ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 54 – 1999

Decomposing Technical Efficiency and Effectiveness of French Urban Transport

Kristiaan KERSTENS*

ABSTRACT. – This article determines the sources of observed technically inefficient behaviour among French urban transit companies using non-parametric deterministic frontier specifications of technology. Decomposing overall technical efficiency yields component-wise efficiency measures reflecting scale, structural and technical inefficiencies. Also the effectiveness of urban transit is evaluated in a similar way. Moreover, the analysis investigates the effect of the selected orientation of measurement. Finally, it is the first study to control for the effect of outliers on the decomposition results.

The empirical results indicate that technical inefficiency is the major source of poor performance, followed by inefficiencies due to inadequacies in scale. Congestion only plays a minor role. The findings also lead us to add some critical notes to this decomposition methodology.

Décomposition de l'efficacité technique et de l'effectivité du transport urbain français

RÉSUMÉ. – Cet article cherche à mesurer l'inefficacité technique des compagnies de transport urbain françaises en utilisant des technologies non paramétriques déterministes. Dans un premier temps, nous utilisons une décomposition permettant de distinguer l'inefficacité purement technique de l'inefficacité d'échelle et de la congestion. Nous analysons l'effectivité de manière similaire. Dans un second temps, nous apprécions l'incidence sur les résultats de l'orientation de la mesure. Enfin, nous contrôlons pour la première fois l'impact des données aberrantes sur cette décomposition.

Les résultats empiriques montrent que l'inefficacité technique constitue la principale source d'inefficacité des compagnies de transport urbain françaises. A l'inverse, la congestion semble jouer peu. Cette étude amène aussi à formuler quelques réserves sur la pertinence de la méthodologie de décomposition de l'inefficacité technique.

---

* K. KERSTENS. LABORES, Université Catholique de Lille. We are grateful to two referees who provided most constructive criticisms. The usual disclaimer applies.

1 Introduction

The public sector is under pressure to show it is making efficient use of its resources. The relative efficiency of urban transit services is also the subject of recent discussions in the literature. In particular, traditional state intervention has been reassessed due to concerns about regulatory failures (see, e.g. BERECHMAN [1993]). Knowledge of the sources and causes of inefficiency in this area is useful in designing new short and long run policies.

This article focuses on the determination of sources of technically inefficient behaviour among urban transit companies in France using nowadays popular non-parametric deterministic frontier specifications of technology.[1] In addition, it evaluates the effectiveness of the sector. In particular, the empirical application distinguishes between technical and scale inefficiencies and congestion. This refined measurement points to possible causes of inferior and superior performance. Furthermore, these different sources determine the time perspective for effective policy changes. Technical inefficiencies and congestion mainly reflect managerial shortcomings and can be remedied in the short run, while scale inefficiencies require a long run viewpoint.

More specifically, this contribution accomplishes four goals. It is the first application of this decomposition methodology in transportation. Furthermore, unlike many other studies it controls for both efficiency and effectiveness of current best practice. Second, this study illustrates for the first time the potential divergence between sample level decomposition findings and individual results. This is accomplished by complementing the traditional sample level results by an analysis of size classes. Third, from a methodological viewpoint it controls for the effect of the selected measurement orientation on efficiency results and, in particular, on the qualitative returns to scale information. Finally, it is, to the best of our knowledge, the first study to assess the impact of outliers on the decomposition results. This methodological concern is warranted. VALDMANIS [1992], for instance, shows in a sensitivity analysis of technical and scale components that their relative importance may well depend on input and output dimensions included in the technology. Thus, our article also aims to critically explore the limits of a frontier methodology that recently gained quite some popularity.

The article is structured as follows. In Section 2, overall efficiency is decomposed into component-wise efficiency measures reflecting allocative, scale, structural and technical inefficiencies (FÄRE, GROSSKOPF and LOVELL [1983, 1985]). Details on the computation of this decomposition using non-parametric deterministic frontier technologies are found in Section 3. In Section 4 the data set of French urban transit operators, analysed earlier in KERSTENS [1996], is briefly introduced and supply- and demand-oriented production models are specified. All empirical performance results are presented in Section 5. A final section concludes.

1. In France, applications of these non-parametric deterministic frontier methods (also known as DEA models) have been mainly limited to agriculture (e.g., PIOT [1994]) and health care (e.g. LELEU and DERVAUX [1997]).

2 Decomposing Overall Efficiency: Concepts

Various methodologies exist to reconstruct the boundary of the production possibility set, or one of its value duals (see, e.g., LOVELL [1993]). The extensive efficiency decomposition, presented here, has only been developed for convex non-parametric deterministic technologies.[2]

First some preliminary definitions, a production technology is defined by the production possibility set containing all feasible input/output vectors: $S = \{(x, y) \mid x \text{ can produce } y\}$. The input requirement set associated with this technology denotes all input vectors x capable of producing a given output vector y: $L(y) = \{x \mid (x, y) \in S\}$. Precise definitions for the non-parametric, deterministic technologies used in the empirical part follow in Section 3.

Turning now to substantial analysis, FARRELL [1957] provided the first measurement scheme for the evaluation of technical and allocative efficiency. More recently, FÄRE, GROSSKOPF and LOVELL [1983, 1985: 3-5] offer a more elaborate efficiency taxonomy and define operational measurement procedures.[3] As their proposals have become the standard way to decompose efficiency (see, e.g., GANLEY and CUBBIN [1992]), this section presents their taxonomy.

Overall efficiency (OE) is defined as a comparison between any production combination and the situation satisfying its behavioural goal. This measure is decomposed to provide information on possible sources of inefficiency. This static decomposition differentiates between private and social goals.[4]

Among private goals – defined in the best interest of the producer – one distinguishes between technical, structural and allocative efficiency and inefficiency. Technical efficiency (TE) is defined as production on the boundary of the production possibility set. A producer is technically inefficient if production occurs in the interior of this set. Allocative (or price) efficiency (AE) requires the specification of a behavioural goal and is defined by a point on the boundary of the production possibility set satisfying this objective given certain constraints on prices and quantities. A producer is allocative inefficient if there is a divergence between observed and optimal costs, revenue, profits of whatever objective the producer is assumed to pursue.[5] Structural efficiency (STE) requires production in the uncongested or "economic" region of production. Structurally inefficiency prevails when

