Research in Transportation Economics 23 (2008) 53–64

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: www.elsevier.com/locate/retrec

Transit costs and cost efficiency: Bootstrapping non-parametric frontiers

Bruno de Borger [a,*], Kristiaan Kerstens [b], Matthias Staat [c]

[a] University of Antwerp, Antwerp, Belgium
[b] CNRS-LEM, IESEG School of Management, Lille, France
[c] DSC GmbH, Schriesheim, Germany

ARTICLE INFO

Article history:
Available online 1 March 2009

ABSTRACT

This paper explores a selection of recently proposed bootstrapping techniques to estimate non-parametric convex (DEA) cost frontiers and efficiency scores for transit firms. Using a sample of Norwegian bus operators, the key results can be summarized as follows: (i) the bias implied by uncorrected cost efficiency measures is numerically important (close to 25%), (ii) the bootstrapped-based test rejects the constant returns to scale hypothesis, and (iii) explaining patterns of efficiency scores using a two-stage bootstrapping approach detects only one significant covariate, in contrast to earlier results highlighting, e.g., the positive impact of high-powered contract types. Finally, comparing the average inefficiency obtained for the Norwegian data set with an analogous estimate for a smaller French sample illustrates how the estimated differences in average efficiency almost disappear once sample size differences are accounted for.

1. Introduction

Estimates of the cost structure and the cost efficiency of transit operators provide highly useful information for transport policy-makers. Whatever the ownership status and the regulatory environment in which transit firms work, proper knowledge of the cost structure (minimal costs, economies of density, economies of scale, input flexibility, network characteristics, services characteristics, etc.) and its determinants (contract type, regulatory environment, subsidies, etc.) is crucial for guiding decisions with respect to pricing, investment, supply adjustments, etc. Moreover, the search for potential efficiency improvements of transit firms has been a constant interest both from a policy and an academic viewpoint. It has become especially relevant over the past decades because, in most western economies, the demand for transit has been almost everywhere declining due to suburbanization tendencies and modal shifts towards private-car transport. Finally, proper cost and efficiency information add useful insights on the desirability of regulatory reforms, it provides information on how to limit cost and subsidy levels, and it contributes to the discussion on the relative merits of private versus public provision.

Several approaches exist to the estimation of the costs and efficiency of transit firms. The parametric frontiers require functional form specifications: flexible functional forms such as the translog or generalized Leontief have been quite popular in empirical applications (for surveys of parametric approaches to cost estimation, see Pels & Rietveld, 2008; Small & Verhoef, 2007: Ch. 3). Non-parametric methods instead determine the cost frontier without assuming a functional form. These methods envelop the data by piecewise linear hyperplanes using mathematical programming methods. Data Envelopment Analysis (DEA) and the Free Disposal Hull (FDH) are the most popular techniques: the former imposes convexity, the latter does not. Both parametric and non-parametric models determine inefficiencies as deviations from the estimated frontier and thereby offer a "benchmarking" perspective. Moreover, depending on the nature of the data, frontier analyses also potentially yield information on productivity changes over time, calculated by considering shifts in the frontier over time. Recent surveys of studies analyzing transit costs, productivity and efficiency include De Borger, Kerstens, and Costa (2002), De Borger and Kerstens (2008), and Waters (2008).[1]

In recent decades, substantial progress has been made in the specification and estimation of parametric cost models for transit firms. For example, problems associated with the heterogeneity of outputs provided by transit firms and the intrinsic spatial nature of the network they operate have been carefully dealt with. This has induced important innovations in output measurement and in modelling economies of scale and scope (see, e.g., Basso & Jara-Díaz, 2005; Spady & Friedlander, 1978, for a survey, see De Borger & Kerstens 2008). Moreover, a variety of econometric studies now

[1] Following up on De Borger, Kerstens, and Costa (2002) with a meta-analysis, Brons et al. (2005) list 33 studies: 15 parametric and 18 non-parametric. This could reveal a close to even popularity of both families of methods in urban transit studies.

* Corresponding author.
E-mail address: bruno.deborger@ua.ac.be (B. de Borger).

0739-8859/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.retrec.2008.10.008

exists that analyze the impact of contractual arrangements and regulatory policies on the cost structure and on the relative efficiency of transit systems (Dalen & Gómez-Lobo, 2003; Gagnepain & Ivaldi, 2002a, 2002b; Kerstens, 1996; Roy & Yvrande-Billon, 2007).[2] In addition, there is now overwhelming evidence of a negative relationship between the levels of subsidies and operating cost efficiency (Kerstens, 1996; Matas & Raymond, 1998; Sakano & Obeng, 1995).

This paper focuses on recent developments in *non-parametric* estimation of cost frontiers, and on the use of these techniques to study transit costs and cost efficiency. Starting with the seminal article of Chu, Fielding, and Lamar (1992), empirical applications of these techniques to transit firms have become quite popular. For instance, Cowie and Asenova (1999) have studied British urban transit, while Kerstens (1996) and Odeck and Alkadi (2001) have analyzed, respectively, French and Norwegian bus transport (see De Borger & Kerstens, 2008 for a more complete overview). Although it has been argued that non-parametric methods have substantial advantages over parametric techniques (e.g., because no a priori functional form is specified), one potential reason why this methodology is sometimes met with skepticism is that the statistical properties of these non-parametric cost models have remained unexplored in the early literature. Essentially, only point estimates of efficiency are obtained from these estimators. For a long time, the lack of statistical tests on essential properties of the cost structure made it difficult to test relevant economic hypotheses with respect to, for example, returns to scale, input substitution, or the significance of inefficiency scores.

Fortunately, however, a recent specialized literature has developed statistical inference tools for non-parametric frontier models (see the Simar & Wilson, 2008 survey). It has been forcefully argued that efficiency estimators derived from such frontiers are intrinsically biased and that the bias depends, among others, on the sample size and on the number of dimensions (outputs and input prices) captured by the model (see, e.g., Simar & Wilson, 2000; Zhang & Bartels, 1998). To correct for these shortcomings, the use of bootstrapping techniques is suggested. Moreover, Simar and Wilson (2007) recently expanded this methodology to the typical semi-parametric two-stage models. These determine non-parametric efficiency scores in a first stage, and then econometrically explain inefficiency patterns using available structural and environmental characteristics of firms. Finally, bootstrapping also seems to offer an elegant solution to the difficulty of comparing efficiency estimates obtained on the basis of widely different sample sizes and specifications (Zhang & Bartels, 1998).

The purpose of this paper is, therefore, to explore a selection of these most recently developed techniques for estimating non-parametric convex (DEA) cost frontiers for transit firms and to show their usefulness to test characteristics of the cost structure and to derive cost efficiency scores.[3] Of course, the specific focus of the paper implies that some other important issues are not dealt with. For instance, transit agencies recently devote substantial resources to curb emissions or to promote other social goals, which may well distract from the production of vehicle kilometers. For example, Nolan, Ritchie, and Rowcroft (2002) attempt to capture urban transit efficiency as well as wider social goals (mobility, energy savings and pollution reduction, among others) and illustrate that the pursuit of these goals affects traditional transit efficiency scores. For the specific goal of emission reduction, McMullen and Noh (2007) use a directional distance function approach to model the joint production of good (vehicle kilometers) and bad (emissions) outputs; they nicely illustrate how different efficiency rankings emerge once bad outputs are included.[4]

This paper sets four precise goals. First, we apply bootstrapping methods to correct for the inherent bias in non-parametric cost efficiency estimates of transit firms. In doing so, we adapt the bootstrap algorithm, originally developed for technical efficiency measures, to the case of overall cost efficiency ratios. The proposed methodology is illustrated exploiting a well-known Norwegian database that has been used several times before in the literature (see, e.g., Jørgensen, Pedersen, & Volden, 1997). We find the bias implied by standard non-parametric cost efficiency measures to be numerically important. It amounts to close to 25%, both for a model that imposes constant returns to scale and for a model allowing for variable returns. Second, we use bootstrapping to test for constant returns to scale of the cost frontier and its underlying technology. Third, we attempt to explain observed cost efficiency patterns by a variety of operating environment and regulatory characteristics, using the recent Simar and Wilson (2007) two-step bootstrapping procedure. Fourth, we illustrate the use of a Monte-Carlo-technique to compare transit cost inefficiencies in two samples of different size. To do so, we study a second sample, previously analyzed by Gagnepain and Ivaldi (2002a, 2002b), providing information on French transit firms. We then empirically compare inefficiencies of Norwegian and French transit firms, correcting for differences in sample size along the lines suggested by Zhang and Bartels (1998). We find that correcting for sample size differences has important implications for estimated average inefficiencies.

The structure of this paper is as follows. To set the stage, in Section 2 we first develop the basic microeconomic framework for estimating cost functions and cost efficiency. We then summarize some of the recent methodological contributions to non-parametric cost frontier estimation. In particular, we intuitively explain the role of bootstrapping in resolving problems of statistical inference on the basis of estimated non-parametric cost frontiers. The techniques employed in the empirical analysis are formally explained in some detail in Section 3. Empirical results are reported in Section 4. Finally, Section 5 concludes.

2. Estimating cost functions: theory and developments in empirical methodologies

In this section, we first briefly review the microeconomic foundations of production and cost functions, and show how technical and cost efficiency follow naturally. We then study recent methodological advances in non-parametric efficiency estimation.

