Efficiency of urban public transit: A meta analysis

Brons, M.R.E.; Nijkamp, P.; Pels, E.; Rietveld, P.

published in
Transportation
2005

DOI (link to publisher)
10.1007/s11116-004-0939-4

document version
Publisher's PDF, also known as Version of record

Transportation (2005) 32: 1–21

Efficiency of urban public transit: A meta analysis

MARTIJN BRONS*, PETER NIJKAMP, ERIC PELS & PIET RIETVELD
Department of Spatial Economics, Free University, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
*(*Author for correspondence, E-mail: mbrons@feweb.vu.nl)*

Key words: efficiency, meta-analysis, urban transit

Abstract. The aim of this paper is twofold. First, to provide a statistical overview of the literature on public transit efficiency performance. Second, to statistically explain the variation in efficiency findings reported in the literature. To this end, first some key concepts of efficiency analysis will be introduced, while next the different frontier methodologies that are used in the literature will be discussed. The empirical part of this paper consists of a statistical summary of the literature as well as meta-regression analyses for different samples of the literature in order to identify key determinants of technical efficiency (TE) of public transit operators. For a broad sample of observations, we found significant and consistent effects of the type of database, region and output measurement method. For the sample of non-parametric studies we found that the type of frontier assumptions also have an impact on the efficiency ratio. Further results show that there is no statistical difference in TE ratio's between parametric and non-parametric studies. Finally, we found a positive univariate relationship between the number of inputs in the estimated specification and the efficiency ratio.

1. Introduction

Urban transit, including bus, ferries, trams, light rail and metros, makes up a major part of the transportation network in an economy. Although travel patterns in most developed countries are increasingly dependent on the car (see Banister 2000), causing a declining trend in transit demand in most industrial economies, urban transit remains an important transport mode. Urban transit services are provided by public, private or mixed companies in a highly regulated environment. Moreover, important constituents of the transportation infrastructure are essentially (semi-) public goods. There are sound economic reasons for a significant degree of state intervention in this field, based mainly on the recognition of a variety of market failures (see, for example, Kerstens 1996). In the past two decades, however, serious concerns about possible regulatory failures have led to a reassessment of the role of the state in the organization of the sector (see Glaister et al. 1990).

In view of these concerns it is of great interest to investigate whether urban transit operators work in a technically efficient way (i.e. reach economic targets such as cost minimization or output maximization conditional on output or input constraints). Solid technical efficiency (TE) measurement can provide a significant contribution to the discussion on the relative merits of private versus public provision of transportation services. From the early 1980s onward various frontier estimation techniques have been developed to determine best practice behaviour in an industry. Frontier methodologies allow for distinguishing between efficient and inefficient production and the estimation of the degree of (in)efficiency. In the transportation literature, frontier methods have been used in efficiency studies on almost all transport modes. A comprehensive survey of frontier methodologies and empirical results for urban public transport has recently been published (De Borger et al. 2002). While this survey provides a good overview of the literature on public transit performance, deeper insights can be obtained by using quantitative-statistical research techniques.

The aim of this paper is to present a statistical overview of the literature on public transit efficiency, and to give a statistical explanation for the variation in TE findings reported in the literature.

In the next section the concepts of TE and efficiency frontiers are introduced. In Section 3, the different frontier specification techniques that are encountered in the literature will be discussed. Section 4 will discuss the feasibility of comparing TE studies in a meta-analytical set-up and will address some of the assumptions underlying this paper. Section 5 consists of a statistical exploration of the methodologies and results that are found in the literature. In Section 6 some meta-regression analyses will be performed in order to identify determinants that may help explain the variation in efficiency results that are reported in the literature. Section 7 will conclude with a brief summary and conclusions.

2. TE and cost or production frontiers

Economists have traditionally distinguished between two sorts of efficiency: TE and allocative efficiency (Viton 1986).[1] In this paper we will focus on TE. This concept relates to the divergence between actual production and production on the boundary of the feasible production set.[2] This set summarizes all technological possibilities of transforming inputs into outputs that are available to the organization. A producer is technically inefficient if production occurs within the interior of this production set. Technical inefficiency can be viewed from two perspectives (Viton 1997). Input-oriented technical inefficiency focuses on the possibility of reducing inputs to produce a given output level. This concept is illustrated for the two input and one output case in Figure 1a.

In Figure 1a, the area above curve y represents the set of feasible input bundles (x_1, x_2) that produce a given output level y. The curve itself represents

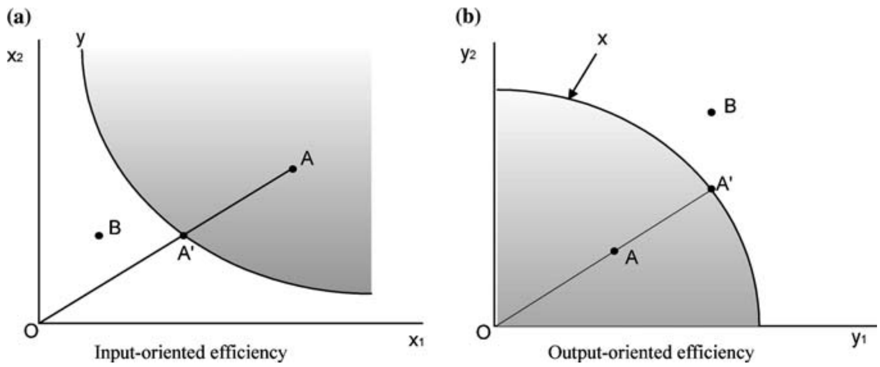

Figure 1. Input- and output-oriented TE.

the set of efficient input bundles (the efficiency frontier). Only points on the curve are technically efficient. Point A is technically inefficient. Point B is not a feasible input vector given the planned output level for this production technology. A simple method to measure the degree of TE is to calculate the scalar (OA′)/(OA). This results of course in a ratio between 0 and 1.

Second, output-oriented technical inefficiency implies the possibility of increasing the output bundle given a fixed input level. This is illustrated in panel b for the two outputs and one input case. The area below curve x represents the set of feasible output bundles (y_1, y_2). The curve itself represents the set of efficient output bundles (the efficiency frontier). Observation points on this frontier are efficient. Point A is not technically efficient. Output bundle B is not feasible given the input level. The degree of inefficiency can be expressed by the ratio (OA)/(OA′).