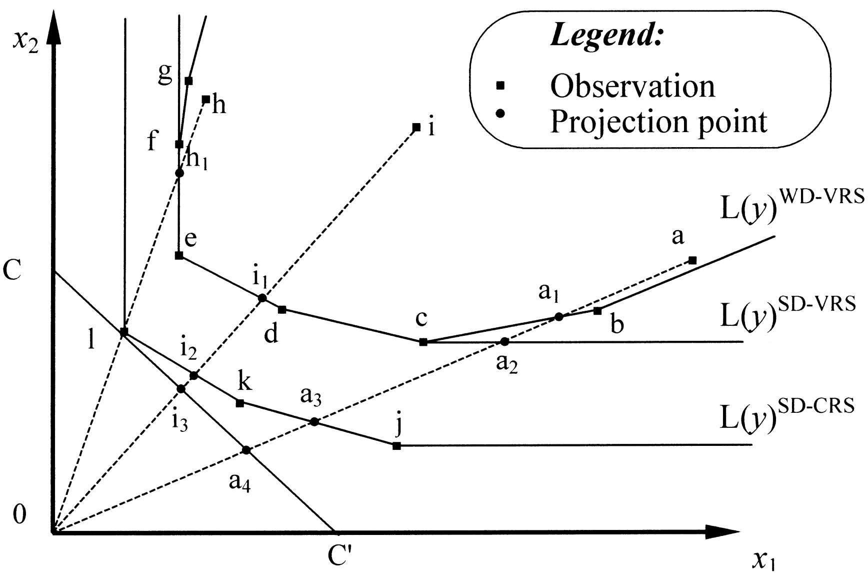
2. For parametric technologies no such detailed decomposition has yet been devised. For instance, KUMBHAKAR, BISWAS and BAILEY [1989] only explore allocative, technical, and scale inefficiencies.
3. Other classifications include BANKER, CHARNES and COOPER [1984] and FØRSUND and HJALMARSSON [1974, 1979].
4. In an intertemporal framework one can allow for technological change. FÄRE, GROSSKOPF and LOVELL [1994], for example, integrate part of this static decomposition into the development of Malmquist productivity indices.
5. This definition can even be adapted to less common, behavioural goals. For instance, FÄRE and LOGAN [1992] present a complete decomposition of efficiency for the rate of return regulated firm.

production is organised in a congested production region, i.e., where some inputs have negative marginal products.[6]

The social goal relates to a possible divergence between the actual and the ideal size of production. The ideal, long run competitive equilibrium configuration requires production at a point where constant returns to scale prevail. A productive activity is scale efficient (SCE) if its scale of production corresponds to that resulting from a long run zero profit competitive equilibrium; it is scale inefficient otherwise.

This static efficiency classification is depicted in Figure 1. It presents three input requirement sets and their boundaries all producing the same output level. The first input set $L(y)^{SD\text{-}CRS}$ is characterised by constant returns to scale (CRS) and it is uncongested. The second input set $L(y)^{SD\text{-}VRS}$ postulates variable returns to scale (VRS) and is also congestion free. The final input set $L(y)^{WD\text{-}VRS}$ assumes VRS, but allows for congestion.

FIGURE 1
Overall Efficiency: Taxonomy Illustrated on an Input Set



The efficiency concepts are illustrated by the positions of certain observations (indicated by squares). First, technical inefficiency (TE) is depicted by observation *i*. As this observation uses more of both inputs to produce exactly the same output vector as, e.g., observations *d* or *e* on the boundary of the input set $L(y)^{WD\text{-}VRS}$, it is technically inefficient. Second, structural inefficiency (STE) is represented by observation *b*. This unit is situated on a backward bending part of the boundary of the input set $L(y)^{WD\text{-}VRS}$, indicating negative marginal productivity or congestion. Third, scale inefficiency (SCE) is illustrated using observation *d*. This observation on the boundary of

6. Modelling congestion requires the assumption of weak instead of strong disposability (see FÄRE, GROSSKOPF and LOVELL [1985] for technical details).

the input set $L(y)^{SD\text{-}VRS}$ is scale inefficient because it needs more inputs to deliver the same output level as, e.g., observations *k* or *l* on the boundary of the CRS technology $L(y)^{SD\text{-}CRS}$. Fourth, allocative efficiency (AE) is discussed using observation *h*. Assuming cost minimisation, then for given input prices total costs are minimised at the intersection of the input set $L(y)^{SD\text{-}CRS}$ and the isocost line CC' (observation *1*). Observation *h* is not allocatively efficient, since it needs a higher budget to produce the same output. Firth, overall efficiency (OE) is discussed using observation *1*. This observation is at the same time allocatively efficient, scale efficient (situated on the boundary of a CRS technology), structurally efficient (not situated on a backward bending part of technology), and technically efficient (not in the interior of the input set). Obviously, all other observations can also be classified in terms of this efficiency taxonomy. For instance, observations *j* to *l* are scale efficient, while observations *a* to *c* and *e* to *i* are not.

Assuming that organisations only control inputs, it is meaningful to measure efficiency in the input dimensions. Efficiency is traditionally measured in a radial or equiproportional way (FARRELL [1957]). The radial input efficiency measure is defined as:

$$DF_i(x,y) = \min\{\lambda \mid \lambda \geqslant 0, \quad \lambda x \in L(y)\}.$$

$DF_i(x,y)$ varies between zero and one, with efficient production on the boundary (isoquant) of the input set represented by unity, and it has a cost interpretation. It can be used relative to different frontiers to operationalise the above efficiency taxonomy.

The traditional way of measuring all five static input efficiency concepts is illustrated on Figure 1 for observation *a* in the interior of input set $L(y)^{WD\text{-}VRS}$. First, technical inefficiency (TE) in the FARRELL [1957] sense is represented by the ratio of distances Oa_1/Oa relative to the weakly disposable technology $L(y)^{WD\text{-}VRS}$. Second, structural inefficiency (STE) is measured by the ratio Oa_2/Oa_1. Reducing the technically efficient vector a_1 down to point a_2 on $L(y)^{SD\text{-}VRS}$ guarantees congestion free production. Third, scale inefficiency (SCE), defined as Oa_3/Oa_2, indicates the smallest input vector a_3 able to produce the same long run output as the technically and structurally efficient combination a_2. Fourth, allocative inefficiency (AE) is captured by the ratio Oa_4/Oa_3. While the inputs of a_4 cannot yield output *y* on the boundary of $L(y)^{SD\text{-}CRS}$, for the same cost the input vector *1* is available that can produce this output level. Finally, overall efficiency (OE) is defined in terms of the ratio Oa_4/Oa.