2.1. Microeconomic foundations of production, cost and efficiency

Technology transforms inputs $x = (x_1, \ldots, x_n) \in \Re^n_+$ into outputs $y = (y_1, \ldots, y_m) \in \Re^m_+$. The production possibility set T is the set of all feasible input and output vectors: $T = \{(x, y) \in \Re^{n+m}_+;\quad x \text{ can produce } y\}$. It is standard to impose the following assumptions on technology (e.g., Färe & Primont, 1995):

[2] Among these, the study by Gagnepain and Ivaldi (2002b) is most firmly grounded in economic theory. They allow for an inefficiency term consisting of two components: an exogenous "pure" inefficiency and an endogenous component, which depends on the optimal effort for cost reduction that the firm exercises. This effort level follows from optimizing behaviour, taking account of both the cost of effort and the productivity of effort. This leads to a cost function incorporating the optimal effort level. The results show that ignoring effort adjustments has fairly limited effects for the cost structure, but it does lead to distorted estimates of efficiency.

[3] Two recent papers applying bootstrapping to efficiency in the transport sector are Boame (2004) and Odeck (2006). Neither of these covers the wide range of methodologies analyzed in this paper.

[4] The directional distance function generalises existing distance functions (being dual to a profit function) and can be extended to model the joint production of good and bad outputs (Chung, Färe, & Grosskopf, 1997).

(T.1) $(0,0)\in T$; $(0,y)\in T \Rightarrow y = 0$, i.e., inaction and no outputs without inputs;
(T.2) T is a bounded set, i.e., infinite outputs are not allowed with a finite input vector;
(T.3) T is a closed set;
(T.4) $\forall (x,y)\in T, (x,-y) \leq (u,-v) \Rightarrow (u,v)\in T$, i.e., fewer outputs can always be produced with more inputs, and inversely;
(T.5) T is a convex set.

The input set associated with the production possibility set T denotes all input vectors $x\in\Re^n_+$ capable of producing a given output vector $y\in\Re^m_+$: $L(y)=\{x|(x,y)\in T\}$.

The input distance function offers a complete characterization of technology; it is defined as

$$D_i(x,y) = \begin{cases} \max\{\lambda : \lambda \geq 0, \quad x/\lambda\in L(y)\} & \text{if } y\neq 0, \\ +\infty & \text{if } y = 0. \end{cases} \tag{1}$$

We next define the radial input efficiency measure as:

$$DF_i(x,y) = \min\{\theta|\theta \geq 0, (\theta\, x)\in L(y)\}. \tag{2}$$

This radial measure of input technical efficiency ($DF_i(x,y)$) is simply the inverse of this input distance function ($DF_i(x,y) = [D_i(x,y)]^{-1}$). It denotes the proportional reduction in inputs that are feasible while maintaining production of a given output vector. Its most important properties are: (i) $0 < DF_i(x,y) \leq 1$, with efficient production on the boundary (isoquant) of $L(y)$ represented by unity; (ii) it has a cost interpretation (see Färe, Grosskopf, & Lovell, 1985 for details).[5]

The cost function, as a dual representation of technology, indicates the minimum expenditures needed to produce an output vector y given a vector of positive input prices $w = (w_1,\ldots,w_n)\in\Re^n_{++}$: $C(y,w)=\min\{wx|x\in L(y)\}$. The same cost function can also be written in terms of the input distance function. This is the basis for the dual relations that establish the foundations for efficiency measurement (e.g., Färe & Primont, 1995). Formally, the duality between the input distance function and the cost function can be written as follows:

$$C(y,w) = \min_x\{wx : D_i(x,y) \geq 1\} \quad w > 0 \tag{3}$$

$$D_i(x,y) = \min_w\{wx : C(y,w) \geq 1\} \quad x\in L(y) \tag{4}$$

These two expressions clearly show the close relation between the cost function and the input distance function. While the cost function can be obtained from the input distance function by optimizing with respect to input quantities, the input distance function can be resolved from the cost function by minimizing with respect to input prices.

The properties of the cost function in prices and outputs are well known. Under minimal regularity conditions, the cost function $C(y,w)$ has the following properties (see Luenberger, 1995):

(a) Homogeneous of degree one in prices w: $C(y,\alpha w)=\alpha C(y,w)$ for $\alpha>0$.
(b) Non-decreasing in prices w: if $w'\geq w$, then $C(y,w')\geq C(y,w)$.
(c) Concave in prices w: $C(y,\alpha w_1+(1-\alpha)w_2)\geq \alpha C(y,w_1)+(1-\alpha)C(y,w_2)$ for $0\leq\alpha\leq 1$.
(d) Non-decreasing in outputs y: if $y'\geq y$, then $C(y',w)\geq C(y,w)$.

To establish a framework for efficiency measurement, we discuss a few points in more detail. It is clear that for each element of the input set ($x\in L(y)$), the following inequality holds:

$$C(w,y) \leq w\left(\frac{x}{D_i(x,y)}\right). \tag{5}$$

Thus, the minimal costs are smaller or equal to the observed cost measured at the isoquant of the input set (i.e., after eliminating any eventual technical inefficiency). This inequality (5) can be rewritten as follows:

$$C(w,y)D_i(x,y) \leq wx. \tag{6}$$

In this form, inequality (6) is known as Mahler's inequality (see Färe & Grosskopf, 2000).

The transformation of this inequality into equality by adding an allocative efficiency component $AE(w,x,y)$ forms the theoretical foundation for the multiplicative Farrell (1957) decomposition for measuring input efficiency:

$$\frac{C(w,y)}{wx} = \frac{1}{D_i(x,y)}AE(w,x,y). \tag{7}$$

The first ratio of mimimal to observed costs $C(w,y)/wx$ defines a cost efficiency component. This is also in general labeled an overall efficiency component. The second ratio $1/D_i(x,y)$ coincides simply to the radial measure of input technical efficiency ($DF_i(x,y)$), as already alluded to above. Finally, the component $AE(w,x,y)$ indicates the allocative efficiency, defined in a residual way. This is formally written in the following definition:

Definition 1. Under the assumptions (T.1)–(T.5) on the input set $L(y)$, the following input-oriented efficiency notions can be distinguished:

(1) Overall Efficiency is the quantity: $OE_i(x,y,w)=C(y,w)/wx$.
(2) Technical Efficiency is the quantity: $TE_i(x,y)=DF_i(x,y)$.
(3) Allocative Efficiency is the quantity: $AE_i(x,y,w)=OE_i(x,y,w)/TE_i(x,y)$.

This analysis immediately makes clear that cost efficiency is a more severe criterion than technical efficiency when benchmarking firms ($OE_i(x,y,w)\leq TE_i(x,y)\leq 1$). Notice that Färe, Grosskopf, and Lovell (1983, 1985: pp. 3–5) offer an even more extended efficiency taxonomy (by splitting up technical efficiency into congestion, scale efficiency and pure technical efficiency).

The above basic microeconomic theory of cost functions and their dual relation to the input distance function as a representation of technology has clearly established that efficiency measurement is firmly grounded on microeconomics.

2.2. Recent methodological advances in non-parametric estimation methods

The recent literature has generated a wide variety of developments in non-parametric estimation of production and cost frontiers.[6] Many of these efforts have concentrated on the statistical properties of the efficiency estimators, which were often naively depicted as deterministic in nature. For surveys on these issues, we refer to Cherchye and Post (2003), Daraio and Simar (2007),

[5] For reasons of convenience, we stick to the traditional radial input efficiency measure, i.e., the inverse of the input distance function. Recently, more general directional distance functions have been introduced to measure profit efficiency (see Chambers, Chung, & Färe, 1998). Apart from the fact that these new measures lead to additive rather than multiplicative decompositions, they can be exactly related to the traditional radial efficiency measures employed in this contribution.

[6] Of course, innovations in parametric estimation have been introduced as well. For example, substantial progress has been made in estimating distance functions that represent multi-input multi-output production processes, in integrating the explanation of efficiency patterns into the one-sided inefficiency error component, in the inclusion of efficiency changes in models decomposing productivity change, etc. The interested reader is referred to Kumbhakar and Lovell (2000) for a summary of the current state of the art in this domain and to Sickles (2005) for a comparison among recent parametric and other estimators.

Grosskopf (1996), and Simar and Wilson (2000, 2008). In what follows, we focus on three recent advances.

The first one is related to small sample bias. Indeed, a crucial result that has been established is that both the convex (DEA) and non-convex (FDH) estimators are consistent, but –unfortunately– they also have very slow rates of convergence. This implies that, when only small samples are available, these cannot generate an adequate representation of technology, and the resulting efficiency estimates are biased. Then, inefficient firms can be wrongly classified as efficient, or the "true inefficiency" of an inefficient observation can be substantially underestimated. This follows from the fact that non-parametric production analysis provides a local and inner bound approximation to the true, unknown and possibly larger production possibility set. This small sample bias depends on specific properties of the underlying data in a given model. In particular, it is related to (i) the number of observations in the sample, (ii) the number of inputs and outputs, and (iii) the density of observations around the relevant segment of the frontier. Since non-parametric estimators only provide an inner approximation of the true frontier, adding more observations can only improve the approximation of the true frontier, hence reducing the eventual gap between efficiency estimates and the true efficiency. Similarly, the more input and output dimensions are included in a given model, the more serious the bias becomes for a given sample size.