Since the degree of TE can only be measured in relation to 'best practice', the efficiency frontier, i.e. the range of values showing the maximum output (minimum input) for a given input (output), must first be constructed. This is the subject matter of the next section.

3. Production and cost frontier methodologies

Efficiency frontiers can be determined by and addressed in an engineering context based on empirical knowledge of technical production operations (parametric specification). This can be done using either a deterministic or a stochastic approach. Furthermore, they can be estimated by observing production operations actually accomplished (non-parametric specification).

The deterministic parametric frontier method assumes a particular functional form for the efficiency boundary. A procedure that is called corrected OLS (COLS)[3] is used to measure technical inefficiency. First an average practice frontier is estimated using OLS. This frontier is corrected by shifting the intercept until all residuals except one become negative. The remaining non-negative residual should be equal to zero. Technical output efficiency is given by the ratio of the observation output value to the fitted frontier output value. Alternatively, one can directly estimate a frontier that envelops all data points by specifying an error term which has a distribution that is truncated at zero, so that there again will be at least one point which coincides with the frontier.

The stochastic parametric frontiers method is similar to the deterministic method, but allows for measurement error in the frontier. The error term therefore consists of two elements: A technical inefficiency component (deviation from the frontier) and a random error term with zero mean (measurement error of the frontier). Table 1 shows a categorization of frontier methodologies.

Deterministic non-parametric methods do not assume a particular production function. The piece-wise linear frontier is directly constructed from the observations themselves by applying mathematical programming techniques. Two main methodologies can be distinguished: (i) the Data Envelopment Analysis (DEA) method and (ii) the Free Disposal Hull (FDH) method (see Kerstens 1996).

DEA modelling was initiated by Charnes et al. (1978). The DEA methodology provides relative measures of efficiency and is increasingly being used in evaluating the performance of public service industries (for an overview we refer to Ganley and Cubbin 1992). The efficiency measures are distances to an empirical production frontier and the values are calculated on the basis of standard Pareto efficiency. No assumption has to be made about the production frontier's functional form, since the frontier is the observed best practice of the raw data set available. The frontier is constructed based on the assumption that any linear combination of observation units is feasible and on the assumption of strong input and output disposability.

Table 1. Categorization of frontier methodologies.

	Deterministic specification	Stochastic specification
Parametric technology	Frontiers based on (corrected) OLS or ML models.	Frontiers with explicit distributional assumptions for TE values
Non-parametric technology	Frontiers based on Free Disposal Hull (FDH) and Data Envelopment Analysis (DEA) techniques.	Resampling, chance constrained programming.

Source: Kerstens (1996).

Strong input disposability means that a feasible output level remains feasible after increasing any input levels. Strong output disposability means that it is always possible to reduce the output level without changing input levels. The variable returns to scale DEA model is illustrated for output maximization in Figure 2a. Points A–E are observational units. The model assumes that an observed output vector can be smaller than the linear combination of observations D and E. All observations to the south-west of the line segment D–E are therefore feasible. This explains the line originating in observation E and extending parallel to the horizontal axis.[4] Using a similar reasoning for all line segments and allowing for all linear combinations yields the set of possible output combinations bounded by the production frontier that consists of the line segments A–B–D–E. Observation point C and F are not efficient according to DEA assumptions. Hence, they are not on the frontier. The degree of technical inefficiency for points C and F is measured by the fraction (OC)/(OC′) and (OF)/(OF′), respectively. Strong *input* disposability and the construction of a cost frontier can be illustrated analogously.

In FDH models (initiated by Deprins et al. 1984) the efficiency frontier is constructed based on the assumptions of strong input disposability and output disposability but without allowing for linear combinations of observational units. Based on these assumptions the FDH frontier typically has a staircase form, as illustrated in panel b. Note that the set of observation points is identical as that in panel a. The set of feasible outputs is bounded by the production frontier that consists of the line segments A–B–C–D–E. Thus, point C, which was inefficient in the DEA model, is efficient in the FDH model. Point F remains inefficient under FDH assumptions. The degree of inefficiency of point F is measured by the ratio (OF)/(OF′). Since FDH models relax the assumptions of DEA models the observational units are on average located closer to the frontier.[5] All other things being equal, TE ratios are thus expected

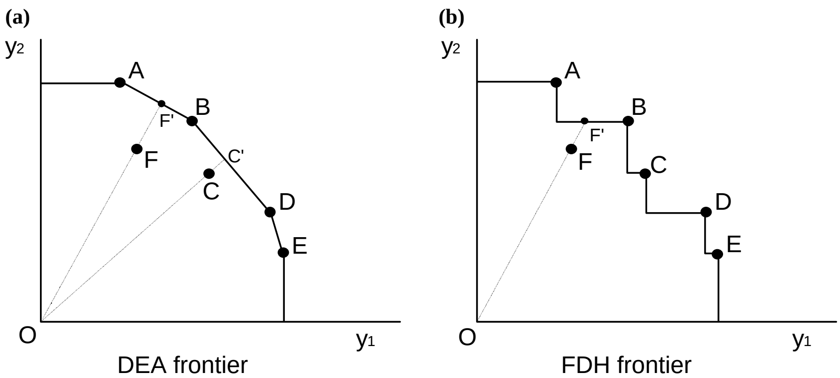

Figure 2. Frontiers of DEA and FDH technology.

to be lower for studies using DEA modelling than for studies using FDH modelling.[6] Strong input disposability and the construction of a cost frontier can be illustrated in an analogous manner.

Stochastic non-parametric frontiers have to our knowledge not yet been used in the literature on public transit efficiency performance.

4. Study comparability

The case studies underlying this meta-analysis report mean TE values; i.e. the mean of a sample of observed TE for individual firms (cross-section data) or time periods (time series data) or both. The reported means are then used as observations in the meta-analytical database. Note that one underlying study can yield multiple observations.[7]

An important issue is related to the fact that TE is a relative measure. The reference point for measuring TE is a case specific efficiency frontier, which is determined based on the actual observation points within the sample. Therefore, and this is the essence of the matter, the efficiency frontier itself may not be fully efficient.