A few remarks on this operationalisation of efficiency concepts are clarifying. First, this decomposition of static efficiency using radial efficiency measures is mutually exclusive and exhaustive. An organisation is overall efficient if it is at the same time technically, structurally, scale and allocatively efficient. A producer can be inefficient in any of these respects. For instance, observation *c* is scale and allocatively inefficient, though it is technically efficient and congestion free. Second, this taxonomy yields a multiplicative decomposition (see FÄRE, GROSSKOPF and LOVELL [1985: 188-191] for details). Overall efficiency is the product of the four composite efficiency measures of technical, structural, scale and allocative efficiency: OE = TE.STE.SCE.AE. Overall technical efficiency (OTE) differs from technical efficiency (TE) in that it is always measured relative to a CRS technology ($L(y)^{SD\text{-}CRS}$). Formally: OTE = TE.STE.SCE. The third remark is that the entire distinction between the various inefficiency sources is to some extent artificial. Theoretically in any case, production decisions are assumed to be taken jointly, not separately. The decomposition is only a conceptual tool to identify a diversity of potential sources generating the inefficiencies.[7] Furthermore, these sources imply different time perspectives regarding organisational decision-making. For instance, technical inefficiency and congestion reflect managerial failures that can be remedied in the short run, while the correction of scale inefficiencies may involve adjustments and investments in a long run perspective.

As a fourth remark, this overall efficiency decomposition presupposes that a strongly disposable CRS technology is a meaningful production model for the evaluated organisations. If this is not the case, then another technology can be selected to provide the basis for an analogous, but simplified decomposition, since one or more of its components equal unity (FÄRE, GROSSKOPF and LOVELL [1994: 81-82]). Finally, the choice of an orientation of measurement should be carefully considered. From an academic viewpoint, FÄRE, GROSSKOPF and LOVELL [1994] and BANKER, CHANG and COOPER [1996] stress that input- and output-based decompositions generate different information. Since they adopt a different perspective, they may provide different answers.[8] From a policy perspective, it is arguable that the different parties involved in organising urban transit are interested in distinct orientations. On the one hand, bus companies probably favour input-based decomposition results, as the outputs are fixed by the contract with the public organising authority. On the other hand, public authorities are perhaps eager to learn whether the public transport output can be increased for given inputs and subsidies. The latter information is helpful when renewing contracts with the companies. Since few comparisons of measurement orientations have yet been reported (for an exception, see FUKUYAMA [1996]), the empirical part illustrates the impact of both orientations.

3 Decomposing Overall Efficiency: Computational Issues

The next section illuminates the sources of inefficient production behaviour of French urban transit companies using the efficiency decomposition

7. See BAUER [1990a: 44] and PESTIEAU and TULKENS [1990: 5].
8. The radial output efficiency measure is defined as: $DF_0(x,y) = \max\{\theta \mid \theta \geqslant 1,\ x \in L(\theta y)\}$. Furthermore, an analogous decomposition is possible in terms of graph efficiency measures which simultaneously reduce inputs and expand outputs (see the formulations in FÄRE, GROSSKOPF and LOVELL [1985, 1994] and the empirical application in BANNISTER and STOLP [1995]). The graph orientation is ignored here because it is unlikely to detect much inefficiency given the small sample and given the fact that the radial graph measure is always no smaller than either the radial input or the inverse of the output efficiency measures (see FÄRE, GROSSKOPF and LOVELL [1985: 136-137]).

discussed in the previous section. Since price information is lacking, however, allocative efficiency cannot be evaluated and the analysis is restricted to decomposing overall technical efficiency (OTE).[9] This section first summarises technical details of the necessary computations (closely following FÄRE, GROSSKOPF and LOVELL [1985]). This review is limited to the input-based decomposition, because the output decomposition is completely similar. Next, the expected similarities and differences in the results from the input- and output-oriented decompositions are commented.

3.1. **Computing Overall Technical Efficiency**

The theoretical analysis is defined in terms of a series of non-parametric, deterministic input correspondences. The first technology, which is closest to FARRELL [1957], is a CRS production model with strong disposability in both inputs and outputs. Its input correspondence is constructed from observed activities in the following way:

$$\mathrm{L}(y)^{\text{SD-CRS}} = \{x \mid x \in \Re_{+}^{m},\ Y'z \geqslant y,\ X'z \leqslant x,\ z \geqslant 0\},$$

where Y is the $k \times n$ matrix of observed outputs, X is the $k \times m$ matrix of observed inputs, z is a $k \times 1$ vector of intensity or activity variables, and y and x are $n \times 1$ and $m \times 1$ vectors of outputs respectively inputs. It imposes CRS, since there is no restriction on the intensity vector z. This last assumption is easily modified to allow for non-increasing returns to scale (NIRS) and VRS by adding the constraint $I_k^t z \leqslant 1$ respectively $I_k^t z = 1$ to this definition (where I_k is a $k \times 1$ unity vector).

To detect structural inefficiencies the technology must impose a theoretical structure allowing observing congested production regions. This is accomplished by imposing weak instead of strong disposal on the CRS input correspondence:

$$\mathrm{L}(y)^{\text{WD-CRS}} = \{x \mid x \in \Re_{+}^{m},\ \mu Y'z = y,\ X'z = \delta x,\ \mu,\delta \in (0,1],\ z \geqslant 0\}.$$

As above, the CRS assumption can be altered.

Efficiency is calculated with respect to these technologies by means of linear programming (LP) techniques (see FÄRE, GROSSKOPF and LOVELL [1985, 1994] for details). First, overall technical efficiency ($\text{OTE} = \lambda^{\text{sd-crs}}$) is computed on a strongly disposable CRS model ($\mathrm{L}(y)^{\text{SD-CRS}}$). For instance, OTE in the inputs requires solving the following LP:

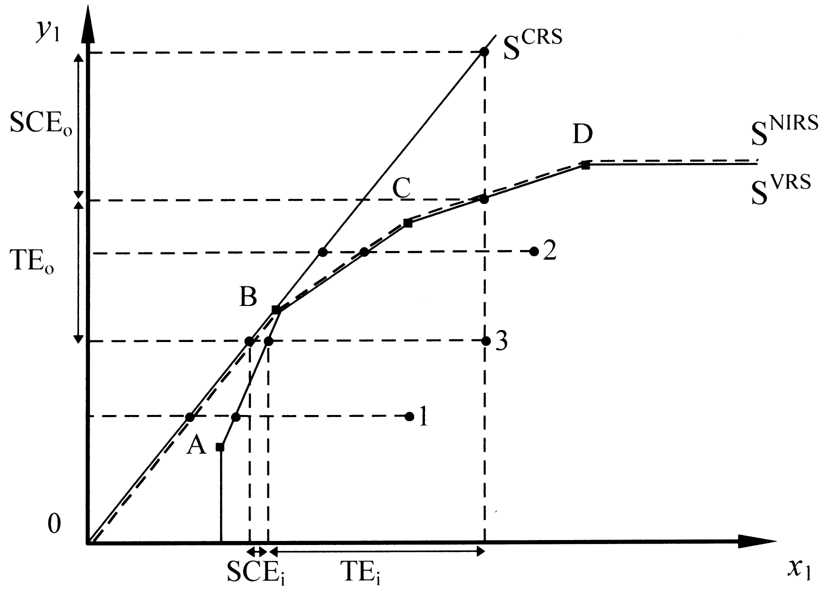
$$\begin{aligned} \mathrm{DF}_i(x^\circ, y^\circ) = {} & \min_{\lambda^{\text{sd-crs}}, z} \lambda^{\text{sd-crs}} \\ \text{s.t.} \quad & Y'z \geqslant y^\circ \\ & X'z \leqslant \lambda^{\text{sd-crs}} x^\circ \\ & \lambda^{\text{sd-crs}} \geqslant 0,\ z \geqslant 0, \end{aligned}$$