The small sample bias can easily be remedied if knowledge of the sampling distribution was available. This would allow constructing confidence intervals, and it is the basis for developing any test statistics. In principle, two approaches exist to obtain sampling distributions of the frontier estimates: (i) theoretical results based on asymptotics, and (ii) bootstrapping techniques. Analytic derivation of the asymptotic sampling distribution has so far yielded a limited number of general results, at least for the more popular convex (DEA) estimators (see, e.g., Gijbels, Mammen, Park, & Simar, 1999). Alternatively, the approximation of the sampling distribution using the bootstrap, a common statistical re-sampling technique, has led to important breakthroughs. Given the linear programming nature of the frontier efficiency estimators, bootstrapping makes it possible to employ brute computer force to overcome analytical intractability. The first bootstrapping procedure tailored to the needs of frontier estimation has been developed in Simar and Wilson (1998).

Given a consistent description of the data generating process, the principle of bootstrapping involves the repeated simulation of this process, and the application of the original estimator to each simulated sample such that the resulting estimators mimic the sampling distribution of the original estimator. Applying this technique to DEA cost efficiency measures can be understood intuitively as follows. First, a standard cost efficiency measure is computed relative to a non-parametric DEA technology. Then, the density $f()$ of the efficiency scores is estimated by a kernel density.[7] Second, this density allows drawing pseudo-scores that follow the same distribution as the scores obtained with the original sample. Third, these simulated scores make it possible to generate a number of B pseudo-data sets, which are then used to obtain B new sets of efficiency scores. Finally, these new efficiency scores enable us to estimate and correct for the bias. Technical details are developed in more detail below.

The intuition is illustrated in Fig. 1. The observations (denoted *a* to *i*; they are represented by squares) support a non-parametric isoquant. The bootstrapping from pseudo-data sets (represented by circles) allows reconstructing each time a new isoquant that can be situated slightly outside (but also in intersection, etc.) the initial non-parametric isoquant. Repeating this process a large number of times

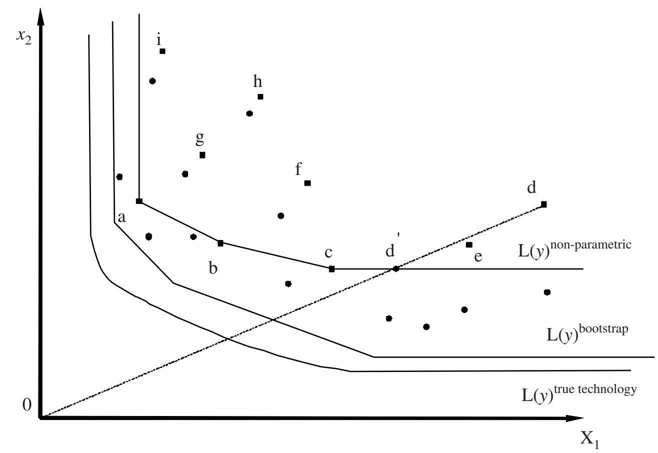

Fig. 1. Bootstrap the frontier isoquant: some intuitions.

provides a clue about the unknown true efficiency relative to the true frontier somewhere below the original non-parametric isoquant. We apply the same method to estimating cost efficiency ratios below.

A second innovation is the application of the bootstrapping methodology to remedy problems related to explaining efficiency patterns in a second stage analysis (see Simar & Wilson, 2007). A standard approach in the transport cost literature is to use "environmental" variables (such as subsidy regulations, contract types, etc.), over which the evaluated transit firm is assumed to have no control, to explain estimated inefficiencies. Although widely used, there are three potential problems with this approach. First, any two-stage approach uses an estimate of the efficiency score as a dependent variable in the second stage. But since the frontier efficiency scores are biased, it is useful to construct bias-corrected estimates, as explained above. Second, most applications of two-stage approaches have relied on Tobit regression to estimate the impact of environmental variables (contract type, etc.) on efficiency at the second stage. However, because of the intrinsic bias in the non-parametric frontier estimates, we can be quite confident that transit operators that are estimated to be inefficient are indeed inefficient, but we can have much less confidence in the status of operators that are estimated to be on the frontier (and hence pronounced efficient). The Tobit approach does value these efficient observations as such, setting their efficiency score equal to one. Truncated regression may therefore be more appropriate, since it concentrates on the inefficient observations solely. Third, since the individual efficiency scores depend on other observations on the frontier, the dependent variable is serially correlated in an unknown way. The efficiency scores are not independent and, since inputs and outputs are correlated with the environmental variables, the error term of a second stage regression of efficiency scores on environmental variables is correlated with the environmental variables as well. While both correlations disappear asymptotically, the slow rate of convergence makes conventional inference invalid in small samples.

Simar and Wilson (2007) suggest a bias-correction and a bootstrap on the second stage to arrive at consistent parameter estimates. The first algorithm is a bootstrap variant of the truncated regression on the efficiency scores obtained at the first stage; the second algorithm comprises an intermediate bias-correction before a final bootstrapped truncated regression is executed. Though they are theoretically equivalent in large samples, the second algorithm is more reliable in small samples.

Finally, a third innovation is related to the difficulty of comparing efficiency estimates resulting from different studies. For example, imagine one compares average cost efficiency estimates from two countries with identical non-parametric specifications of

[7] Since the input-oriented scores are bounded from above by unity, a reflection technique (Silverman, 1986) needs to be applied to obtain a proper kernel density.

the cost frontier that differ only in sample size, and assume one country using competitive tendering procedures while the other does not. Suppose one finds that the country with the smaller sample enjoys higher cost efficiency and employs competitive tendering, this does not tell us anything about the potential impact of such a policy because the bias of the non-parametric frontier estimates depends, among others, on sample size.

To allow comparability, several methods have been proposed. A non-exhaustive overview of some of these methods follows.[8] The first method is simply to estimate a common frontier for samples from several countries. This requires an identical specification and perfect similarity in data, which are rarely available in practice. This solution has to the best of our knowledge not been applied in a transit context. Second, another solution is to use meta-analysis and to regress the efficiency estimates from a variety of studies to a set of control variables, including characteristics of the samples and of the specification used, together with environmental and policy variables. The trouble is that it is not clear to which extent a traditional meta-regression can accommodate variations between estimates resulting from radically different estimators (e.g., parametric vs. non-parametric frontiers). One example in the transit setting is the study by Brons, Nijkamp, Pels, and Rietveld (2005). Finally, a third method to compare results from different samples is due to Zhang and Bartels (1998). They demonstrated that average efficiency in a non-parametric model decreases both in the number of observations and in the number of dimensions included. They argue in favour of a Monte-Carlo-type approach, limiting the size of larger samples to the size of the smaller samples, to obtain average sample efficiencies that are comparable across samples. In a more or less similar way, it is almost always possible to handle the fact that different models use different numbers of parameters by adjusting the number of observations in the samples accordingly. Notice that this method does not correct for bias, but simply ensures that results share a similar degree of bias. Again, we are unaware of any application in a transit context.

To conclude this brief methodological overview, note that additional innovations have been developed. We mention two of them. First, some new estimators focus on "partial" rather than traditional "full" frontiers enveloping all data. Instead of trying to estimate the absolute lowest technically (or allocatively) achievable input for a given level of output, the goal is just to obtain a *robust* estimate that is "rather close" to these optimal quantities by focusing on a local reconstruction of technology and cost function (see Simar & Wilson, 2008). So far, two families of partial frontiers are available: (i) order-m frontiers where m functions like a trimming parameter defining the number of observations for which a local frontier is estimated, and (ii) order-α quantile frontiers where the α-parameter is analogous to a quantile regression function. Estimating partial frontiers avoids many statistical problems plaguing full frontier estimators (e.g., these enjoy standard parametric rates of convergence). Second, note that the frontier methodology assumes that observed inputs and outputs are accurately measured. However, data can be contaminated by errors-in-variables (e.g., accounting data can generate a flawed view on economic value because of a questionable depreciation scheme). Since frontiers rely on comparisons among extreme observations, efficiency results are very sensitive to such errors: in fact, even a single outlier can substantially affect the outcomes for any given sample. This errors-in-variable problem is different from the impact of small sample bias and has been analyzed in Kneip and Simar (1996) and Post, Cherchye, and Kuosmanen (2002), among others.

A final remark is in order. It is obvious that the availability of statistical inference for these non-parametric frontier estimators has repercussions for other methods that are also based on these frontier estimates. For instance, the Malmquist productivity index uses input and output information solely and allows disentangling frontier change and technical efficiency change (see Färe, Grosskopf, Lindgren, & Roos, 1994 for the seminal article and Boame & Obeng, 2005 for a transit study). Thus, it is also possible to employ bootstrapping when assessing productivity change (see, e.g., Odeck, 2006 for a recent urban transit application). However, these extensions are not further discussed in this chapter.

3. Non-parametric cost estimation with bootstrapping

Having briefly reviewed recent developments in non-parametric approaches to efficiency measurement, we are now ready to explain in more detail the tools that are used in the empirical analysis below.

3.1. Basic non-parametric cost frontiers

Assuming there are K observations in the sample, a unified algebraic representation of the convex technologies with various returns to scale assumptions is:

$$L^{\Lambda}(y) = \left\{ x : x \geq \sum_{k=1}^{K} x_k z_k, y \leq \sum_{k=1}^{K} y_k z_k, z_k \in \Lambda \right\} \tag{8}$$

where $\Lambda \in \{CRS, VRS\}$, with (i) $CRS = \{z_k \in \Re_+^K : z_k \geq 0\}$, and (ii) $VRS = \{z_k \in \Re_+^K : z_k \geq 0, \sum_{k=1}^{K} z_k = 1\}$, and z is the activity vector. These technologies basically impose strong input and output disposability and either constant (CRS) or variable returns to scale (VRS). Computing a cost function amounts to solving for each observation in the sample the following optimization program defined relative to the above technologies:

$$\begin{aligned} C(y, w) = \min_{x,z} & \sum_{n=1}^{N} w_n x_n \\ \text{s.t.} & \sum_{k=1}^{K} y_{km} z_k \geq y_{km}^o \quad m = 1, \ldots, M, \\ & -x_n + \sum_{k=1}^{K} x_{kn} z_k \leq 0 \quad n = 1, \ldots, N, \\ & z_k \in \Lambda, x_n \geq 0, \quad k = 1, \ldots, K, \end{aligned} \tag{9}$$

where $\Lambda \in \{CRS, VRS\}$. These linear programming models have become common knowledge.

