabitants/skm	564.7 (541.9)	859.1 (797.3)	68.8 (68.8)	3881.1 (3881.1)
Train-km	TR_KM	million km	25.97 (22.63)	24.33 (22.71)	1.81 (1.81)	123.3 (123.3)
Passenger-km	P_KM	million km	1892.5 (1585.0)	1999.1 (1875.7)	61.9 (61.9)	8902.0 (8902.0)
Punctuality	PUNCT	%	92.3 n.a.	3.8 n.a.	76.9 n.a.	99.5 n.a.
Traveller satisfaction	SAT		2.57 n.a.	0.44 n.a.	4.32 n.a.	3.62 n.a.
Share of tendered train-km	TEND	%	22.5 (23.4)	24.8 (25.5)	0.0 (0.0)	100.0 (100.0)
Share of train-km under net contracts	NET	%	81.2 (80.4)	30.3 (31.7)	0.0 (0.0)	100.0 (100.0)
Average contract size	SIZE	million train-km p.a.	2.98 (2.68)	2.09 (1.99)	0.52 (0.42)	10.14 (10.14)
Average contract duration	DUR	years	11.9 (11.8)	2.9 (2.8)	5.9 (4.5)	18.2 (18.2)
Operation subsidies*	S_op	million €	91.33 (80.25)	84.01 (78.37)	3.18 (3.18)	463.93 (463.92)
Investment subsidies	S_inv	million €	24.84 (21.94)	33.07 (30.14)	0.00 (0.00)	254.68 (254.68)
Share of infrastructure charges** in total subsidies	INFRA	%	56.7 (56.8)	8.9 (9.2)	35.4 (35.4)	80.6 (82.6)

Figures for 28 PTAs in brackets.- N=264 for dataset with 22 PTAs, N=336 for dataset with 28 PTAs.- . *) At 2010 prices. Adapted by fare revenues from gross contracts.- **)Infrastructure charges paid for the use of tracks and stations represent items which are paid by rail operators to rail rack providers (typically DB Netz). They are passed through to PTAs and are part of the subsidies paid by PTAs to operators within franchise contracts.

The dataset finally used for this paper covered the period from 2003 to 2014. The decision of not using the data for the full period since 1996 was taken for two reasons. First, QoS indicators were only available since 2003. Second, the period since 2003 is of special interest since in 2003 PTAs have closed large long-term RRPS contracts with DB within direct negotiations due to size problems and the risk of having no sufficient bidders for such large contracts. However, all of these long-term contracts contain a time path of reducing train-km

within this contracts and tendering them stepwise in order to achieve a smooth transfer to competitive tendering and to guarantee a sufficient number of bids.

Since not for all PTAs QoS indicators were available, the final dataset covered 22 out of 27 PTAs. The PTA Berlin-Brandenburg was finally split and treated as two PTAs since the greater Berlin area and the Brandenburg area are different federal states and show large differences regarding population density, and in particular regarding the QoS indicators where the Berlin S-Bahn crisis with very low punctuality figures and low customer satisfaction would overlay and distort the figures for the Brandenburg area.

Table 2 shows the descriptive statistics for the relevant variables for all PTAs (where 28 PTAs instead of 27 are shown due to the split up of the PTA for Berlin/Brandenburg) and for the final dataset with 22 PTAs (where again the PTA for Berlin/Brandenburg was split up). The differences in variable means between the two datasets (full and 22 PTAs) are not large and thus it seems feasible to proceed with the reduced 22 PTAs dataset. However, some caution is necessary when interpreting disaggregated figures in the previous section, for example for the co-called city states because one of the three city state PTAs (Bremen) is not included (lack of QoS variables).

Table 3: DEA results

	VRS				CRS			
	ES		ES– bias corrected		ES		ES – bias corrected	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
V0	0.8924	0.0808	0.8722	0.1233	0.8391	0.0888	0.8443	0.0930
V1	0.9189	0.0789	0.9189	0.0789	0.8459	0.0899	0.8488	0.0859
V2	0.9235	0.0753	0.9283	0.0720	0.9083	0.0800	0.9103	0.0788

V0: Base model - DEA model with two outputs (train-km and passenger-km).- V1: DEA model with four outputs (train-km, passenger-km, punctuality, traveler satisfaction).

6. Estimation results

Table 3 summarises the estimated and the bias-corrected scores for technical efficiency obtained with the three DEA models under the assumptions of variable and constant returns to scale. As to be expected, the average bias-corrected technical efficiency scores are higher for models V1 and V2 where two QoS variables are included since the number of outputs

increase by two and by one respectively. The slightly higher efficiency scores for model V2 are somewhat surprising since here only one additional output is considered and at the same time one input is added. The overall level of DEA scores is in line with previous literature on rail efficiency (see Growitsch and Wetzel, 2009; Merkert *et al.*, 2010, Cantos *et al.*, 2010; Link, 2016).