9. To be precise, when costs and outputs are available, then this limited information can be used to evaluate OE, but AE and TE can no longer be disentangled (see FÄRE and PRIMONT [1988]).

for each observation (x°, y°). Second, the measure of technical efficiency $(TE = \lambda^{wd\text{-}vrs})$ is evaluated on a weakly disposable technology with VRS $(L(y)^{WD\text{-}VRS})$. Third, structural efficiency $(STE = \lambda^{sd\text{-}vrs}/\lambda^{wd\text{-}vrs})$ is a ratio of two efficiency measures. The efficiency measure in the numerator $(\lambda^{sd\text{-}vrs})$ is calculated on a VRS technology that is congestion free by assumption $(L(y)^{SD\text{-}VRS})$; the denominator $(\lambda^{wd\text{-}vrs})$ is determined on a VRS technology that allows for input congestion $(L(y)^{WD\text{-}VRS})$. Since $\lambda^{sd\text{-}vrs} \leqslant \lambda^{wd\text{-}vrs}$, $0 < STE \leqslant 1$.

Finally, scale efficiency $(SCE = \lambda^{sd\text{-}crs}/\lambda^{sd\text{-}vrs})$ is also a ratio of two efficiency measures, but both evaluated on strongly disposable technologies. Its numerator $(\lambda^{sd\text{-}crs})$ is calculated on a CRS technology $(L(y)^{SD\text{-}CRS})$; the denominator $(\lambda^{sd\text{-}vrs})$ on a VRS technology $(L(y)^{SD\text{-}VRS})$. Because $\lambda^{sd\text{-}crs} \leqslant \lambda^{sd\text{-}vrs}$, evidently $0 < SCE \leqslant 1$. If $SCE = 1$, then the technology exhibits CRS at the observation under evaluation or at its projection point. When $SCE < 1$, it is possible to obtain for each observation qualitative information about returns to scale of its bounding hyperplane by comparing both components with a third efficiency measure evaluated on a technology imposing NIRS. $(\lambda^{sd\text{-}nirs})$.[10] Since these three technologies are nested, the input efficiency measures are ordered *a priori* $(\lambda^{sd\text{-}crs} \leqslant \lambda^{sd\text{-}nirs} \leqslant \lambda^{sd\text{-}vrs})$. On the one hand, if technical efficiency on CRS and NIRS technologies are equal $(\lambda^{sd\text{-}crs} = \lambda^{sd\text{-}nirs} < \lambda^{sd\text{-}vrs})$, then the observation is scale inefficient due to increasing returns to scale (IRS). On the other hand, if technical efficiency on the NIRS and VRS technologies are equal $(\lambda^{sd\text{-}crs} < \lambda^{sd\text{-}nirs} = \lambda^{sd\text{-}vrs})$, then the observation is scale inefficient due to decreasing returns to scale (DRS). This reasoning is illustrated on Figure 2 for observations 1 and 2.

FIGURE 2
Returns to Scale Characterisation of Individual Observations



10. Alternative methods for determining returns to scale are discussed in KERSTENS and VANDEN EECKAUT [1999].

Information on the congestion phenomenon can be refined. Following FÄRE *et al* [1985], it is possible to develop specific models testing whether each dimension separately contributes to congestion or not. To know the degree to which each input dimension contributes to congestion, it suffices to compute a radial efficiency measure relative to a VRS production model with strong input and output disposability, except for a single input dimension for which weak disposability is postulated. Repeating this computation for each input dimension and computing ratios as above yields partial input measures of structural efficiency.[11]

3.2. **The Impact of Measurement Orientation**

Obviously, input- and output-based decompositions may yield different information regarding the relative importance of the different sources of overall technical efficiency. For observation 3 in Figure 2, for example, the projection in the inputs would indicate technical efficiency as the major source $(\mathrm{TE}_i \leqslant \mathrm{SCE}_i)$, while when measuring into the output dimensions scale efficiency is of about equal importance $(\mathrm{TE}_o \cong \mathrm{SCE}_o)$.

Systematic informational similarities and differences of the decomposition of overall technical efficiency based on either input or output radial efficiency measures can be summarised in the following remarks (see FÄRE, GROSSKOPF and LOVELL [1994: 122-123]).

To start with the differences, input- and output-based measures of structural efficiency (STE) are not comparable. They measure the amount of congestion in the inputs and/or the outputs, depending on which dimensions are assumed strongly or weakly disposable, in either the input or output orientation.

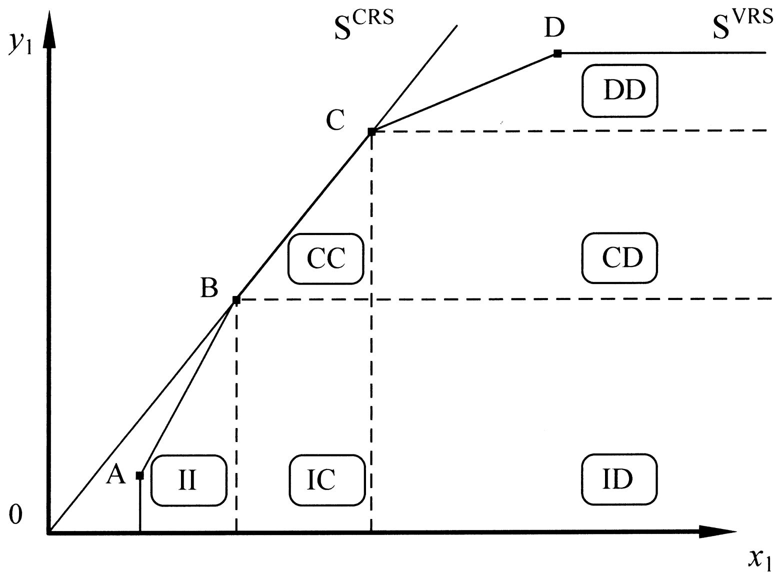
As to the similarities, the input and output measures of overall technical efficiency (OTE) and of scale efficiency (SCE) are related to each other. First, input and output measures of overall technical efficiency (OTE) are identical, since they are calculated on the same strongly disposable CRS technology $\mathrm{L}(y)^{\text{SD-CRS}}$.[12]

Second, since input and output technical efficiency measures calculated on a strongly disposable CRS technology $\mathrm{L}(y)^{\text{SD-CRS}}$ are identical, the ratios defining the scale efficiency measures (SCE) are only identical (or better, equal in reciprocal terms) if their denominators are identical (i.e., if input- and output-based efficiency measures relative to the short run technology $(\mathrm{L}(y)^{\text{SD-VRS}}$ are identical (equal in reciprocal terms)). Figure 3 displays a section in input-output space of technologies with CRS respectively VRS. Input and output based efficiency measures relative to the short run technology $(\mathrm{L}(y)^{\text{SD-VRS}})$ are equal (reciprocally) if, either observations are projected on a part of this technology exhibiting CRS (the region CC), or input and output efficiency measures both equal unity along the upward sloping segments of this VRS frontier (line segments AB and CD).