3.2. Bootstrapping efficiency scores and developing test statistics

As argued above, applying the homogeneous bootstrap boils down to using smoothing techniques to approximate the distribution of the efficiency scores, and repeatedly constructing samples of pseudo-data to estimate bootstrap efficiency scores. The algorithm to derive bias-corrected efficiency scores is described in detail in Simar and Wilson (1998). An overview of the algorithm is in Appendix 1. Here, we only point out the essentials of the approach.

The technical efficiency estimates $\widehat{\theta}_k$, their unknown true values θ_k, and the bootstrap estimates $\widehat{\theta}_k^*$ are related in the following way:

$$\left(\widehat{\theta}_k - \theta_k\right) | \mathcal{S} \overset{\text{approx.}}{\sim} \left(\widehat{\theta}_k^* - \widehat{\theta}_k\right) \Big| \mathcal{S}^* \tag{10}$$

where $\mathcal{S}$ and $\mathcal{S}^*$ indicate the initial sample and the bootstrap sample of pseudo-data, respectively, and $\widehat{\theta}_k^*$ is a bootstrap estimate of the efficiency for observation k. Expression (10) means that the relation

[8] Indeed, other methodologies have been proposed (see, e.g., Atkinson & Wilson, 1995).

of the original estimate $\widehat{\theta}_k$ to the true value θ_k can be approximated by the relation between the bootstrapped estimate $\widehat{\theta}_k^*$ and the original estimate $\widehat{\theta}_k$. Therefore, the bias of the convex non-parametric (DEA) estimator in the general setting, $\text{bias}_{\mathcal{S},k} = E_{\mathcal{S}}(\widehat{\theta}_k) - \theta_k$, can be estimated by its bootstrap counterpart $\widehat{\text{bias}}_{\mathcal{S}^*,k} = E_{\mathcal{S}^*}(\widehat{\theta}_k^*) - \widehat{\theta}_k$. Hence, bias-corrected estimates $\tilde{\theta}_k$ can be obtained by applying the correction $\tilde{\theta}_k = \widehat{\theta}_k - \widehat{\text{bias}}_k = 2\widehat{\theta}_k - \overline{\theta}_k^*$, with $\overline{\theta}_k^* = B^{-1}\sum_B \overline{\theta}_k^*$.[9]

Finally, the bootstrap enables researchers to test certain hypotheses concerning the specification of DEA models. One such test is the test of *CRS* vs. *VRS* introduced in Simar and Wilson (2002). A test for the hypothesis of *VRS* against *CRS* may be carried out at the level of the individual observation, or globally at the level of the technology. Results for the individual level may be interesting for firms themselves, but for purposes of regulation or optimization of the transit system we prefer testing at the level of the technology itself. Several test statistics are available. We chose the average ratio between the means of the overall efficiencies (cost ratios) computed over all observations as a test statistic: $\widehat{S}^{CRS} = K^{-1}\sum_{i=1}^{n}\widehat{OE}_i^{CRS}(x,y,w)/\widehat{OE}_i^{VRS}(x,y,w)$. The hat (^) symbol indicates an estimate. Critical values for this test at a given significance level can be obtained from the corresponding percentile of the bootstrap distribution for this test statistic (using the bootstrap method described above).

3.3. Bootstrapping a second stage explanatory analysis of efficiency

As previously explained, the standard second stage of an efficiency analysis, in which efficiency differences between observations are explained on the basis of a set of environmental variables, suffers from various deficiencies. Simar and Wilson (2007) do suggest a bias-correction and a bootstrap on the second stage to arrive at consistent parameter estimates. The method assumes that the sample observations (x_i,y_i,z_i) are realizations of *i.i.d.* random variables (X,Y,Z) with probability density function $f(x,y,z)$ where, as before, the x_i are inputs, the y_i are outputs, the z_i are environmental variables, and observations are indexed by i. The probability density is assumed to have support over $T \times R^r$ where $T = \{(x,y)|x$ can produce $y\}$ is the production possibility set and r is the index for the environmental variables. The assumption regarding the probability that $(X,Y) \in T$ is $\Pr((X,Y) \in T) = 1$.

The relation between efficiency and environmental variables is assumed to be linear, $\delta_i = z_i\beta + \varepsilon_i$ where ε_i is a random *i.i.d.* variable independent of z_i, β is a parameter vector, and $\varepsilon_i \sim N(0,\sigma)$ with right-truncation at $1 - z_i\beta$. A separability assumption between the space of inputs and outputs on the one hand and the environmental variables on the other is implied by these assumptions. To overcome the problems with standard estimation procedures, the regression parameters are estimated by truncated regression with a bootstrap method. A prior round of bootstrapping non-parametric frontier efficiency scores is applied to arrive at bias-corrected estimates. The algorithm to be carried out is given in Appendix 2. It is employed to derive the empirical results presented below in Section 4.4.

3.4. Monte-Carlo simulation to alleviate the effects of different sample sizes

Zhang and Bartels (1998) demonstrate that average efficiency decreases as the number of observations in a model increases for a given number of variables used. They argue in favour of a Monte-Carlo-type approach limiting the size of larger samples to the size of the smaller samples to derive average sample efficiencies that are comparable across samples. We use their approach to compare efficiency scores derived for two samples of transit firms from two different countries, where the samples differ substantially in size. We draw random sub-samples (without replacement) from the larger data set of urban transit companies for, say, country A. The sub-samples match the size of the smaller samples from country B. By using a sufficiently large number of replications and averaging over the results we obtain the expected efficiency for larger samples if only a smaller data set had been available. In this way, we are able to separate the sample size effect from efficiency differences across countries.

Notice that the Zhang and Bartels (1998) method provides no correction for bias in a technical sense. It simply ensures that the results share a similar degree of bias. Note also that the application of this approach artificially limits the precision of the estimates. Indeed, reducing the number of observations decreases the level of precision to the one for the sample with the smallest size. Thus, the gain in one desirable property –increased comparability– comes at the loss of another desirable property –the overall precision of the estimates. Furthermore, the Zhang and Bartels (1998) approach only remedies differences in sample size for models with an identical number of parameters. Since these non-parametric estimators have a convergence rate that is inversely related to the number of model parameters (e.g., Kneip , Park & Simar, 1998), the bias increases with the number of parameters. To maintain precision of estimates when parameters are added to a given model, the number of observations must increase considerably. Simulation results by, e.g., Pedraja-Chaparro, Salinaz-Jiménez, and Smith (1999) support the theoretical results obtained by Kneip et al. (1998) that the number of observations must ideally double for each parameter added to a specific model to retain the same precision level for the estimates. Thus, to deal with the fact that different models are estimated using different numbers of parameters, it is always possible to adjust the number of observations in the samples accordingly.

An alternative for adjusting the number of observations is to simply drop some parameters from the models containing the higher number of parameters, or to aggregate some parameters into a single one. However, Orme and Smith (1996) demonstrate that dropping a parameter that is highly correlated with another parameter from the model or dropping a parameter that is basically uncorrelated with the rest of the parameters may generate quite different effects on the results. Therefore, it is not obvious how dropping or aggregating parameters contributes to the solution of the underlying problem.

4. Empirical application

In this section, we first apply the bootstrapping methodologies described before to a set of data on Norwegian transit firms. These data have been repeatedly used before (see Jørgensen, Pedersen, & Solvoll, 1995; Jørgensen et al., 1997; Holvad, Hougaard, Kronborg, & Kvist, 2004). This facilitates comparisons with previous results which, given the methodological nature of our contribution, is useful in itself. Second, to illustrate how one can compare efficiency estimates for samples of transit firms from different countries, we also employ a small sample of French urban transit data, earlier analyzed in Gagnepain and Ivaldi (2002a, 2002b), and compare it with the Norwegian results.

4.1. Data description: Norwegian transit operators

The data were derived from official reports from the bus companies to the county councils for the calendar year 1991. All 175

[9] For a bias-correction to be an improvement upon the standard estimates, the bias-corrected estimates should not have a mean square error (MSE) larger than the standard estimates. For this to be the case, the condition $\text{bias}_k^2/3 > \text{var}(\theta_k^*)$ must hold (see Simar & Wilson, 2000).

subsidized Norwegian bus companies providing local bus services in that year are contained in the initial database. However, quite a few companies were discarded due to extreme observations or missing data for key variables. For instance, four companies seemed to have reported inaccurate data, while six other companies operated under special conditions in reference to the other companies in the database (for instance, one of these is the main bus operator in Oslo, another is a small company with very low costs because some routes are served by hired taxi cabs). In the end, 154 observations were used.