Under the theoretical assumption that there is some efficiency frontier with universal validity, it is not (in)efficiency values that we are comparing but rather sample heterogeneity. The issue is illustrated in Figure 3 for the one-input-one-output case under parametric assumptions. The homogeneous sample A will yield a higher mean TE value than the heterogeneous sample B.

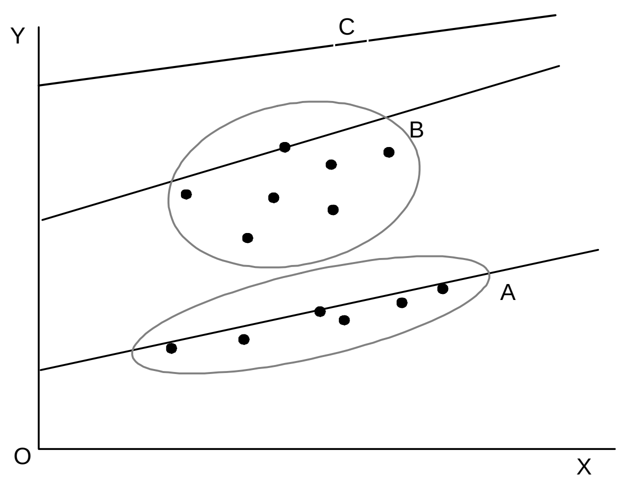

Figure 3. Comparison of mean TE values from different studies.

Yet, if we assume that there is an overall efficiency frontier, represented by C, it is obvious that the real TE is higher for sample B than for sample A.[8]

Alternatively, under the assumption that the study specific efficiency frontiers indeed represent the maximum attainable efficiency relations given the study-specific conditions we can interpret the comparison of mean TE values as the comparison of actual efficiency scores.[9]

Obviously, neither of the two assumptions will hold in practice. Firms that are 'efficient' within a specific sample might still be able to do better given the environmental conditions.[10] On the other hand, the observed TE values of the sample *do* provide some indication of the maximum attainable efficiency under the case specific conditions. Comparing reported TE values will therefore mean comparing values that consist of a heterogeneity component and a TE component.

There are several ways to deal with this problem. One could assume revealed optimal efficiency (i.e. the observed frontier is the actual frontier). Obviously, this is a very strong assumption. A weaker assumption might be based on the notion that the observed efficiency values give some indication of optimal efficiency given the circumstances (i.e. the observed frontier is related to the actual frontier). The assumption then is that, there is a fixed linear relation between the observed and the actual frontier that is identical for all observations. In other words, the possibilities for improving efficiency are similar for all best practice firms within a sample. Clearly, the strong assumption of revealed optimal efficiency is a special case of the weaker assumption of a fixed linear relation between revealed and optimal efficiency, viz. the case where constant and coefficients are zero.

The previous discussion suggests that, although comparing average TE values between studies may lead to biased results, comparing average TE values that come from the same study will prove less problematic.[11] Therefore, in a meta-regression context where the aim is to explain the variation in TE values by the variation in a number of moderator variables, the specification of the econometric model in such a way that only the within study variation is used in order to explain the variation in the TE estimations is another way to deal with the problem of study comparability. This can be done by the use of dummy variables for different studies. Obviously, the efficacy and feasibility of such a solution depends on the average number of observations per study.

A third way to circumvent the comparability problem involves a re-interpretation of the dependent variable. Although the actual TE frontier is not observed, the average TE value as it is measured does however provide an indication of the relative variation in TE values, and thus of the possibilities to improve TE.[12] If the dependent variable is interpreted as such then comparison between studies in a meta-analytical format is indeed valid. The requirement that the individual firm with the best efficiency within a study have a TE value of 1 is not necessary to be able to interpret individual TE values as an indicator of relative efficiency within a study.

Finally, assuming that there is indeed some universal efficiency frontier one could transform the study specific efficiency frontiers in such a way that they become directly comparable. Theoretically, this could be possible for parametric frontier studies if one has sufficient knowledge about the intercept and slope of the input–output curve (in the absence of such data information on input and output levels of all the firms in the original studies is needed). In practice such data is missing for a great number of studies. Moreover, such transformation calculations become increasingly complicated when the number of input and output goods increases and even more so when assumptions about the functional form of the production technology differ among studies. For non-parametric studies such transformations are generally not possible in the first place because the frontier is not expressed in terms of a functional form.

Since the database underlying this paper contains multiple studies with only one observation, the remedy of focusing on the within-study variation only is not workable in practice. Too much statistical information will be discarded and the econometric model collapses due to near-singularity of the regression matrix. Also, the transformation of the dependent variable values is not workable in practice, mainly due to a lack of information. The assumption of revealed optimal efficiency (or a fixed linear relationship) may be too strong. The reinterpretation of the dependent variable has similar implications for the further analysis: no additional *statistical* procedures are needed to remedy the comparability problem. Therefore, in the remainder of the paper we will interpret the mean TE value as a measure of relative TE. When discussing the results of the empirical analysis in this paper we will pay attention to the close link between the mean relative efficiency value and the variation in individual efficiency values from the underlying study.

5. Quantitative overview of the literature on public transit efficiency performance

In order to discuss the literature on public transit efficiency we use a sample of parametric and non-parametric case studies. Most of these studies have been collected and used for a survey by De Borger et al. (2002). For a qualitative description and discussion of those studies we refer to this study. In Table 2 a quantitative summary of the literature is presented. The average TE ratio of the total set of studies is 0.825. This ratio is slightly higher for parametric studies (0.847) than for non-parametric studies (0.814).

Furthermore, we see that 70% of the parametric studies use panel data, either balanced or unbalanced. Among the non-parametric studies this is about 50%,

Table 2. Statistical overview of case studies used in the meta-analysis.