Table 4: Disaggregated efficiency scores

	DEA-V0		DEA-V1		DEA-V2	
	ES	SD	ES	SD	ES	SD
Hamburg, Berlin	0.9219	0.0714	0.9351	0.0674	0.9379	0.0628
Other PTAs	0.8672	0.1262	0.9175	0.0798	0.9221	0.0762
Out of these:						
East German PTAs	0.7636	0.1437	0.8773	0.0866	0.8855	0.0858
West German PTAs	0.9129	0.0841	0.9339	0.0696	0.9378	0.0653

V0: Base model - DEA model with two outputs (train-km and passenger-km).- V1: DEA model with four outputs (train-km, passenger-km, punctuality, traveler satisfaction.) - V2: DAE model with three outputs (train-km, passenger-km, traveler satisfaction) and punctuality as bad input.-All figures are based on VRS assumption and are bias-corrected scores obtained from 1000 bootstrapping iterations.

Based on the preferred specification of variable returns to scale and without considering QoS variables (base model V0) the results suggest that the average PTA could save subsidies in a range of around 13 per cent, when considering QoS variables this saving potential reduces at 8%. Generally, these potential savings should be interpreted with caution. First of all, including QoS variables increases the number of outputs and more DMUs are deemed efficient. Second and more general, the saving potential has to be interpreted as the maximum possible saving if all PTAs would face identical conditions (population density, level of infrastructure charges to be paid etc.) as the most efficient PTAs which is not the case (see the considerable variation of indicators in table 2). The disaggregated technical efficiency scores for different types of PTAs in table 4 underline this: First, the disaggregated efficiency scores for the city states⁷ (the PTAs of Hamburg and Berlin) are higher than for the remaining PTAs. This is an expected effect because the city states represent large urban areas with high population density allowing a bundling of rail traffic and interchanges with other public transport which makes regional rail traffic attractive. Second, similar to the findings in Link (2016) the PTAs in the Western part of Germany are more efficient in using rail subsidies than the East German PTAs. Interestingly, these East-West differences are still persistent in

⁷ The PTA for Bremen, the third of the three so-called German city states had to be excluded from the analysis since QoS variables were only available for the last three years.

the period from 2003-20014. Amongst the potential reasons for such differences is the transformation process with the related economic problems in East Germany and the lower population density. Furthermore, the East German PTAs had during the major part of the observation period to pay higher infrastructure charges than most of the West German PTAs due to the regional surcharges levied by DB Netz until 2011 for low-density lines in particular in East Germany.

Table 5: Rankings of PTAs

PTA number	Model V0	Model V1	Model V2	$\Delta V1, V0$	$\Delta V2, V0$
1	20	19	18	-1	-2
2	6	8	9	2	3
3	22	22	22	0	0
4	19	16	16	-3	-3
5	18	15	14	-3	-4
6	17	20	20	3	3
7	11	14	15	3	4
8	8	11	12	3	4
9	7	10	11	3	4
10	16	17	17	1	1
11	14	9	7	-5	-7
12	15	21	21	6	6
13	21	18	19	-3	-2
14	4	3	3	-1	-1
15	5	5	5	0	0
16	1	1	1	0	0
17	9	13	13	4	4
18	3	4	4	1	1
19	10	6	6	-4	-4
20	2	2	2	0	0
21	12	12	10	0	-2
22	13	7	8	-6	-5

The inclusion of QoS variables had two effects. First, the difference between city states and remaining PTAs and between Eastern and Western PTAs decrease considerably. Apparently, including punctuality and traveller satisfaction introduces an important dimension into efficiency analysis, and punctuality problems but also the loss of customer satisfaction due to unreliable and crowded trains in the large conurbations and also in some Western PTAs change efficiency scores. Second, the rankings of PTAs change (table 5). In model V1 only

five and in model V2 only four out of 22 PTAs keep their rank from the basic model while for remaining PTAs the ranking changes partly considerably. In particular East German PTAs improve their rankings due to better punctuality and higher customer satisfaction. This all strengthens the argument that QoS indicators should be included into efficiency and benchmarking studies.