Table 1 shows descriptive statistics for the final data set comprising 154 Norwegian local bus transport companies for the year 1991. Several remarks on the data set are in order. First, note that company size varies considerably. Second, it is clear that, consistent with much of the earlier literature, several variables describe the output characteristics of the bus companies. The bus services provided are captured by different measures, including vehicle-kilometers, passenger-kilometers and the number of seat kilometers. The latter correct the kilometers driven by the number of seats on the individual buses. In other words, seat kilometer captures differences in bus size capacity. This turns out to be the preferred output specification in the models estimated below. Third, the table provides three cost components (adding up to total cost), namely the outlays for fuel, for bus drivers, and expenditures on other inputs (the latter including depreciation). All financial variables are measured in Norwegian Kroner (NOK). Finally, the data contain a series of operating characteristics and environmental variables for all companies, including variables describing the rolling stock and the area they operate in. Density pertains to the population density, which varies by a factor of about 100. The dummy D1 indicates whether the company engages in sea transport or not, which is the case for about 10% of the companies. The dummy D2 relates to its operation in a coastal area, which is a characteristic of nearly half of the companies. The next three dummies describe contract types and subsidy arrangements: (i) H1 indicates a bus company that is publicly owned and faces a subsidy policy based on a cost norm or not; (ii) H2 represents a bus company that is privately owned and has the ability to negotiate with the county council over the subsidy size or not; and (iii) H3 indicates a bus company that is privately owned and facing a subsidy policy based on a cost norm or not.

Note that, with the exception of passenger-kilometers, the outputs available are "pure supply" output indicators (e.g., seat kilometers, vehicle-kilometers). Some authors have argued that costs should be related to pure supply indicators, because it is supply that directly causes transit costs. However, it is now widely believed that the complexity of transit firms' objectives and the heterogeneity of transport output in terms of temporal, spatial, and quality attributes imply that supply characteristics should be accompanied by output attributes. For example, companies may operate a highly dense or a sparse network, they may differ in terms of peak-to-base ratios, and their services may differ in quality (as reflected in, e.g., speed, punctuality, frequencies, travel linkages, cleanliness of vehicles, drivers' attitudes). Although the data do not contain proxies for many of these characteristics, we do have information on density, several operating characteristics (coastal area or not, etc.) and the regulatory environment.[10] While these variables will not be used to construct the frontier, they are clearly relevant in the second stage when explaining variations in efficiency.

Notice that Norway has quite a history in the estimation of cost and efficiency models for transit firms. For example, Jørgensen et al. (1995) and Jørgensen et al. (1997) estimate the average cost per vehicle kilometer, using both the number of vehicle-kilometers and the number of passengers per vehicle-kilometer as outputs. They find very mild economies of scale for increases in vehicle kilometers. These authors also provide estimates of efficiency, using the standard two-stage procedure: it first estimates efficiency scores based on the cost frontier; next, these scores are explained in a separate regression analysis, using contract type and ownership as the main explanatory variables. It is found that the standard cost norm contract improves cost efficiency over the individually negotiated contract by between 1.7% and 3.5%, depending on the distributional assumptions made for the inefficiency term. Holvad et al. (2004) expand the analysis: they identify two types of vehicle kilometers as outputs (supply in an urban and in a regional environment) and also obtain mild economies of scale. Dalen and Gómez-Lobo (2003) estimate a cost frontier for an 11-year panel of Norwegian bus companies (1136 company-year observations) using the methodology proposed by Battese and Coelli (1995). Finally, Odeck and Alkadi (2001) apply non-parametric frontier analysis to examine the performance of Norwegian bus companies using data from 1994.

Importantly, the database only contains information on the total expenditures on three inputs: expenditures on drivers, on energy and on other inputs. Of course, the computation of a cost function necessitates the availability of information on both outputs and input prices. However, it has been argued by previous users of this data set that it is not an unreasonable working hypothesis to assume that, since all firms have access to the same input markets, they face the same prices (Jørgensen et al., 1997). Under the hypothesis that all firms indeed face the same input prices, Färe, Grosskopf, and Lee (1990) demonstrate that the optimal costs calculated with prices or without remain identical. Denoting by $C_n = w_n x_n$ the cost of input n, computing a cost function without input prices now amounts to solving for each observation in the sample the following optimization program:

Table 1
Descriptive statistics Norwegian bus operators (1991).

Variable	Mean	Std. Dev.	Min	Max
Alternative outputs				
Vehicle-kilometers	1,633,445	1,822,368	16,037	8,863,117
Passenger-kilometers	1.74E+07	2.65E+07	62,210	2.08E+08
Seat kilometers	9.31E+07	1.17E+08	460,800	6.20E+08
Cost components				
Fuel costs	1,571,908	1,902,667	16,151	9,775,000
Other costs	1.23E+07	1.60E+07	113,646	9.61E+07
Driver costs	9,392,579	1.29E+07	64,000	7.21E+07
Network characteristics & environmental variables				
Population density	37.2857	48.9299	2	194
D1: sea transport	0.0974	0.2975	0	1
D2: coastal area	0.4740	0.5009	0	1
H1: Public & subsidy based on cost-norm	0.0909	0.2884	0	1
H2: Private and subsidy negotiable	0.3377	0.4745	0	1
H3: Private and subsidy based on cost-norm	0.4286	0.4965	0	1

[10] In parametric applications, the relevance of output quality characteristics and operating attributes has induced Spady and Friedlander (1978) to suggest the use of hedonic aggregators to correct the generic output vehicle-kilometers for variations in spatial, temporal, and quality characteristics. Work by Jara-Díaz and his collaborators (see the seminal paper Jara-Díaz, 1982; recent developments include, among many others, Basso & Jara-Díaz, 2005) focuses on the effect of the network structure for the proper specification of measures of economies of scale and (spatial) scope. For non-parametric technologies, similar techniques are in principle applicable. However, if a large number of additional attributes are thought to be relevant, the nature of the non-parametric approach implies that a very large number of observations used in constructing the frontier will be situated on the frontier. This undermines the discriminatory power of the analysis, and using this frontier to estimate the cost structure and to determine efficiency of individual operators may become difficult. We therefore stick to the single output model, using seat kilometers as the preferred indicator.

$$\begin{aligned}
&C(y,w) = \min_{C,z} \sum_{n=1}^{N} C_n \\
&\text{subject to} \\
&\sum_{k=1}^{K} y_{km} z_k \geq y_{km}^{o} \quad m = 1,\ldots,M, \\
&-C_n + \sum_{k=1}^{K} C_{kn} z_k \leq 0 \quad n = 1,\ldots,N, \\
&z_k \in \Lambda, C_n \geq 0, \quad k = 1,\ldots,K,
\end{aligned} \tag{11}$$

where $\Lambda \in \{CRS, VRS\}$. This leads to some modifications in the bootstrapping procedures outlined above. These are taken into account in the empirical analysis below.

4.2. Basic cost efficiency estimates

For purposes of comparison, we not only estimated a cost frontier, but also a production frontier that measures technical efficiency in an input-orientation. As argued above, the cost frontier methodology differs somewhat from the estimation of the standard cost models, due to the assumption of input prices identical across observations (see mathematical program (11)). Minimal costs divided by the observed costs provide a cost ratio as an indicator of overall efficiency. A value of unity implies full efficiency. The production frontier uses the cost for fuel, drivers and other resources as inputs and the seat kilometers as the single output.[11] Application of the model implies a search for the minimal radial reduction of all inputs that is feasible for each company. For efficient companies, the reduction is zero and their efficiency score is unity. For inefficient companies, the score below unity indicates the fraction of its current input use that would lead it to be efficient.

Table 2 presents some summary results for two versions of each model: one resting on the assumption of *CRS*, the other resting on *VRS*. Note that these are the results obtained without any correction for the bias referred to above. This correction is discussed below. Before we do so, however, we briefly comment on the results presented in Table 2. The difference between the *CRS* and the *VRS* versions of the respective models amounts to almost 8% for the cost efficiency ratio (mean efficiencies are 57% and 65%) and 9% for the radial efficiency measure based on the production frontier (mean efficiencies are 63% and 72%, respectively). The cost efficiency ratios are—as expected—somewhat lower, since overall efficiency is a more demanding criterion. These rather low levels of relative efficiency indicate considerable heterogeneity with respect to the efficiency with which individual bus companies provide their services. Given the heterogeneity of the data presented above, this is not that surprising.

For the cost efficiency model imposing *VRS*, we only found seven observations to be fully efficient. Under *CRS*, only one firm could provide a benchmark of excellence. The number of efficient firms is higher under the technical efficiency criterion. It is important to check later on to which extent this efficiency status is robust when correcting for bias.

In Fig. 2, we summarize the efficiency results for the *VRS* cost frontier using a Salter diagram. On the vertical axis are the cost ratios obtained with the *VRS* specification. They are sorted in ascending order, from the least efficient operator to the efficient ones. On the horizontal axis we represent the cumulative number of seat kilometers calculated for these sorted operators. The gaps between the vertical lines are proportional to the seat kilometers of the respective individual operators. Wide gaps, or wide bars, therefore represent operators providing a comparatively large number of seat kilometers.

Close inspection of this Salter diagram in Fig. 2 reveals some interesting information. First, it is immediately clear that quite a few of the smaller companies are experiencing low cost efficiency ratios.[12] To see this, note that the left part of the figure, the most inefficient operators, consists to a very large extent of operators offering relatively small numbers of seat kilometers (small gaps between vertical lines). There are exceptions, of course. For example, efficiency scores of less than 0.6 are observed for several firms offering high numbers of seat kilometers. Second, although some of the larger firms in terms of seat kilometers are among the more efficient ones (many of the wider gaps on the figure are situated towards the right-hand-side), quite a few small operators have efficiency scores of 0.8 and more. In fact, the correlation between inefficiency and firm size is quite low, less than 0.2.