	Parametric studies	Non-parametric studies	Total sample
General			
Number of studies	15	18	33
Number of observations	30	63	93
Average TE	0.847	0.814	0.825
Database type			
Cross-section	16.7%	36.5%	30.1%
Time series	13.3%	14.3%	14.0%
Panel	60.0%	49.2%	52.7%
Unbalanced panel	10.0%	0.0%	3.2%
Region			
Europe	56.7%	60.3%	59.1%
USA	10.0%	33.3%	25.8%
Asia	33.3%	6.3%	15.1%
Parametric specifications			
Deterministic	32.3%		
Stochastic	66.7%		
Production frontier	32.3%		
Cost frontier	66.7%		
Non-parametric specification			
DEA		88.9%	
FDH		11.0%	
Output indicator related to			
Passengers	50.0%	22.2%	31.2%
Seats	40.0%	9.5%	19.4%
Vehicles	16.7%	69.8%	52.7%
Revenues	3.3%	22.2%	16.1%

all of them being balanced panel data sets. Apparently, among non-parametric studies cross-section databases seem to be preferred. About 60% of the studies use European data. In about 25% of the studies US data are used. Parametric studies use relatively more data from Asian countries while non-parametric studies use relatively more US data. Two thirds of the parametric studies use stochastic frontier methodologies. Only 33% use deterministic specifications. In two thirds of the parametric studies cost frontiers are used to derive TE. Production frontiers are used in the remaining studies. Most of the non-parametric studies use DEA techniques in order to construct an efficiency frontier (89%). FDH assumptions are used in the remaining 11% of these studies.

There is a large variability in the use of output measures, suggesting that there is no generally accepted set of appropriate variables in the urban transit sector. Most studies use a combination of different output measures. The majority of the measures used are related to numbers of passengers, seats and/or vehicles.[13] Table 2 shows that in parametric studies passenger and vehicle related output measures are more often used than in non-parametric studies. The latter focus more frequently on vehicle related output measures. In non-parametric studies combinations of different output measures are used more often than in parametric studies.

Figure 4 shows the distribution of TE ratios of the complete sample. We see that the ratios have a right-skewed distribution with a peak around 0.9. The shape of the distribution implies that, on average, there are relatively few very inefficient companies, and that in general urban transit companies tend towards efficiency.

Figure 5 shows the distribution of TE ratios of the set of parametric observations. Here the efficiency ratios are more or less uniformly distributed along the range 0.7–1.

Table 3 shows the average TE ratios for different subsets of the set of public transit efficiency observations. The observations are categorized according to study characteristics such as database type, geographic region, deterministic versus stochastic frontier specification, cost versus production frontier and non-parametric model type. As the table shows, the use of different study characteristics does generally not lead to large differences in efficiency ratios. Parametric studies that use stochastic frontier specifications find a slightly lower degree of efficiency on average compared with deterministic studies, which is somewhat surprising as deterministic studies attribute all errors to inefficiency. Furthermore, the set of parametric cost frontier studies shows a slightly lower efficiency ratio compared to the set of parametric studies that use production frontiers. This could be due to the fact that cost frontier inefficiencies include both technical and allocative effects. Observations based on time-series analysis and observations from studies based on USA data generally

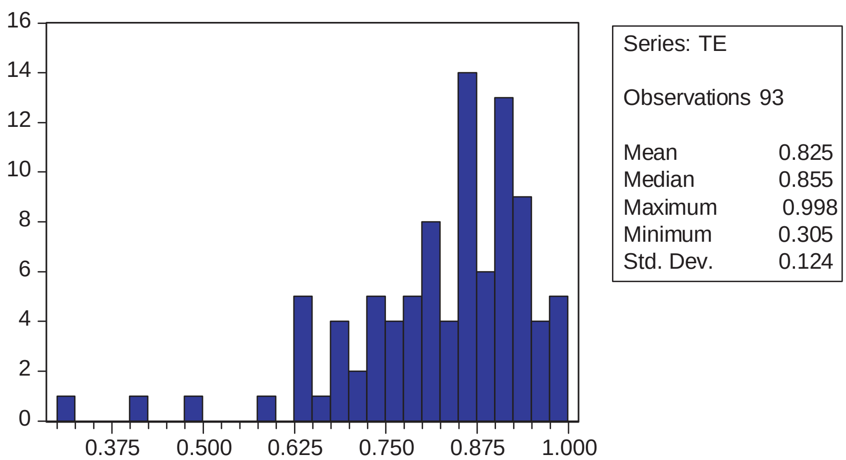

Figure 4. The distribution of mean TE values.

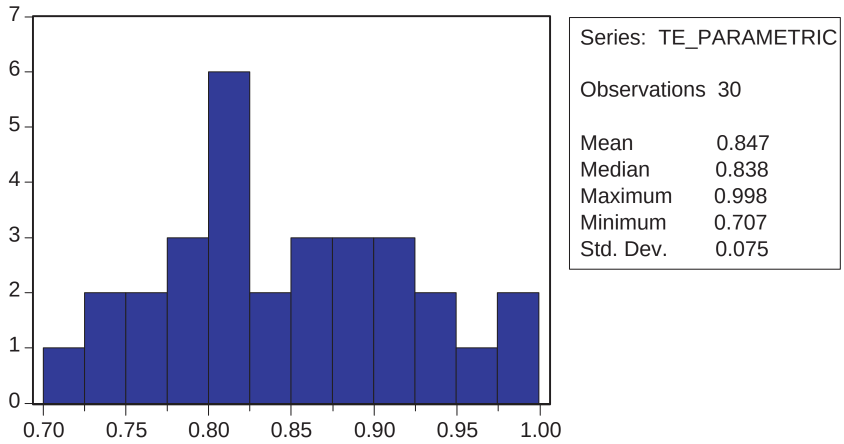

Figure 5. The distribution of mean TE values in parametric studies.

find higher relative efficiency values than their reference categories. This holds for both parametric and non-parametric sets of observations. Among the non-parametric studies, DEA studies report on average slightly lower relative efficiency levels than FDH studies. This can be explained from the fact that in DEA studies assumptions with respect to input–output relations are stronger

Table 3. Average TE for various study categories.