11. Further refinements are proposed in BYRNES *et al.* [1988] and FÄRE, GROSSKOPF and LOVELL [1994].
12. More precisely, $\mathrm{DF}_i(x,y) = 1/\mathrm{DF}_o(x,y)$ on CRS technologies (see FÄRE, GROSSKOPF and LOVELL [1985], or FØRSUND and HJALMARSSON [1979]).

FIGURE 3

Input-based and Output-based Measures of Scale Efficiency (SCE)



Also the determination for each observation of the exact nature of returns to scale at its bounding hyperplane may yield conflicting information. Referring to Figure 3, both decompositions predict CRS, IRS and DRS for all observations in respectively regions CC, II, and DD.[13] But for activities located in other regions input and output projections yield contradictory information. These decompositions predict respectively IRS and CRS for region IC, IRS, and DRS for region ID, and CRS and DRS for region CD. These qualitative differences result from the fact that returns to scale of inefficient units are conditional on a move to the frontier (FÄRE, GROSSKOPF and LOVELL [1994], BANKER, CHANG and COOPER [1996]).

3.3. Conclusions

Concluding, implementing these decompositions of overall technical efficiency results in detailed information regarding its possible sources. Furthermore, contrasting the results of input-based and output-based decompositions sheds light on the similarities and differences in the information generated.

---

13. Adapted from Figure 4.10 in FÄRE, GROSSKOPF and LOVELL [1994: 124].

4 Description of the Sample: Urban Transit in France

The empirical analysis is based on a date set of French urban transit firms that operate outside the Paris region. The institutional environment is aptly sketched as follows (MITRIC [1988]).[14] During a certain period, an urban transport operator supplies transport services within a transport perimeter agreed upon with a public organising authority (a municipality or group of municipalities). This authority most often owns the infrastructure, the equipment and the rolling stock. The transport perimeter is not limited by territorial boundaries, and only serves to distinguish urban from interurban transport. The operator can be a private, public or mixed company. In France the private sector plays an important role in urban transit. It is one of the few countries where transit ridership has grown in the last two decades through a strategy of expanding both the range and quality of the transit services supplied. This success required only a moderate increase in subsidy levels relative to costs (see PUCHER [1988]).

Recently, urban transit performance has been evaluated using frontier methodologies. The studies of CHU, FIELDING and LAMAR [1992], GATHON [1989], KERSTENS [1996], OBENG [1994], TONE and SAWADA [1990], TULKENS [1993], and VITON [1997] among others, are based on non-parametric deterministic production models. KERSTENS [1996] analysed the same sample, but focused on causally explaining technical efficiency only. This study decomposes overall technical efficiency into its sources. To the best of our knowledge, no urban transit study has so far reported such extensive decomposition results.[15]

The sample contains 114 single mode urban transport companies in 1990 driving buses (all other modes are excluded).[16] Two types of outputs and three traditional inputs are selected to specify two separate models of production technology. The output is the number of vehicle kilometres in one model (in 1000), and the number of passengers in another model (in 1000).[17] The first output, which is a pure supply indicator, is a classical units times distance per unit time concept. The second, demand-related output measures the utilisation of the services being offered. This specification reflects the economic motive for providing the services. Ideally, one would like to evaluate the allocative efficiency of transit operators, but this is very difficult for a regulated

14. The institutional organisation of French urban transport is summarised in greater detail in KERSTENS [1996].
15. VITON [1997] is the first urban transit study aiming to decompose OTE into all its components. It uses nonradial instead of radial efficiency measures to obey the Koopmans instead of the Farrell definition of efficiency. But this requires specific modifications to capture congestion properly (see DERVAUX, KERSTENS and VANDEN EECKAUT [1998]). A methodological error prevents VITON [1997] to detect any congestion at all. Thus, no urban transit study properly accounted for the possibility of congestion.
16. Note that some observations were discarded: see KERSTENS [1996] for details.
17. KERSTENS [1996] only used pure supply indicators. BERECHMAN [1993: 97, 152-153] argues against the use of demand-related outputs.

industry. Furthermore, the lack of adequate price information prevents pursuing this possibility. The second output specification can therefore be interpreted as a short-cut for a more extensive evaluation. It is sometimes referred to as effectiveness (see CHU, FIELDING and LAMAR [1992]), COSTA [1998] and TONE and SAWADA [1990] in the transport literature). Following BRADFORD, MALT and OATES [1969], another way to classify both models is to distinguish between direct outputs (D-output) and consumer related outputs (C-output).

Each of these two outputs is combined with the following inputs to yield two specifications of technology: average number of vehicles used over the year; average number of employees over the year; and yearly total fuel consumption (in m^3). This definition of the inputs closely follows tradition in the transportation literature.

In the traditional parametric literature, the above output specifications are often complemented with variables accounting for spatial, temporal and quality characteristics of urban transit services (JARA DÍAZ [1982]). Some non-parametric deterministic efficiency studies (e.g., VITON [1997]) include additional dimensions representing these network characteristics. However, these reference technologies have serious problems to account for these characteristics (see KERSTENS and VANDEN EECKAUT [1995] for details). In particular, there is no general way to determine sign and significance of any additional dimensions. Therefore, we refrain from adding any network characteristics to the inputs and outputs in the technology. Consequently, we cannot distinguish between economies of scale, of traffic density, and other scale notions often estimated in parametric urban transit studies (see BERECHMAN [1993]).

5 Decomposing Overall Technical Efficiency: Results

5.1. The Impact of Outliers and the Structure of Results

The technologies used for the decomposition are specified in a deterministic way, disregarding any measurement error in the data. As argued by, among others, WILSON [1993] it is important to control for the effect of outliers. Also recall that the scale efficiency measure and the determination of scale economies for the individual observations are based on, among others, a CRS technology. It is intuitively clear that this envelopment of observations by a conical hull (i.e., $L(y)^{SD\text{-}CRS}$) is relatively more sensitive to outliers than other model specifications. If this intuition turns out to be correct, then the efficiency decomposition results could well be unreliable. Consequently, it is useful to check for outliers.