Table 6 reports the results of the Tobit panel models which use the efficiency scores from the three DEA models with and without QoS variables under the VRS assumption as dependent variable. Since only a random effects model has produced significant and plausible estimates, this type of Tobit panel model is the preferred one and reported here. The scale factors for obtaining marginal effects are given at the bottom of tables 4 and 5 and are near unity, indicating that the coefficients can be interpreted as almost equal to the marginal effects (evaluated at the mean of the exogenous variables).

Table 6: Results of the second stage Tobit regressions – random effects models

Variable	Random effects – basic model		Random effects – model V1				Random effects – model V2		
	coefficient	Std error	coefficient	Std error	P(z)>Z	P(z)>Z	coefficient	Std error	P(z)>Z
P_SKM	-0.00011***	0.00006	0.0000	-0.00039***	0.00006	0.0000	-0.00038*	0.00002	0.0763
DUR	-0.00229**	0.00115	0.0473	-0.00515***	0.00166	0.0019	-0.00450**	0.00217	0.0382
NET	0.02273	0.02170	0.2950	0.09914***	0.02206	0.0000	0.08436***	0.02925	0.0038
TEND	- 0.17421***	0.02128	0.0000	-0.12049***	0.02478	0.0000	- 0.12298***	0.02961	0.0000
SIZE	0.02890***	0.00243	0.0000	0.01394***	0.00210	0.0014	0.01165***	0.00209	0.0000
Const	0.18103***	0.02792	0.0000	0.08180**	0.03389	0.0158	0.08038*	0.04646	0.0836
Sigma(v)	0.10155***	0.00591	0.0000	0.05558***	0.00188	0.0000	0.05282***	0.00247	0.0000
Sigma(u)	0.09436***	0.01948	0.0000	0.0712***	0.00622	0.0000	0.06007***	0.00597	0.0000
Scale factor			0.9903			0.9986			0.9371
LogL			203.4			203.3			212.1
*** significant at 1% level.- **significant at 5% level.									

Since the dependent variable was obtained from the transformed efficiency scores, negative signs of the parameters indicate that the respective variables have a positive impact on efficiency. In both the models with and without considering QoS, all coefficients (except *NET*

in the base model) are significant at 1per cent level and have the same sign⁸. However, the sizes of coefficients differ considerably between the three models.

Overall, the estimation results are intuitively plausible and confirm expectations from theory. They are also in line with the findings for the German federal states during the period from 1996 to 2010 reported in Link (2016). As reported there, a higher share of tendered train-km, a lower share of train-km operated under net contracts, a longer contract duration and a smaller contract size explain higher efficiency scores. When using DEA scores obtained from models with QoS variables the size of coefficients changes while keeping the sign and thus the general direction of explanation. Furthermore, the approach of including QoS variables (model V1 versus model V2) does not lead to larger deviations: The coefficients in both model V1 and V2 are in a similar range. In both model V1 and V2 the coefficients for *SIZE* and for *TEND* are smaller than those obtained with the basic model (around half and around two thirds respectively, and the coefficients for *DURATION* and *NET* are higher (doubling and even quadrupling in size). This suggests that the efficiency enhancing effect of tendering is smaller when considering QoS variables but contractual features are gaining in importance. The efficiency enhancing effect of a longer contract duration and of a lower share of net contracts is strengthened in both models with QoS variables while the efficiency enhancing effect of a smaller contract size is lower. These results indicate that QoS is rather driven by contractual arrangements although competitive tendering keeps its positive effect on efficiency.

As in Link (2016) the negative impact of net contracts on efficiency in both the model with and without consideration of QoS seems to be counterintuitive at a first glance. Net contracts would be expected to give incentives to the operator to increase patronage and, in order to do so, also improve quality of service. An explanation might be the special German situation where in 2003 all PTAs closed large long-term contracts with DB within direct negotiations as net contracts. This might have led to a situation where DB was able to receive rather high net subsidies, which is reflected in lower efficiency scores. PTAs had finally to spend more public money for net contracts than for gross contracts for which the fare revenues collected by the PTA reduce the subsidy level - even though the net subsidy per train-km is per definition lower than the gross subsidy when not considering fare revenues.

⁸ Models which included the variables car density suffered from collinearity problems and therefore these variables were removed from the models.

As expected and in line with Link (2016), a higher population density has a positive effect on the efficient use of subsidies. Further variables (dummies for urban areas and for East German PTAs) were not significant and are not reported here.