	Parametric studies	Non-parametric studies	Complete sample
Complete set of studies	0.847	0.814	0.825
Cross-section studies	0.823	0.767	0.777
Time series studies	0.919	0.895	0.903
Panel data studies	0.847	0.825	0.833
Unbalanced panel studies	0.791		0.791
European studies	0.842	0.826	0.792
USA studies	0.896	0.904	0.893
Asian studies	0.841	0.843	0.834
Stochastic studies	0.836		
Deterministic studies	0.870		
Cost frontier studies	0.836		
Production frontier studies	0.869		
DEA		0.809	
FDH		0.847	
Studies using passengers as output measure			0.802
Studies using vehicles as output measure			0.837
Studies using seats as output measure			0.821
Studies using revenues as output measure			0.844

than in FDH studies. This results in the construction of a frontier that is more efficient in terms of the input–output relationship. All other things being equal, observational units in DEA studies on average will be located farther away from the frontier than observational units in FDH studies.

In addition to the summarized relationships between TE and the various categorical variables as shown in the Table above, it might be interesting to investigate the univariate relationship between efficiency and some of the continuous variables that we use in our database by means of various scatter plots.

In Figure 6 the univariate relationship between GDP per capita and the TE ratio is shown. The correlation coefficient is approximately zero. Apparently, higher GDP and the associated increased car ownership, which in turn implies a lower demand for public transit does not affect efficiency ratios.

From the scatter plot in Figure 7, we see that the correlation between the year of data and TE is negative. As the previous scatter plot showed, this cannot be explained from the GDP growth over time and associated consequences for urban transit demand. Another reason for the negative sign of the latter coefficient might be that due to technological innovation and outsourcing the frontiers may have shifted to new efficiency standards. Thus, when the technical and institutional environments of urban transit become more heterogeneous, one may expect that variations in efficiency increase. Rigid public transit organizations may not have adjusted fully to these new standards.

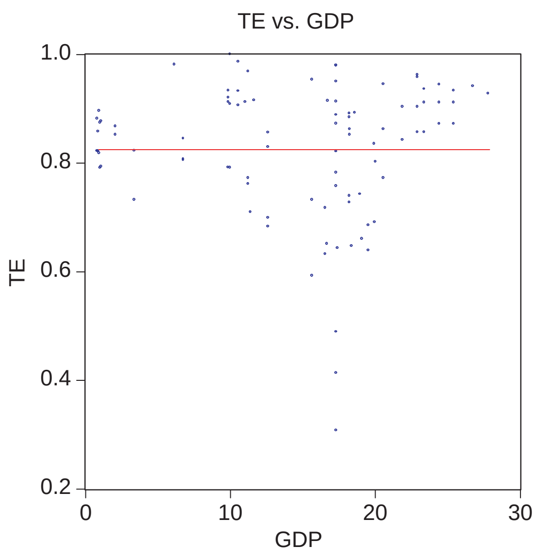

Figure 6. Correlation between TE and GDP per capita.

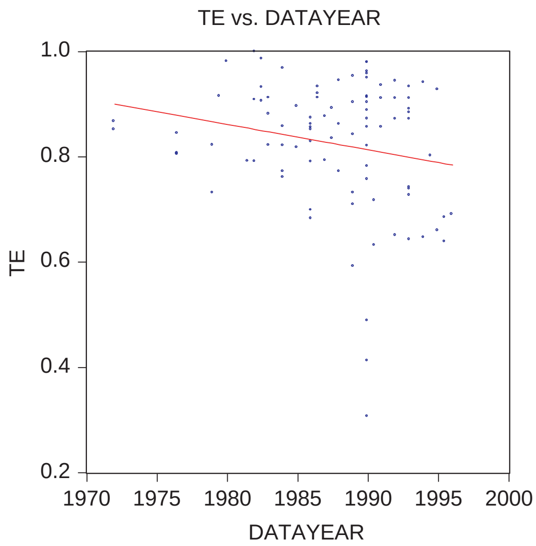

Figure 7. Correlation between TE and year of data.

Due to the nature of the dependent variable (i.e. the interpretation as a distribution measure), it is interesting to investigate the relationship between the mean TE value and the sample size of the underlying study (see also Zhang & Bartels 1998). Figure 8 shows the correlation between the two variables. The correlation is negative but not very strong at −0.220.

In Figure 9 the mean TE values are plotted against the number of inputs that are used in the original studies to estimate TE. The correlation coefficient is positive which may be explained from the fact that, in general, if more inputs are used in the econometric model, the explained variation in the systematic part of the cost or production function will be higher. Because of the way the mean TE value is constructed, an increased number of inputs may therefore lead to higher values.

6. Meta-analytical experiments

A major aim of this paper is to provide a statistical explanation for the variation in TE findings reported in the literature. In order to identify moderator variables that may explain such a variation we will apply meta-regression techniques. Compared to the exploratory analysis in the previous section the multi-variate techniques in this section provide us with more possibilities to correct for various factors while assessing the partial effects of certain variables. For this analysis, we pooled the sets of parametric and non-parametric

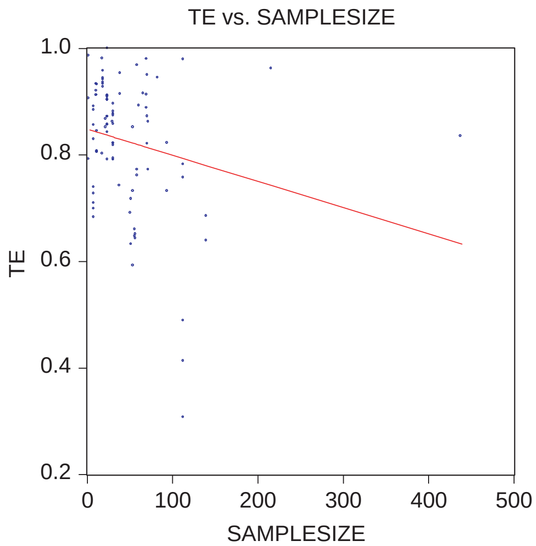

Figure 8. Correlation between TE and sample size.

studies and used a dummy variable to correct for fixed effects between the two sets as well as several other dummy variables to address systematic differences within other categorical variables. We also included a number of continuous variables.