The analysis of outliers in this contribution relies on robust Mahalanobis distances. The classical Mahalanobis distance measures how far a random vector is situated from the centre of its distribution taking account of the shape of the multivariate cloud. To avoid a bias introduced by groups of outliers (masking effect) the first two moments – needed to compute the Mahalanobis distance – are estimated for each model specification separately using a robust minimum volume ellipsoid (MVE).[18] It turns out that for both model specifications 3 out of the 4 (respectively 5) efficient observations characterised by CRS (see fourth line in Table 2) are diagnosed as outliers. In addition, 40 (respectively 33) inefficient observations are potential outliers. Judgements on the main sources of inefficiency, based on sample averages, could well be affected by their presence.

Once outliers are detected several options remain open: data can be re-examined and eventually corrected, models can be reformulated, observations can be discarded, etc. In our situation, only the latter, more radical response was feasible. To limit the impact on both the determination of the frontier and average sample sources of inefficiency, all potentially outlying observations are excluded from the sample. Both decompositions are then recomputed. This leaves us with an adjusted sample size of 71 and 78 observations for respectively the vehicle kilometres and the passengers output specifications. Since both samples without outliers have 70 observations in common, the determination of outliers does not seem to depend on the selected production model.

Table 1 offers descriptive statistics for both model specifications and for both the complete sample and the sample without outliers. Since the latter sample differs in size for the two model specifications, the second part of the

TABLE 1
Descriptive Statistics on Inputs and Outputs

	Mean	Standard Deviation	Minimum Value	Maximum Value
Full sample (*N* = 114)				
Vehicle Kms (y_1)	2416.1	3290.7	110	213.5E+02
Passengers (y_2)	8195.8	12235.3	124	67726
Vehicles (x_1)	63.44	85.46	4	521
Employees (x_2)	149.82	248.44	5	1662
Fuel (x_3)	972.37	1433.3	39	8986
Observations without Outliers (*N* = 71 & *N* = 78)				
Vehicle Kms (y_1)	728.70	527.82	110	2570
Passengers (y_2)	8392.18	13158.52	124	67726
Vehicles (x_1)*	21.11	14.40	4	70
	62	90.69	4	521
Employees (x_2)*	41.02	31.84	5	135
	154.44	272.86	5	1662
Fuel (x_3)*	283.77	224.47	39	1020
	1002.76	1548.21	39	8986

* Numbers on the first and the second line refer to the y_1 respectively y_2 output specifications.
Source: Ministère de l'Equipement, du Logement, des Transports et de l'Espace (1991) *Annuaire Statistique sur les Transports Collectifs Urbains: Statistiques 1983-1990*, Paris, Direction des Transports Terrestres.

18. Details on this methodology are outlined in KERSTENS [1996] and especially in SEAVER and TRIANTIS [1995].

table reports two lines for each input. Clearly, French transit companies differ widely in size. Give these size differences, it is useful to determine the precise impact of both technical and scale efficiencies. This is exactly the aim of the overall technical efficiency decomposition. Comparing averages between full and adjusted sample, however, reveals that mean vehicle kilometres fall drastically. This indicates that especially the larger companies with a relative low passenger loading have been eliminated as outliers. While the 70 observations in common are supplemented with some large companies in the passengers model, the latter operators seem to be missing in the other model. Though *a priori* more weight should be given to the analysis without outliers, the different composition of both samples should make us extremely cautious when discussing matters regarding size.

Empirical results are based on two model specifications and on output and input efficiency measures. Furthermore, findings are reported on the initial sample of 114 operators and on the samples without outliers. Moreover, all output oriented efficiency measures are defined to be no larger than unity to facilitate the comparison of both decompositions.[19]

First, decomposition results are analysed at the sample level. Attention focuses on the amount of efficient and inefficient observations, the resulting distributions and correlations among the different model variants. Then, more detailed results are discussed over size classes. Finally, our results are related to the existing body of evidence on urban transit performance.

5.2. Decomposition Results at the Sample Level

First, as this decomposition is based on the computation of four efficiency measures for each observation (i.e., $\lambda^{\text{sd-vrs}}$, $\lambda^{\text{sd-crs}}$, $\lambda^{\text{sd-nirs}}$ and $\lambda^{\text{wd-vrs}}$), it is instructive to analyse the impact of different disposability and returns to scale assumptions on the elementary classification of efficient versus inefficient observations. Table 2 reports the number of efficient observations for the four postulated production models. Three comments can be made. First, comparing the models with weak and strong disposability, there is evidence of congestion since the number of efficient observations more or less doubles. This not only implies that almost half of the efficient observations operate in a congested way, but also that there is ample score for projecting inefficient observations on these backward bending parts of technology. Second, the number of efficient observations spanning the technology decreases as the returns to scale assumptions become more restrictive. Third, eliminating outliers has a large, positive impact on the relative number of efficient observations for all reference technologies. A notable exception is the CRS vehicle kilometers model where technology is spanned by barely 4% of the sample.

Next, these decompositions of overall technical efficiency are applied for both sample sizes of urban transit companies. Table 3 reports the following detailed results in both orientations at the sample level: overall technical efficiency (OTE), scale efficiency (SCE), technical efficiency (TE), and

19. For this purpose, the radial output efficiency measure is redefined as: $DF_o^*(x,y) = \min\{\theta^* \mid \theta^* \geqslant 0,\ x \in L(y/\theta^*)\}$. Hence, the relations between both decompositions mentioned earlier hold with identity, not reciprocally.

TABLE 2
Efficient Observations Spanning the Different Production Models

	Vehicle kilometres		Passengers	
Full sample (N = 114)				
wd ∧ vrs	20	17.54%	25	21.93%
sd ∧ vrs	11	9.65%	14	12.28%
sd ∧ nirs	7	6.14%	9	7.89%
sd ∧ crs	4	3.51%	5	4.39%
Adjusted sample	(N = 71)		(N = 78)	
wd ∧ vrs	29	40.85%	25	32.05%
sd ∧ vrs	18	25.35%	13	16.67%
sd ∧ nirs	14	19.72%	7	8.97%
sd ∧ crs	3	4.23%	6	7.69%

wd = weak input and output disposability; sd = strong input and output disposability; vrs = variable returns to scale; nirs = non-increasing returns to scale; crs = constant returns to scale.