7. Conclusion

This paper has for the first time provided an efficiency analysis for subsidising regional rail services in Germany which combines conventional input and output measures with QoS variables. It considers a subsample of 22 out of 27 PTAs over the period from 2003 to 2014. As in any econometric application, the results naturally reflect the specific empirical situation under study, in this paper the German rail franchising during a period which was characterised by the closure of large long-term contracts with DB without competitive tendering and a stepwise tendering of parts of these services in subsequent years. Bearing this note of caution in mind, the following general policy conclusions can be drawn which conform those obtained earlier in Link (2016): First, the findings suggest that a higher share of tendered rail services would significantly increase the efficiency in using the available public funds to provide these services. Second, the paper provides evidence that a higher share of gross contracts has a positive impact on efficiency. One potential reason for this is the fact that the majority of net contracts was awarded without any competitive tendering, and the effect could therefore be overlapped with the effect of lacking competitive pressure.

Third, the analysis suggests that contract duration, even though longer when compared with the period from 1996-2010, should be longer than currently (on average 11.9 years) to guarantee a more efficient use of public funds, a finding which confirms ongoing debates and critiques in Germany on too short contract periods for amortising rolling stock, in particular when keeping the problems since the financial crises in mind. Furthermore, smaller contracts than in the period of analysis (on average 2.3 million train-km p.a.) would potentially increase efficiency.

The most important finding is that although including quality of service into the analysis changes the ranking of DMus partly considerably, the general explanatory factors for efficiency remain the same. However, while keeping their sign and thus the direction of impact, their absolute size (and thus their weight of impact) changes. Apparently, rather

contractual features gain in importance while keeping the positive impact of competitive tendering. These findings are robust regarding the approach of taking into account quality of service. However, more research is needed on QoS, in particular a systematic approach which quality indicators should be considered and how they should be included into the models. Finally, for future research it would be of interest to study how an increase of efficiency in the sense of reducing subsidies and maintaining rail services is reflected in an increase of companies' efficiency in using their factor inputs. With new data collection in future this could be tested and serve as a complementary evidence of the effect of rail franchising.

References

- Affuso, L. and Newbery, D. (2002): 'The impact of structural and contractual arrangements on a vertically separated railway', *The Economic and Social Review*, Vol. 33, No.1, Spring, 83-92.
- Ajodhia, V., Petrov, K., Scarsi, G.C. (2004): 'Economic benchmarking and its Applications, Paper presented at the 3rd International Conference on Applied Infrastructure Research, 9 October, rechnical University of Berlin, Germany.
- Assaf, A.G., Josiassen, A., Gillen, D.. (2014): 'Measuring firm performance: Bayesian estimates with good and bad outputs', *Journal of Business research*, Vol. 67, pp. 1249-1256.
- Beck, A. (2011): 'Barriers to entry in rail passenger services: Empirical evidence for tendering procedures in Germany', *European Journal of Transport and Infrastructure Research* 11(1), 20-41.
- Bitzan, J (2003): 'Railroad costs and competition', *Journal of Transport Economics and Policy*. 37 (2) 201-225.
- Cantos P., Pastor, J.M., and Serrano, L. (1999): ' Productivity, efficiency and technical change in the European Railways: a non-parametric approach', *Transportation*, vol. 26, 337-357.
- Cantos P., Pastor, J.M., and Serrano, L. (2010): 'Vertical and horizontal separation in the European railway sector and its effects on productivity', *Journal of Transport Economics and Policy*, 44 (2), 139-160.
- CEPA (2003): Background to Work on Assessing Efficiency for the 2005 Distribution Price Control Review. Scoping study, Final report prepared for OFGEM, September 2003.
- Coelli, T. J., Rao, D.S.P., O'Donnell, C.J. and Battese, G.E (2005): *An introduction to efficiency and productivity analysis*. 2nd ed.
- Coelli, T. and Perelman, S. (1999): 'A comparison of parametric and non-parametric distance functions: With application to European railways', *European Journal of Operational Research* 117 (1999), 326-339.
- De Borger, B. and Kerstens, K. (1996): 'Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches'. *Regional Science and Urban Economics* 26(1996), 145-170.