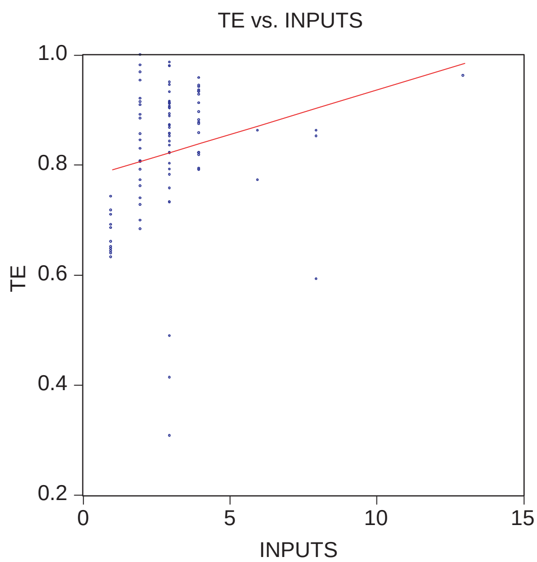

Figure 9. Correlation between TE and number of inputs.

In the econometric specification we use a transformation of the mean TE values as a dependent variable so that the estimated values will be within the range from zero to one.[14] In comparison to other econometric specifications this model has a better goodness of fit, i.e. a higher *R*-square.

Table 4 shows the results of a meta-regression where we used the following variables to explain the variation in the mean TE values: dummy variables for time series studies and panel data studies,[15] the year of data collection, GDP per capita, the share of government expenditure as a percentage of GDP, a dummy variable for stochastic parametric studies, dummies for USA and Asian data, a dummy for studies that use cost frontiers, a dummy for non-parametric DEA studies, a dummy for parametric studies, dummies for the type of output measures that are used in the estimation model, the sample size and the number of inputs.[16]

First we estimated the model using ordinary least squares. The results in column 1 show that the type of database that has been used in the original study affects the mean TE value. TE values for time series and panel data studies are significantly higher than those of cross-section studies. Apparently, increased TE, primarily due to technological change, has resulted in a variation in TE values that is large compared to the variation in TE values of a sample of

Table 4. Meta-regression results.

	(1)		(2)	
	Coefficient	Std. Error	Coefficient	Std. Error
Constant	−21.778	77.318	−85.507	134.537
Time series	1.528**	0.371	2.767**	0.779
Panel	0.952**	0.303	1.694**	0.393
Datayear	0.013	0.039	0.045	0.068
GDP	−0.077	0.057	−0.006	0.100
Gov_share	−0.062	0.031	−0.059	0.064
USA	1.378*	0.548	0.823	0.823
Asia	−0.633	0.627	0.151	0.978
Stochastic	−0.625	0.440	0.338	0.719
Cost frontier	−0.347	0.433	−1.905**	0.645
DEA	−1.401**	0.357	−1.590**	0.260
Parametric	−0.147	0.463	−0.049	0.714
Passengers	−0.710*	0.276	−1.073**	0.366
Vehicles	−0.902**	0.280	−1.521**	0.219
Seats	−1.078**	0.390	−0.830	0.621
Number of inputs	0.209**	0.075	0.443**	0.082
Sample size	−0.003	0.002	−0.007**	0.001
R-squared	0.511		0.884	
Adjusted *R*-squared	0.408		0.859	

*Significant at the 5% level; **Significant to the 1% level.

firms at a specific moment in time. The coefficient for the government share variable is negative but barely significant (the *t*-value is −1.982). Under the assumption that the government share is related to the degree of government intervention in the field of public transit this negative coefficient indicates that the efficiency of public transit operators benefits from a privatized environment and thus that the concerns about possible regulatory failures are valid. Furthermore we see that US studies report higher TE values compared to European studies. Perhaps the reason might be that in a market like the US, where the degree of privatization is higher and subsidies are lower, the relative importance of profit maximization versus accessibility as the industry's goal is higher. Since car ownership is higher in the USA and the infrastructure accommodates private transport to a larger degree, public transit may be less developed on low profit segments of the public transport network. Obviously, the concerns about possible regulatory failures that were briefly mentioned in the introduction play a role here. Furthermore, the dummy for DEA studies is significantly negative.[17] As previously discussed, this makes sense because of the stronger frontier assumptions of DEA studies compared to FDH studies. The dummies related to output measures enter significantly negative. The reason may be that the inclusion of revenues in the estimation model leads to a better goodness of fit; the reference category here contains primarily studies that use revenues as an output measure. The number of inputs variable enters positive and significant, which confirms our expectations as to the positive relationship between goodness of fit and the mean TE value. Finally, the coefficient of the sample size variable enters insignificantly.

The second estimation uses the same set of explanatory variables but uses weighted least squares (WLS), weighting for sample size. The results are in column 2. The US dummy and the dummy for observations using seats as output lose their significance. Cost frontier studies and the sample size enter significantly, both negative. This result corresponds with the correlation between TE value and sample size shown in Figure 8.

7. Summary and concluding remarks

The results of the paper show that TE values are similar for parametric and non-parametric studies. Estimations based on time series or panel data result in a higher mean TE value. Furthermore, DEA assumptions lead to lower efficiency values than FDH assumptions. The variables that are used to measure output also affect the TE value; the use of passenger-, seats-, and vehicle-related output indicators leads to lower efficiency estimations compared to studies based on revenue measures. The sample size of the original case study does not affect the TE value when using OLS estimation but has a significantly negative coefficient with WLS. The number of inputs that are used in the estimation model has a positive effect on the mean TE value.

Among parametric studies, there is no significant difference between stochastic and deterministic specifications although one would expect the stochastic specification to lead to higher TE values since part of the variation is attributed to measurement error, whereas for deterministic specifications all the variation is assumed to reflect inefficiency. The use of cost frontiers versus production frontiers does not affect the TE value when using OLS estimation. When using WLS the cost frontier dummy has a significantly negative coefficient. The GDP per capita does not show a significant influence on efficiency values.