Table 2
Basic comparison of cost and production efficiency results.

Variable	Mean	Std. Dev.	Min	Max	# Efficient Obs.
Cost frontiers $OE_i(x,y,w)$					
CRS	0.5712	0.1570	0.2279	1	1
VRS	0.6508	0.1639	0.2882	1	7
Production frontiers $TE_i(x,y)$					
CRS	0.6341	0.1530	0.2526	1	3
VRS	0.7235	0.1588	0.3618	1	15

4.3. Bootstrapped cost efficiency estimates: estimating the efficiency bias and testing for returns to scale

As outlined above, the results reported in Table 2 and illustrated in Fig. 2 are potentially biased. To analyze the bias involved in these estimates, we ran bootstrap routines for the cost efficiency model, using the techniques outlined above (Section 3.2). To save space, we only report results on the *VRS* version of the cost efficiency model.

The importance of correcting for the inherent bias in non-parametric efficiency scores is best illustrated using Fig. 3. The distribution on the right of the figure is the density of the uncorrected estimates; the corresponding distribution of the bias-corrected estimates is situated on the left part of the figure. Compared to the original estimates and reflecting the size of the bias, the density indeed shifts remarkably to the left. We found that the average bias amounted to 25.20% for the *VRS* case (and 23.73% for the *CRS* model). This suggests that the bias is large. This comes as no real surprise, given the data heterogeneity.

Since presenting more detailed results on each of the individual 154 observations is not very informative, we restrict the representation to a small selection of specific observations. We give results for the most efficient observations (the seven fully efficient ones) and for the five most inefficient observations. In addition, we report the results for the five observations located closest to the 25th, 50th and 75th percentiles. This yields 27 observations all together. The information for these 27 observations is summarized in Table 3 and in the box plots in Fig. 4.

Considering Table 3, the observations are sorted on the original cost efficiency ratio. The table reports the uncorrected estimate, the magnitude of the bias, and the confidence interval for the

[11] Solving for technical efficiency in this case amounts to:

$$\begin{aligned}
DF_i(x,y) = \min_{\theta,z} \theta \quad \text{subject to} \quad &\sum_{k=1}^{K} y_{km} z_k \geq y_{km}^{o} \quad m = 1,\ldots,M, \\
&\sum_{k=1}^{K} C_{kn} z_k \leq \theta C_{kn}^{o} \quad n = 1, \\
&\ldots,N, z_k \in \Lambda, \theta \geq 0, \quad k = 1,\ldots,K,
\end{aligned}$$

where $\Lambda \in \{CRS, VRS\}$.

[12] This is consistent with the findings of Dalen and Gómez-Lobo (2003) and Odeck and Alkadi (2001). Notice, though, that Jørgensen et al. (1995) as well as Jørgensen et al. (1997) obtain almost constant returns to scale in Norwegian transit.

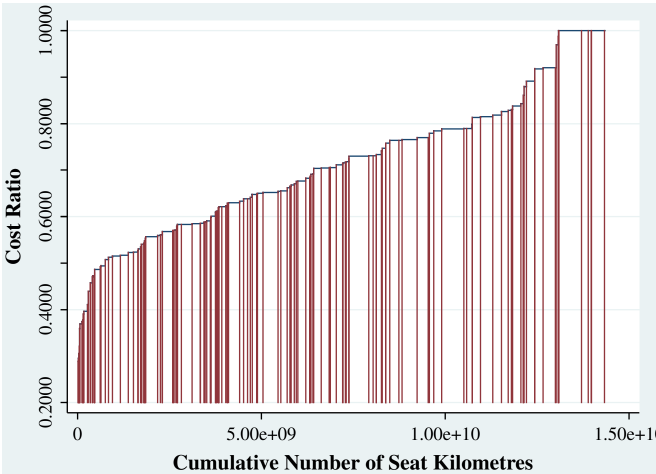

Fig. 2. Salter diagram of cost efficiency (*VRS* model).

bias-corrected estimate. Several observations stand out from the table. First, it confirms that for most observations the bias is large. Second, the width of the confidence intervals indicates that the bias-corrected efficiency estimates are imprecise. Moreover, in many cases the uncorrected estimate is not within the standard confidence interval for the bias-corrected estimate. This underscores the risk of using uncorrected estimates. Third, it is interesting to specifically look at the results for the seven observations that were pronounced efficient, based on the uncorrected estimates. We see that for five out of the seven observations, the value of 1 is outside the confidence interval for the bias-corrected estimates. In other words, bootstrapping techniques suggest that these transit operators are not fully efficient after all; these just seemed to be efficient, given the size of the sample available for the analysis. Two observations are potentially efficient; the unit value is contained in the bias-corrected confidence interval.

Fig. 4 has similar information, but presented in a box plot diagram. Here the observations are sorted by the 50% value of the distribution of bootstrap scores. Note that for only two observations (the two observations on the far right of the box plot) the upper whisker is above the line for the value 1. This corresponds to the two above-mentioned observations in Table 3 for which the upper bound of the confidence interval (97.5%) overlaps the value of one (observations 54 and 96).

Table 3
Bootstrap results for selected observations.

	Cost Ratio	Bias	2.5%	97.5%
1st Percentile				
64	0.2882	0.1051	0.1444	0.2299
50	0.2913	0.1102	0.1422	0.2311
22	0.2952	0.1050	0.1547	0.2324
103	0.3053	0.1123	0.1517	0.2447
100	0.3167	0.1154	0.1592	0.2550
25th Percentile				
86	0.5483	0.2139	0.2311	0.4668
39	0.5532	0.2331	0.2491	0.4183
24	0.5565	0.2234	0.2606	0.4303
56	0.5597	0.1122	0.3628	0.5624
57	0.5611	0.2514	0.2341	0.4169
50th Percentile				
110	0.6333	0.1522	0.3973	0.5936
72	0.6383	0.2859	0.2589	0.4819
91	0.6386	0.2857	0.2610	0.4772
79	0.6415	0.2868	0.2667	0.4768
138	0.6474	0.2868	0.2753	0.4810
75th Percentile				
126	0.7509	0.3358	0.3075	0.5592
43	0.7577	0.2896	0.3450	0.6339
52	0.7584	0.2443	0.3516	0.7310
12	0.7641	0.3421	0.3107	0.5751
90	0.7643	0.2860	0.3737	0.6350
100th Percentile				
8	1.0000	0.3221	0.4636	0.9638
10	1.0000	0.4455	0.3987	0.8001
16	1.0000	0.3684	0.4962	0.8024
54	1.0000	0.1278	0.6950	1.0824
96	1.0000	0.2080	0.8203	1.8481
118	1.0000	0.3505	0.5331	0.7960
149	1.0000	0.4464	0.4147	0.7434

Finally, we can use the results to test whether the hypothesis of constant returns to scale is tenable. This can be done along the lines of several test procedures introduced by Simar and Wilson (2002). Although tests at the level of the individual operator are also available, we concentrate on a test at the global level. Specifically, our test statistic is the ratio of the average efficiency of the *CRS* model and the *VRS* model. The ratio of average *CRS* efficiency over average *VRS* efficiency is 0.8869. Therefore, the null hypothesis that the average *CRS* efficiency score does not significantly differ from the average *VRS* efficiency measure can be rejected at any conventional level of significance. The critical 1% (5%, respectively,

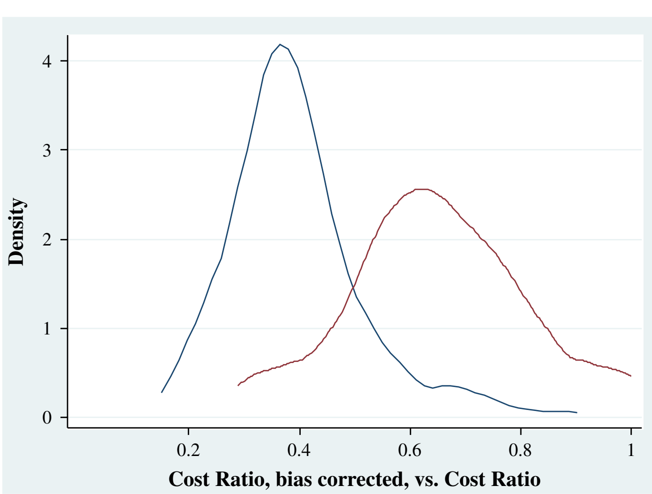

Fig. 3. Density of cost efficiency (*VRS* model): original estimates (left distribution) and bias-corrected estimates (right distribution).

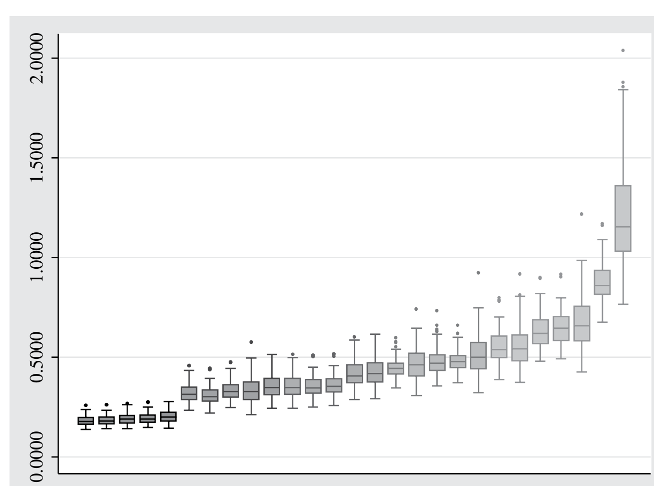

Fig. 4. Box plot of 1st, 25th, 50th, 75th and 100th percentiles of *VRS* cost efficiency.