TABLE 3
Decompositions of Overall Technical Efficiency: Sample Level

	OTE_o OTE_i	SCE_o SCE_i	TE_o TE_i	STE_o STE_i	IRS_o IRS_i	CRS_o CRS_i	DRS_o DRS_i
Vehicle kilometres							
Full sample	0.682(4)	0.916(4)	0.780(20)	0.970(31)	58	4	52
(N = 114)		0.921(4)	0.775(21)	0.970(27)	65	4	45
				x_1 0.999			
				x_2 0.998			
				x_3 0.984			
Adjusted sample	0.748(3)	0.847(3)	0.911(29)	0.974(45)	12	3	56
(N = 71)		0.858(4)	0.903(29)	0.971(45)	31	3	37
				x_1 0.999			
				x_2 0.984			
				x_3 0.979			
Passengers							
Full sample	0.669(5)	0.916(5)	0.778(25)	0.954(40)	53	5	56
(N = 114)		0.887(5)	0.799(26)	0.955(46)	63	5	46
				x_1 0.999			
				x_2 0.977			
				x_3 0.995			
Adjusted sample	0.714(6)	0.923(6)	0.821(25)	0.957(42)	51	6	21
(N = 78)		0.898(6)	0.835(24)	0.964(47)	57	6	15
				x_1 0.996			
				x_2 0.991			
				x_3 0.980			

IRS = increasing returns to scale; CRS = constant returns to scale; DRS = decreasing returns to scale.

structural efficiency (STE). Furthermore, the sources of scale efficiency are determined by reporting for each observation the returns to scale of its bounding hyperplane: increasing (IRS), constant (CRS) and decreasing (DRS) returns to scale. Each first line in the table contains the following information:

average output efficiency measures, number of efficient observations (between brackets) and returns to scale determination. Each second line does the same for the input orientation, and in addition provides congestion measures per input component. The second line in the first column reports the number of observations in the sample or size class (between brackets). A single number is reported for overall technical efficiency (OTE), as both orientations are by definition identical.

For the initial samples the major source of deviations from overall technical efficiency (OTE) is technical efficiency (TE), followed by deviations from scale efficiency (SCE), and finally structural efficiency (STE). Differences between output and input orientations are minor, except for the amount of scale efficiency in the passengers model. For the sample without outliers the level of overall technical efficiency (OTE) is higher in both specifications. The ordering of the efficiency sources remains the same for the passengers model. By contrast, scale efficiency (SCE) becomes the primary source of poor performance for the vehicle kilometres specification. Output-based and input-based decompositions diverge little, again with exception of scale efficiency in the passengers specification.

In addition to these sample averages, it is important to check how many observations are actually affected by these inefficiencies. In the original sample, under both models and both orientations, about 20 to 26 observations are technically efficient (TE), between 27 and 46 observations are efficient from the viewpoint of congestion (STE), but only 4 to 5 observations are scale efficient (SCE). This last outcome is due to the fact that CRS conical hulls are often spanned by few observations. For the adjusted samples, irrespective of models and orientations the relative number of efficient observations for technical efficiency (TE) and congestion (STE) components increases substantially, while scale components (SCE) change barely.

Having discussed averages and one extreme side of the distribution, i.e., the efficient observations, it is useful to represent the complete distribution of the four components. For the adjusted samples, the distributions of components for both decompositions are shown on Figure 4 to 7. Reasons of space preclude showing the complete sample distributions. All distributions are highly skewed to the right. For the samples without outliers, for instance, about 65 observations have structural efficiency scores situated in the (0.90-l) interval whereas about 4 units have scores in the (0.70-0.80) interval. The apparent strong similarities between these distributions require qualification. Non-parametric Wilcoxon signed-ranks test statistics reveal that input and output orientations yield different distributions, except for congestion components and, sometimes, scale efficiencies.[20] The same test statistics also indicate that both output specifications in general do follow a common distribution on the initial samples. The same is true for the reduced samples (70 common observations), except for the technical and scale efficiency components.

20. For reasons of space, details of all non-parametric test statistics are suppressed. Results are always reported at the 5% significance level. As noted by a referee, selecting proper statistical test procedures is not evident, since the traditional independence assumption between observations within each set of measurements is invalid. Lacking a consensus on this issue, we adopt the tests discussed in GROSSKOPF's [1996] review of statistical inference in the area.

FIGURE 4
Output-based OTE Decomposition: Vehicle Kms. (N = 71)

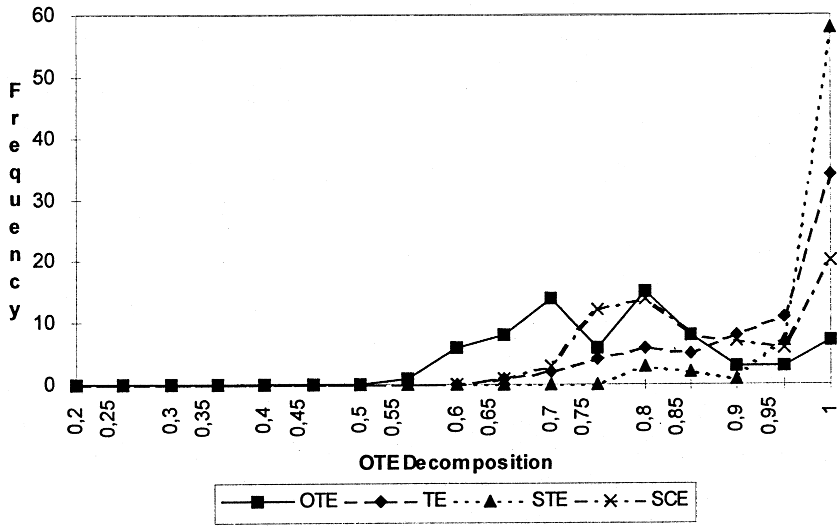
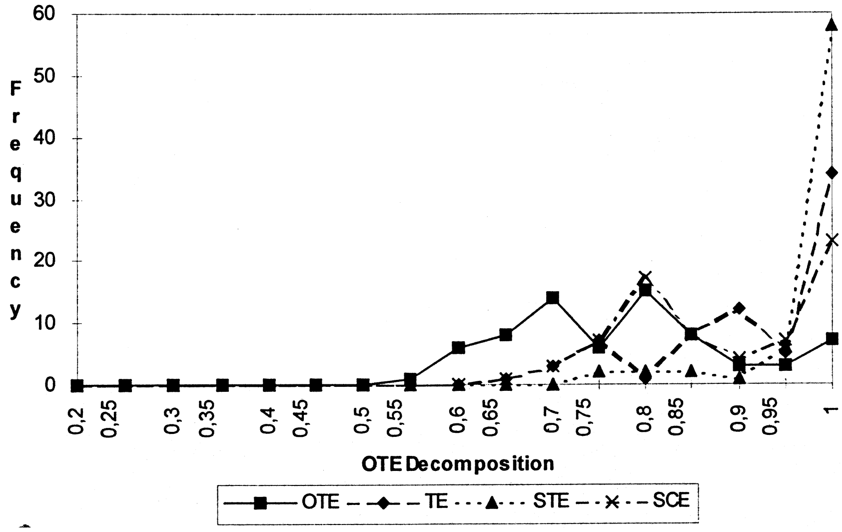


FIGURE 5
Input-based OTE Decomposition: Vehicle Kms. (N = 71)