Two variables that are of particular interest are the dummy variable for USA based studies and the variable for government share. When using OLS estimation, the dummy for USA studies enters positive and significant. An explanation could be that in a market like the USA, with a relatively high degree of privatization, the importance of profit maximization versus accessibility is higher than for instance in Europe. Furthermore, since car ownership is higher in the USA, public transit may be underdeveloped on low profit segments of the public transport network. As such, a higher degree of privatization allows a firm to reach a higher level of TE. The government share variable enters with a negative coefficient. Assuming that there is a relation between the government share and the degree of government intervention in the field of public transit this coefficient also implies that a privatized environment is beneficiary for the efficiency of public transit operators. Obviously, the concerns about possible regulatory failures that were mentioned in the introduction play a role here. However, with WLS both the USA dummy and the government share variable lose their significance. This is probably due to two USA based studies (Nolan 1996; Nolan et al. 2002) that show high TE ratio's and are based on smaller than average sample sizes (20 and 25 respectively).

Notes

1. Other efficiency concepts frequently encountered in the literature include scale efficiency, which relates to a divergence between actual and ideal production size, and structural inefficiency, which relates to possible congestion of production regions (see De Borger et al. 2002).
2. Allocative efficiency requires the specification of a behavioural goal and is defined by a specific point on the boundary of the production possibility set that satisfies this objective, given certain constraints on prices and quantities.
3. This method is sometimes referred to as Displaced OLS (DOLS).
4. Note that, by definition, any point on the efficiency frontier between E and the y_1-axis is efficient, although the production bundle is dominated by the bundle in point E for all positive prices of y_1. Such differences between optimal and non-optimal points on the frontier are referred to as 'slacks'. The undesirable properties of slacks may be circumvented by adding an additional coefficient to the Linear Programming problem, which prevents the frontier from running parallel to the axis.

5. Note that the set of observation points is the same in panel and panel. Point C, which is located on the efficiency frontier under FDH assumptions, is not efficient under DEA assumptions.
6. To see this, compare the degree of inefficiency of observation point F under DEA and under FDH assumptions.
7. A list of all case studies used can be found in the references section of this paper.
8. In the figure the frontiers are located as they are for educational purposes. Note that in practice we do not directly observe the difference between intercepts (although under certain conditions they can be derived indirectly).
9. Comparison is also possible under the assumption that the *actual* (non-observed) TE of 'efficient' firms is similar among studies.
10. This includes conditions related to the institutional, cultural, geographic and economic environment.
11. Note that this only holds if the TE values from one study are all based on the same set of observations or are all based on different subsets of the same set of observations.
12. Consider the following measure of variation: $\sum_n |1 - x_n|$. Because x in our case is a measure of efficiency, it is positive and smaller than or equal to 1. Dividing this measure by n then yields $1{-}x_n$, which yields the same information as x_n, the average TE-value.
13. For example, the dummy variable "passengers" is a dummy for studies that include in the econometric model an output measure related in one way or another to passengers (e.g. number of passengers per km (or mile), number of passengers per month etc.).
14. We use the transformation $\ln[y/(1-y)]$ to estimate the following model: $y = \exp(X'\beta)/[1 + \exp(X'\beta)] + \mu$
15. This dummy refers to both balanced and unbalanced panel data.
16. By using dummy variables for parametric studies, stochastic studies and DEA studies, we basically follow the taxonomy of methodologies in Table 1. Note that these dummies are not directly related to each other in the sense that they aim to form a mutually exclusive set, so that multi-collinearity may be a problem. We experimented with an alternative dummy structure that is based on the notion that ultimately there are four frontier methodologies that can be made directly comparable by including dummies for three of them and using the remaining one as a reference category. The results of such a dummy structure are essentially the same as the one we use here. We extended this experiment by further dividing the methodologies based on the distinction between cost- and production-based studies, resulting in six frontier methodologies and thus five dummies. The results were very similar to our findings (results are available on request).
17. Although formally the reference category is non-DEA studies, we can interpret the result as if the reference category would be FDH studies. This is because we use a dummy to correct for parametric studies and both DEA and FDH studies are non-parametric studies.

Case studies

Bhattacharyya A, Kumbhakar S & Bhattacharyya A (1995) Ownership structure and cost efficiency: A study of publicly owned passenger-bus transportation companies in India. *Journal of Productivity Analysis* 6: 47–61.

Button K & Costa A (1999) Economic efficiency gains from urban public transport regulatory reform: Two case studies of changes in Europe. *Annals of Regional Science* 33: 425–438.

Chang K-P & Kao P-H (1992) The relative efficiency of public versus private municipal bus firms: An application of data envelopment analysis. *Journal of Productivity Analysis* 3: 67–84.

Chu X, Fielding G & Lamar B (1992) Measuring transit performance using data envelopment analysis. *Transportation Research* 26A: 223–230.

Costa Á (1998) Public transport efficiency and effectiveness: Metro de Madrid. In: Button K, Nijkamp P, & Priemus H (eds) *Transport Networks in Europe: Concepts, Analysis and Policies* (pp. 248–264). Cheltenham: Edward Elgar.

Costa Á & Markellos RN (1997) Evaluating public transport efficiency with neural network models. *Transportation Research* 5C: 301–312.

Cowie J & Asenova D (1999) Organization form, scale effects and efficiency in the British bus industry. *Transportation* 26: 231–248.

De Jong G & Cheung F (1999) Stochastic frontier models for public transport. In: Meersman H, van de Voorde E & Winkelmans W (eds) *World Transport Research: Selected Proceedings of the 8th World Conference on Transport Research Vol.1*: *Transport Modes and Systems* (pp. 373–386). New York: Pergamon.

Delhause B, Perelman S & Thiry B (1992) Substituabilité partielle des facteurs etefficacité-coût: L'example des transports urbain et vicinal Belges. *Economie et Prévision* 32: 105–115.

Fazioli R, Filippini M & Prioni P (1993) Cost-structure and efficiency of local public transport: The case of Emilia Romagna bus companies. *International Journal of Transport Economics* 20: 305–324.

Filippini M, Maggi R & Prioni P (1992) Inefficiency in a regulated industry: The case of Swiss regional bus companies. *Annals of Public and Cooperative Economics* 63: 437–455.

Gathon H-J (1989) Indicators of partial productivity and technical efficiency in the European urban transit sector. *Annals of Public and Cooperative Economics* 60: 43–59.