Table 4
Two-stage results on network characteristics and environmental variables.

Variable	Parameters	2.5%	97.5%
Population density	0.37E−04	−0.0005	0.0009
D1	−0.0703	−0.0979	0.1164
D2	0.1145	0.0820	0.2109
H1	−0.0160	−0.1112	0.1696
H2	−0.0968	−0.1745	0.0238
H3	−0.0202	−0.0992	0.0879
Constant	0.6279	0.6168	0.7838

10%) values for this one-sided test are 0.7878 (0.8267, respectively, 0.8416). This result is in line with the previously mentioned literature on Norwegian transit.

4.4. Bootstrapped cost efficiency estimates: explaining efficiency patterns

In this section, we turn to the question whether any of the variables that were not considered in the estimation of efficiency scores (namely the variables characterizing the area in which the companies operate, etc.) are significantly related to the level of relative efficiency. In line with earlier studies (Jørgensen et al., 1997), the second stage model includes the following variables into a linear specification: population density, the dummy variables characterizing the operating environment (operating in sea transport (D1) or in a coastal area (D2)), and the contract types (H1–H3).

We applied the methodology outlined in Section 3.3 to obtain bootstrapped second stage parameter estimates. In Table 4 we provide the estimates along with their confidence intervals.

Since we are simply testing whether a parameter estimate is significantly different from zero, we must look in Table 4 for parameters where the confidence interval does not cover the zero (i.e., lies on either side of zero). Hence, apart from the intercept, the only variable that is significantly associated with the variations in the cost ratio seems to be the dummy variable D2. Thus, operating in a coastal area ($D2 = 1$) has a positive impact on overall efficiency: these bus operators have on average a cost ratio that is 11.5% higher than other operators. No other variable seems to explain the rather wide dispersion in overall efficiency observed within this heterogeneous sample.

Notice that the earlier studies using the same data set and parametric methods (Jørgensen et al., 1995, 1997) confirm the positive impact of operating in the coastal area and find a significant impact of some of the contract type variables (notably H1 and H3). Dalen and Gómez-Lobo (2003), using slightly different data, also report a significant coefficient of high-powered contracts. Interestingly, we do not find such results. The non-parametric approach, which attempts to minimise the number of assumptions maintained, implies very large variability in inefficiency scores. Apparently, the contractual dummy variables do not systematically contribute to explaining the wide variations in overall efficiency among the operators. The earlier results may thus be less robust than previously believed. This matter merits further investigation.

Table 5
Descriptive statistics for French bus operators (1991).

Variable	Min	Max
Fuel costs		
Driver costs	3000.0000	
Other costs	5326.0000	
Seat kilometers (in 1000)		1114.2410

Table 6
Comparing mean cost efficiency results for Norway and France.

	CRS	*VRS*
France	0.6854	0.7186
Norway	0.6027 (0.5712)	0.6954 (0.6508)

4.5. Comparing estimates from different sample sizes: Norway versus France

We finally illustrate the use of Monte-Carlo bootstrapping techniques to correct for differences in sample size when comparing two samples of transit operators from different countries. We use the Norwegian data described before together with a small French data set.

In Table 5, the relevant summary data for 55 French bus operators are reported (also see Gagnepain & Ivaldi, 2002a, 2002b). Note that the cost data (first three rows) are given in local currency (French Franc) in 1991 prices. Therefore, the cost figures for the two data sets cannot be readily compared. We limit ourselves to some brief observations. While the number of seat kilometers driven differs substantially between the two data sets, the cost shares are very similar. The highest share is the cost for drivers in both data sets, followed by other costs, while fuel costs are by far the smallest cost component. Precisely, driver cost amounts to 59% of the total cost in France and to 63% in Norway. Other costs make up almost exactly one-third of total cost in Norway whereas the corresponding share in France exceeds 35%. Finally, fuel costs amount to slightly more than 5.5% in the case of France and are close to 4.25% in Norway.

Table 6 presents results for both the French and the Norwegian data. Both data sets pertain to the year 1991. The row labeled France lists results for the standard overall efficiency (cost ratio) model introduced above. Here, the *CRS* as well as the *VRS* results seem somewhat higher than the corresponding Norwegian results from the standard model listed in parentheses in the row Norway. However, as explained above, these two sets of results cannot be readily compared. Even though both models comprise the same output, namely seat kilometers, and the same number of inputs (three cost components: fuel cost, personnel cost, and other costs), the difference in sample size makes a direct comparison impossible. Recall that the number of observations is 154 for Norway, but only 55 for France.

To correct for this difference in sample size, we repeatedly generated results for the Norwegian case based on samples of 55 observations each. This results in a Monte-Carlo analysis, whereby samples were drawn without replacement. The results are listed in the row "Norway". They are higher than the standard results listed in parentheses in the same row and therefore, as expected, closer to the results for the French data. Indeed, while the standard results make it appear as though the Norwegian bus operators are substantially less efficient than the French ones, the results based on the Monte-Carlo analysis convey a different scenario. Here, the gap between the French and the Norwegian bus operators is almost closed (less than 2.5%) for the *VRS* model, while it remains rather substantial in the case of *CRS*. However, note that the *VRS* assumption is the relevant one for Norway, as indicated by the test on returns to scale reported above (we did not test for returns to scale in the French case).

5. Conclusions

This paper exploits recent advances in statistical inference applied to non-parametric transit cost frontiers. First, the literature has convincingly shown that efficiency scores derived from non-parametric cost frontiers are inherently biased. We use bootstrapping techniques on a convex (DEA) cost frontier to correct for this bias. The proposed methodology is applied to a well-known Norwegian database that has been employed several times before in the literature. The model used to illustrate the strength of the bootstrapping methodology uses seat kilometers as the relevant output measure and takes account of three inputs (drivers, fuel, and other inputs). The bias implied by uncorrected non-parametric cost efficiency measures is found to be numerically important, amounting to close to 25%. Bootstrapping also allows us to test the hypothesis of constant returns to scale. We decisively reject constant returns to scale against an alternative model allowing for variable returns.

Second, an attempt to explain patterns of efficiency scores using the Simar and Wilson (2007) two-stage bootstrapping approach detects only one significant covariate. This contrasts notably with parametric results that earlier highlighted the positive impact of high-powered contract types. Third, we illustrate the use of bootstrapping techniques to compare inefficiencies in two samples of different size. Specifically, we study how to compare the average inefficiency obtained for the Norwegian data set with the analogous estimate for a second sample of 55 French transit firms. Using the variable returns to scale model, we find that the estimated difference in average efficiency almost disappears once the correction for sample size differences is accounted for.

Acknowledgements

We are grateful to an anonymous referee and the editors of this book for useful comments on an earlier version. A special word of thanks to T. Holvad and D. Kronborg for providing us with the Norwegian data set, and to Philippe Gagnepain for giving access to the French data set.

Appendix 1. Algorithm for homogeneous bootstrap

The algorithm for the homogeneous bootstrap of technical efficiency scores ($DF_i(x,y)$) involves the following steps:

1. Calculate the estimates $\widehat{\delta}_i$ of the efficiency scores based on a convex non-parametric model.
2. Use a suitable method to calculate the optimal bandwidth h for a kernel smoother used for the approximation of the density of the $\widehat{\delta}_i$, using an appropriate method to account for the boundary problem (at the value of 1 in the input-oriented case). Based on the estimate of the optimal bandwidth, generate B draws from the respective density, $\tilde{\delta}_{ib}$.
3. The $\tilde{\delta}_{ib}$ from step 2 allows the construction of pseudo data $y_i^* = y_i, x_{ib}^* = x_i\widehat{\delta}_i/\tilde{\delta}_{ib}$. A large number of pseudo-data sets can be used to obtain a distribution of efficiency estimates for each observation i, with the single elements $\widehat{\delta}_{ib}^*$.
4. If appropriate, carry out bias-correction.
5. The distribution obtained in step 3 makes it possible to construct confidence intervals for the estimates obtained in step 1 by selecting the appropriate percentiles from the distributions obtained.

To modify this algorithm for the case of overall cost efficiency $OE_i(x,y,w)$, we simply substitute the estimate of the technical efficiency score $\widehat{\delta}_i$ by the estimate $\widehat{OE}_i(x,y,w)$ in steps 1–3. Notice that also the generation of pseudo-data employs the estimate $\widehat{OE}_i(x,y,w)$ for the perturbation.