- Driessen, G., Liejesen, M. and Mulder, M. (2006): *'The impact of competition on productive efficiency in European railways'*, CPB Discussion Paper no. 71, The Hague.
- Farell, M.J. (1957): 'The measurement of Productive Efficiency', *Journal of Royal Statistical Society*, Series A, 120(3), 253-281.
- Friebel, G.; Ivaldi, M, and Vibes, C. (2010): 'Railway (De)Regulation: A European efficiency comparison', *Economica* [Volume 77, Issue 305](#), 77–91.
- Giannakis, D., Jamasb, T., Pollit, M. (2005): 'Benchmarking and incentive regulation of quality of service: an application to the UK electricity distribution networks', *Energy Polivy*, Vol. 33, pp. 2256-2271.
- Greene, W.H. (1993): The econometric approach to efficiency analysis, in Fried, H., Lovel, C.A.K. and Schmidt, S. S. (eds.) *The measurement of productive efficiency and application*, New York , Oxford University Press.
- Greene, W. H. (2002): *Econometric Analysis*. Prentice Hall.
- Greene, W. H. (2008): *Limited Dependent Variables – Truncation, censoring and sample selection*. <http://people.stern.nyu.edu/wgreene/Lugano2013/Greene-Chapter-19.pdf>, accessed at 20 April 2014.
- Growitsch, C. and Wetzel, H. (2009): 'Testing for economies of scope in European railways: an efficiency analysis', *Journal of Transport Economics and Policy*, 43 (1), 1-24.
- Growitsch, C., Jamasb, T., Pollit, M. (2005): 'Quality of Service, Efficiency and Scale in Network Industries: An Analysis of European Electricity Distribution', Discussion Paper No. 3, Halle Institute for Economic Research.
- Hunold, M., Wolf, C. (2012): *Competitive procurement design: Evidence from regional passenger railway services in Germany*. Paper presented at the Kuhmo Nectar Conference on Transport Economics, Berlin.
- Ivaldi, M and McCullough G J (2001): 'Density and integration effects on class 1 U S freight railroads', *Journal of Regulatory Economics*, 19, 161-182.
- Jensen, A. Stelling, P. (2007): 'Economic impacts of Swedish railway deregulation: A longitudinal study', *Transportation Research E*, Vol 43, No. 5, 516-534.
- Korhonen, P.J., Syrjänen, M.J. (2003): 'Evaluation of Cost Efficiency in Finnish Electricity Distribution', *Annals of Operations Research*, Vol. 121, pp. 105-122.
- Lalive, R., Schmutzler, A. (2008): 'Entry in liberalised railway markets: The German experience', *Review of Network Economics*, 7 (1), 37-52.
- Lalive, R., Schmutzler, A. (2011): *Auctions versus negotiations in public procurement: which works better?* CEPR discussion papers, No. 8538.
- Link, H., Merkert, R. (2011). Success factors and pitfalls of the German rail franchising approach. *International Journal of Transport Economics*, XXXVIII(2), pp. 171-198.
- Link (2016). A Two-stage Efficiency Analysis of Rail Passenger Franchising in Germany. *Journal of Transport Economics and Policy* 50, part 1, pp. 76-92.
- Merkert, R, Smith A and Nash C.A. (2010): 'Benchmarking of train operating firms – a transaction cost efficiency analysis', *Transportation Planning and Technology*, 33:1, 35-53.
- Merkert, R., Assaf, A.G. (2015): 'Using DEA models to jointly estimate service quality perception and profitability – Evidence from international airports', *Transportation Research Part A*, Vol. 75, pp. 42-50.

- Mituzani, F. and Uranishi, S. (2010): ‘Does vertical separation reduce cost? An empirical analysis of the rail industry in OECD countries’, Discussion paper 2010-48, Kobe University.
- Mouwen, A., Rietveld, P. (2013): ‘Does competitive tendering improve customer satisfaction with public transport? A case study for the Netherlands’, *Transportation Research Part A*, Vol. 51, pp. 29-45.
- Nash, C. A., Nilsson, J.E., Link, H. (2013). Comparing three models for introduction of competition into railways. *Journal of Transport Economics and Policy*, Vol. 47, Part 2, May 2013, pp. 191-206..
- Olsen, R. (1978): ‘A note on the uniqueness of the maximum likelihood estimator in the Tobit model’, *Econometrica* 46,1211-1215.
- Paha, J., Rompf, D., Warnecke, C. (2013). Customer choice patterns in passenger rail competition- [Transportation Research Part A: Policy and Practice](#), Vol. 50, April 2013, pp. 209–227.
- Simar, L. and Wilson, P.W. (1998): ‘Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric Frontier models.’ *Management Science* 44(19), 49-61.
- Simar, L. and Wilson, P.W. (2000): ‘A general Methodology for Bootstrapping in Non-parametric Frontier Models’, *Journal of Applied Statistics*, 27 (6), 779-802.
- Tobin, J. (1958): ‘Estimation of relationships for limited dependent variables’, *Econometrica* 26 (1), 24–36.
- Woldridge, J. M. (2010): *An econometric analysis of cross section and panel data*. The MIT Press. Cambridge, Massachusetts.
- Zhang, C., Juan, Z., Lu, W., Xiao, G. (2016): ‘Do the organisational form affect passenger satisfaction? Evidence from Chinese public transport service’, *Transportation Research Part A*, Vol. 94, pp. 129-148.