Hanusch H & Cantner U (1991) Produktion öffentlicher Leistungen: Effizienz und Technischer Fortschritt. *Jahrbücher für Nationalökonomie und Statistik* 208: 269–384.

Kerstens K (1996) Technical efficiency measurement and explanation of French urban transit companies. *Transportation Research* 30A: 431–452.

Khumbhakar S & Bhattacharyya A (1996) Productivity growth in passenger-bus transportation: A heteroskedastic error component model with unbalanced panel data. *Empirical Economics* 21: 557–573.

Levaggi R (1994) Parametric and non-parametric approach to efficiency: The case of urban transport in Italy. *Studi Economici* 49: 67–88.

Lijesen M (1998) Analyzing cost structures of public sector activities, with an application to regional public transport. In: de Graan GBK and Volmer FG (eds), *Performance Budgeting: A Perspective on Modelling and Strategic Planning in the Public Sector in Holland* (pp. 178–187). Delft: Eburon.

Loizides I & Giahalis B (1995) The performance of public enterprises: A case of the Greek railway organization. *International Journal of Transport Economics* 22: 283–306.

Matas A & Raymond J-L (1998) Technical characteristics and efficiency of urban bus companies: The case of Spain. *Transportation* 25: 243–263.

Nolan JF (1996) Determinants of productive efficiency in urban transit. *Logistics and Transportation Review* 32: 319–342.

Nolan JF Ritchie PC & Rowcroft JE (2002) Identifying and measuring public policy goals: ISTEA and the US bus transit industry. *Journal of Economic Behavior and Organization* 48: 291–304.

Obeng K (1994) The economic cost of subsidy-induced technical inefficiency. *International Journal of Transport Economics* 21: 3–20.

Sakano R & Obeng K (1995) Re-examination of inefficiencies in urban transit systems: A stochastic frontier approach. *Logistics and Transportation Review* 31: 377–392.

Sakano R Obeng K & Azam G (1997) Subsidies and inefficiency: Stochastic frontier approach. *Contemporary Economic Policy* 15: 113–127.

Thiry B & Tulkens H (1992) Allowing for inefficiency in parametric estimation of production functions for urban transit firms. *Journal of Productivity Analysis* 3: 45–65.

Tone K & Sawada T (1990) An efficiency analysis of public vs. private bus transportation enterprises. In: H.Bradley (ed), *Operational research '90* (pp. 357–365). New York: Pergamon.

Tulkens H & Vanden Eeckaut P (1995) Non-frontier measures of efficiency, progress and regress for time series data. *International Journal of Production Economics* 39: 83–97.
Tulkens H & Wunsch P (1994) Les Performance économiques de la STIB au mois le mois. *Revue Suisse d'Économie Politique et de Statistique* 130: 627–646.
Tulkens H (1993) On FDH efficiency analysis: Some methodological issues and applications to retail banking, courts and urban transit. *Journal of Productivity Analysis* 4: 183–210.
Tulkens H, Thiry B & Palm A (1988) Mesure de l'efficacité productive: methodologies et applications aux sociétés de transports intercommunaux de Liege, Charleroi et Verviers. In: Thiry B and Tulkens H (eds) *La Performance Economique des Sociétés Belges de Transport Urbains* (pp. 81–136). Charleroi: CIRIEC.
Viton P (1986) The question of efficiency in urban bus transportation. *Journal of Regional Science* 26: 499–513.
Viton P (1992) Consolidations of scale and scope in urban transit. *Regional Science and Urban Economics* 22: 25–49.
Viton P (1997) Technical efficiency in multimode bus transit: A production frontier analysis. *Transportation Research* 31B: 23–39.
Viton P (1998) Changes in multimode bus transit efficiency, 1988–1992. *Transportation* 25: 1–21.
Wunsch P (1994) Costing busses: Back to basics. Brussels *FUSL* (SMASH Cahier 9405).

References

Banister D (2000) Sustainable urban development and transport – a Eurovision for 2020. *Transport Reviews* 20: 113–130.
Charnes A Cooper WW & Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operations Research* 2: 429–444.
De Borger B Kerstens K & Costa Á (2002) Public transit performance: What does one learn from frontier studies? *Transport Reviews* 22: 1–38.
Deprins D Simar L & Tulkens H (1984) Measuring labor efficiency in post offices. In: *The Performance of Public Enterprises: Concepts and Measurement* (pp. 234–267), Amsterdam: Elsevier Science.
Farrell M (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society* 120: 253––281.
Ganley J & Cubbin J (1992), Public sector efficiency measurement. In: *Applications of Data Envelopment Analysis*. Amsterdam: Elsevier Science.
Glaister S, Starkie D & Thompson D (1990) The assessment: Economic policy for transport. *Oxford Review of Economic Policy* 6: 1–21.
Nollet C, Thiry B & Tulkens H (1988) Mesure de l'efficacité productive: Application a la société de transports intercommunaux de Bruxelles. In: Thiry B and Tulkens H (eds) *La Performance Economique des Sociétés Belges de Transport Urbains* (pp. 137–170). Charleroi: CIRIEC.
Zhang Y & Bartels (1998) The effect of sample size on the mean efficiency in DEA: With an application to electricity distribution in Australia, Sweden and New Zealand. *Journal of Productivity Analysis* 9: 187–204.

About the authors

Martijn Brons is a Ph.D. candidate at the Department of Spatial Economics, Free University of Amsterdam and the Tinbergen Institute, Amsterdam. His research themes include transportation economics, transport and the environment, economic growth.

Peter Nijkamp is professor in regional and urban economics and in economic geography at the Free University of Amsterdam. Since 2002 he is president of the governing board of the Netherlands Research Council (NWO). He is a fellow at the Tinbergen Institute, Amsterdam. His research themes include regional and urban planning, transport systems analysis, technological innovation.

Eric Pels holds a *post- doc* position at the Department of Spatial Economics, Free University of Amsterdam. His research themes include aviation economics, transport economics, transport and the environment.

Piet Rietveld is professor in transportation economics at the Department of Spatial Economics, Free University of Amsterdam. He is a fellow at the Tinbergen Institute, Amsterdam. His research themes include transport and regional development, valuing quality of transport services, economics of public transport.