Appendix 2. Algorithm for a second stage analysis of efficiency patterns

The Simar and Wilson (2007) algorithm (algorithm 2, pp. 42–43) for a second stage analysis of technical efficiency scores ($DF_i(x,y)$) comprises the following steps:

1. Calculate the estimates $\widehat{\delta}_i$ of the efficiency scores based on a convex non-parametric model.
2. Use maximum likelihood to estimate a truncated regression of $\widehat{\delta}_i$ on the environmental variables to obtain estimates $\widehat{\beta}$ as well as $\widehat{\sigma}$.
3. Estimates from step 2 permit drawing error terms ε_i^* from the $N(0,\widehat{\sigma})$ distribution truncated at $(1-z_i\widehat{\beta})$, from which one can obtain $\delta_i^* = z_i\widehat{\beta}+\varepsilon_i^*$. These δ_i^* allow the construction of pseudo data $y_i^* = y_i, x_i^* = x_i\widehat{\delta}_i/\delta_i^*$. A large number of pseudo-data sets can be used to obtain efficiency estimates $\widehat{\delta}_{ib}^*$.
4. The $\widehat{\delta}_{ib}^*$ obtained in step 3 and the $\widehat{\delta}_i$ from step 1 are used to calculate bias-corrected estimates $\widehat{\widehat{\delta}}_i$.
5. A further truncated regression of the bias-corrected scores $\widehat{\widehat{\delta}}_i$ on the environmental variables gives $(\widehat{\widehat{\beta}}, \widehat{\widehat{\sigma}})$.
6. The parameter estimates from step 5 are used to draw ε_i^{**} from the $N(0,\widehat{\widehat{\sigma}})$ distribution truncated at $(1-z_i\widehat{\widehat{\beta}})$, from which $\delta_i^{**} = z_i\widehat{\widehat{\beta}}+\varepsilon_i^*$ can be computed. A final truncated regression of the δ_i^{**} on the environmental variables gives $(\widehat{\widehat{\beta}}^*, \widehat{\widehat{\sigma}}^*)$. Drawing $\varepsilon_i^{**}, \delta_i^{**}$ and carrying out the regression is repeated a large number of times, which makes it possible to derive confidence intervals for the parameters of interest.

It is straightforward to modify this algorithm for the case of $OE_i(x,y,w)$ along the lines presented for algorithm 1.

References

Atkinson, S. E., & Wilson, P. W. (1995). Comparing mean efficiency and productivity scores from small samples: A bootstrap methodology. *Journal of Productivity Analysis, 6*(2), 137–152.
Basso, L., & Jara-Díaz, S. R. (2005). Calculation of economies of spatial scope from transport cost functions with aggregate output. *Journal of Transport Economics and Policy, 39*(1), 25–52.
Battese, G. E., & Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20*(2), 325–332.
Boame, A. K. (2004). The technical efficiency of Canadian urban transit systems. *Transportation Research, 40E*(5), 401–416.
Boame, A. K., & Obeng, K. (2005). Sources of productivity change: A Malmquist total factor productivity approach. *Transport Reviews, 25*(1), 103–116.
Brons, M., Nijkamp, P., Pels, E., & Rietveld, P. (2005). Efficiency of urban public transit: A meta analysis. *Transportation, 32*(1), 1–21.
Chambers, R. G., Chung, Y., & Färe, R. (1998). Profit, directional distance functions, and Nerlovian efficiency. *Journal of Optimization Theory and Applications, 98*(2), 351–364.
Cherchye, L., & Post, T. (2003). Methodological advances in DEA: A survey and an application for the Dutch electricity sector. *Statistica Neerlandica, 57*(4), 410–438.
Chu, X., Fielding, G., & Lamar, B. (1992). Measuring transit performance using Data Envelopment Analysis. *Transportation Research, 26A*(3), 223–230.
Chung, Y., Färe, R., & Grosskopf, S. (1997). Productivity and undesirable outputs: A directional function approach. *Journal of Environmental Management, 51*(3), 229–240.
Cowie, J., & Asenova, D. (1999). Organisation form, scale effects and efficiency in the British bus industry. *Transportation, 26*(3), 231–248.
Dalen, D. M., & Gómez-Lobo, A. (2003). Yardsticks on the road: Regulatory contracts and cost efficiency in the Norwegian bus industry. *Transportation, 30*(4), 371–386.
Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Berlin: Springer.
De Borger, B., & Kerstens, K. (2008). The performance of bus-transit operators. In D. A. Hensher, & K. J. Button (Eds.), *Handbook of transport modelling* (2nd ed.). (pp. 693–714) Amsterdam: Elsevier.
De Borger, B., Kerstens, K., & Costa, Á. (2002). Public transit performance: What does one learn from frontier studies? *Transport Reviews, 22*(1), 1–38.
Färe, R., & Grosskopf, S. (2000). Notes on some inequalities in economics. *Economic Theory, 15*(1), 227–233.
Färe, R., Grosskopf, S., & Lee, H. (1990). A nonparametric approach to expenditure-constrained profit maximization. *American Journal of Agricultural Economics, 72*(3), 574–581.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1983). The structure of technical efficiency. *Scandinavian Journal of Economics, 85*(2), 181–190.

Färe, R., Grosskopf, S., & Lovell, C. A. K. (1985). *The measurement of efficiency of production*. Boston: Kluwer.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994). Productivity developments in Swedish hospitals: A Malmquist output index approach. In A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford (Eds.), *Data Envelopment Analysis: Theory, Methodology and Applications* (pp. 253–272). Boston: Kluwer.

Färe, R., & Primont, D. (1995). *Multi-output production and duality: Theory and applications*. Boston: Kluwer.

Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, 120A*(3), 253–281.

Gagnepain, P., & Ivaldi, M. (2002a). Stochastic frontiers and asymmetric information models. *Journal of Productivity Analysis, 18*(2), 145–159.

Gagnepain, P., & Ivaldi, M. (2002b). Incentive regulatory policies: The case of public transit systems in France. *Rand Journal of Economics, 33*(4), 605–629.

Gijbels, I., Mammen, E., Park, B. U., & Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association, 94*(445), 220–228.

Grosskopf, S. (1996). Statistical inference and nonparametric efficiency: A selective survey. *Journal of Productivity Analysis, 7*(2–3), 161–176.

Holvad, T., Hougaard, J. L., Kronborg, D., & Kvist, H. K. (2004). Measuring inefficiency in the Norwegian bus industry using multi-directional efficiency analysis. *Transportation, 31*(3), 349–369.

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

Jørgensen, F., Pedersen, P. A., & Solvoll, G. (1995). The costs of bus operations in Norway. *Journal of Transport Economics and Policy, 29*(3), 253–262.

Jørgensen, F., Pedersen, P. A., & Volden, R. (1997). Estimating the inefficiency in the Norwegian bus industry from stochastic cost frontier models. *Transportation, 24*(4), 421–433.

Kerstens, K. (1996). Technical efficiency measurement and explanation of French urban transit companies. *Transportation Research, 30A*(6), 431–452.

Kneip, A., & Simar, L. (1996). A general framework for frontier estimation with panel data. *Journal of Productivity Analysis, 7*(2-3), 187–212.

Kneip, A., Park, B. U., & Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory, 14*(6), 783–793.

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic frontier analysis: An econometric approach*. Cambridge: Cambridge University Press.

Luenberger, D. G. (1995). *Microeconomic theory*. New York: McGraw-Hill.

Matas, A., & Raymond, J.-L. (1998). Technical characteristics and efficiency of urban bus companies: The case of Spain. *Transportation, 25*(3), 243–263.

McMullen, B. S., & Noh, D.-W. (2007). Accounting for emissions in the measurement of transit agency efficiency: A directional distance function approach. *Transportation Research, 12D*(1), 1–9.

Nolan, J. F., Ritchie, P. C., & Rowcroft, J. E. (2002). Identifying and measuring public policy goals: ISTEA and the US bus transit industry. *Journal of Economic Behavior and Organization, 48*(3), 291–304.

Odeck, J. (2006). Congestion, ownership, region of operation, and scale: Their impact on bus operator performance in Norway. *Socio-Economic Planning Sciences, 40*(1), 52–69.

Odeck, J., & Alkadi, A. (2001). Evaluating efficiency in the Norwegian bus industry using Data Envelopment Analysis. *Transportation, 28*(3), 211–232.

Orme, C., & Smith, P. C. (1996). The potential for endogeneity bias in Data Envelopment Analysis. *Journal of the Operational Research Society, 47*(1), 73–83.

Pedraja-Chaparro, F., Salinas-Jiménez, J., & Smith, P. C. (1999). On the quality of the Data Envelopment Analysis model. *Journal of the Operational Research Society, 50*(6), 636–644.

Pels, E., & Rietveld, P. (2008). Cost functions in transport. In D. A. Hensher, & K. J. Button (Eds.), *Handbook of transport modelling* (2nd ed.). (pp. 381–393) Amsterdam: Elsevier.

Post, T., Cherchye, L., & Kuosmanen, T. (2002). Non-parametric efficiency estimation in stochastic environments. *Operations Research, 50*(4), 645–655.

Roy, W., & Yvrande-Billon, A. (2007). Contractual practices and technical efficiency: The case of urban public transport in France. *Journal of Transport Economics and Policy, 41*(2), 257–282.

Sakano, R., & Obeng, K. (1995). Re-examination of inefficiencies in urban transit systems: A stochastic frontier approach. *Logistics and Transportation Review, 31*(4), 377–392.

Sickles, R. C. (2005). Panel estimators and the identification of firm-specific efficiency levels in parametric, semiparametric and nonparametric settings. *Journal of Econometrics, 126*(2), 305–334.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science, 44*(1), 49–61.

Simar, L., & Wilson, P. W. (2000). Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis, 13*(1), 49–78.

Simar, L., & Wilson, P. W. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research, 139*(1), 115–132.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics, 136*(1), 31–64.

Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In H. Fried, C. A. K. Lovell, & S. Schmidt (Eds.), *The measurement of productive efficiency and productivity change* (pp. 421–521). New York: Oxford University Press.

Small, K. A., & Verhoef, E. T. (2007). *The economics of urban transportation*. London: Routledge.

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

Waters, W. G., II (2008). Productivity measurement. In D. A. Hensher, & K. J. Button (Eds.), *Handbook of transport modelling* (2nd ed.). (pp. 395–415) Amsterdam: Elsevier.

Zhang, Y., & Bartels, R. (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*(3), 187–204.