Although the relative importance of congestion is small, the results nevertheless imply that an important number of transit companies provides transportation services in a somehow congesting way. Additional information on the contribution of each input dimension separately is available. Although the amounts are minor, the second labour input and the third energy input clearly contribute most to congestion. Repeating this detailed analysis for the adjusted sample confirms the full sample results. A Friedman test for related samples reveals that the global and partial congestion measures do not share a

FIGURE 6

Output-based OTE Decomposition: Passengers (N = 78)

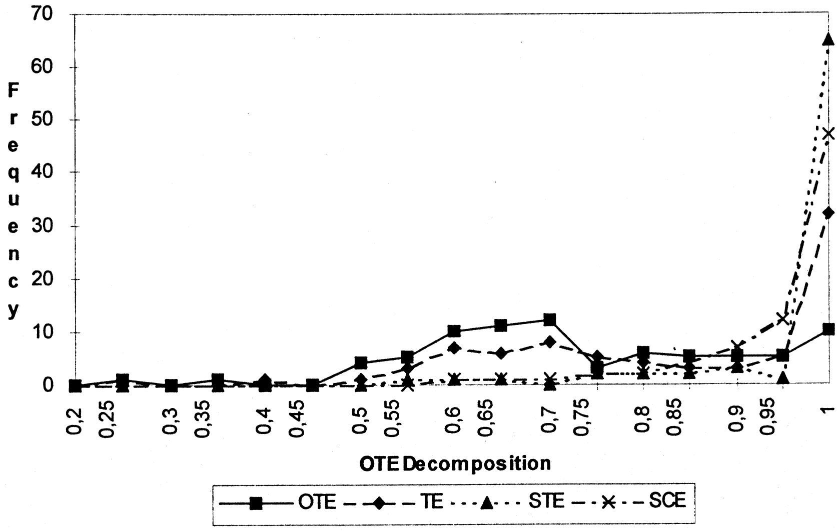
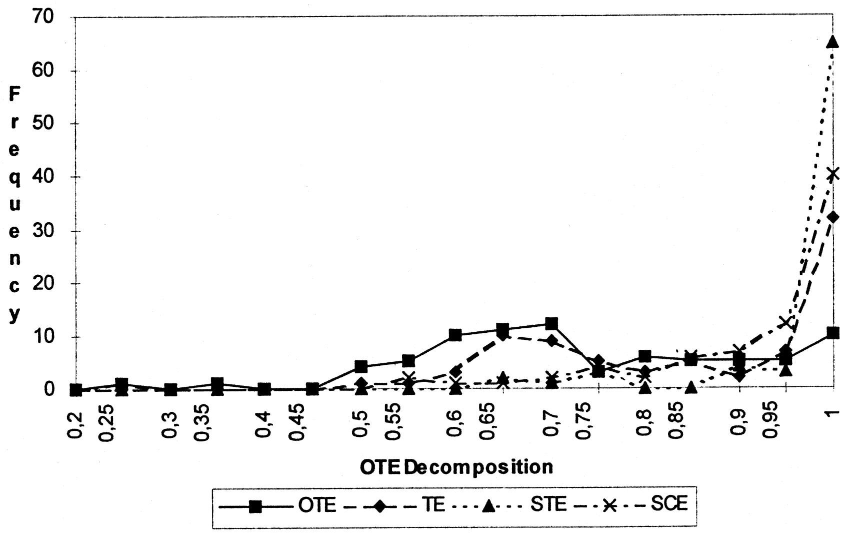


FIGURE 7

Input-based OTE Decomposition: Passengers (N = 78)



common distribution. Pairwise Wilcoxon signed-ranks test confirm this result, except for the second and third input pair in the original samples.

The determination of returns to scale at the bounding hyperplanes of each observation provides qualitative information about the origin of any scale inefficiencies. For the original samples, slightly more than half of the observations are characterised by IRS. These differences can be considered in greater detail: 7 and 10 observations have a different classification under output- and input-based decompositions for respectively the vehicle kilometres and the passengers specification. In terms of Figure 3, they are situated in region ID. The samples without outliers show less coherence. For the passengers model, the majority of observations are again classified as experiencing IRS, but the revere holds true for the vehicle kilometres output. Hence, it is difficult to formulate firm conclusions regarding returns to scale. Furthermore, the choice of measurement orientation has a relative large effect on this classification in the vehicle kilometres case: 19 observations have a different classification. For the passengers output, both orientations yield conflicting suggestions for only 6 observations. All are situated in the ID region. Non-parametric tests confirm that both measurement orientations always yield different classifications. These differences are in line with the findings of FUKUYAMA [1996].

To explore the similarities between both output specifications further, SPEARMAN rank correlations are reported in Table 4 for both output- and input-based decompositions. While for the original sample size the similarities in ranking are fairly acceptable (with the notorious exception of OTE), the 70 common observations in the samples without outliers diverge widely in their rankings. There is almost no correlation and for scale efficiency the correlations even turn out to be mildly negative. Correlations between the vehicle kilometres model in input orientation and the passengers model in output orientation – of potential interest from a policy viewpoint – are not much better. This clearly indicates that all conclusions are conditional on the type of output specified. In other words, assessing efficiency and effectiveness need not coincide, at least in our sample. These conflicting results for the reduced sample, in line with the results at the sample level and the returns to scale information, may be due to the relatively small number of remaining

TABLE 4
Spearman Rank Correlation Analysis

	OTE	SCE	TE	STE
Correlation between Vehicle Kms. and Passengers OTE Decompositions				
Output orientation				
Full sample (N = 114)	– 0.010	0.674	0.499	0.298
Adjusted sample (N = 70)	0.049	– 0.106	0.172	0.341
Input orientation				
Full sample (N = 114)	– 0.010	0.557	0.485	0.310
Adjusted sample (N = 70)	0.049	– 0.137	0.110	0.342
Vehicle Kms: Input orientation/Passengers: Output orientation				
Full sample (N = 114)	– 0.010	0.681	0.478	0.301
Adjusted sample (N = 70)	0.049	– 0.160	0.146	0.292
Correlation between Output and Input OTE Decompositions				
Vehicle kilometres				
Full sample (N = 114)	1.000	0.829	0.966	0.912
Adjusted sample (N = 71)	1.000	0.937	0.994	0.614
Passengers				
Full sample (N = 114)	1.000	0.803	0.877	0.781
Adjusted sample (N = 78)	1.000	0.937	0.915	0.827

observations. As noted earlier, the few observations spanning the CRS technology in the vehicle kilometres case may also partially explain these results.

Taking a closer look to the similarity between output and input decompositions, we compute the number of observations for which output- and input-based components are identical and assess the similarities in their rankings.

First, we illustrate the number of identical observations in terms of output- and input-based components for the adjusted sample only. The number of observations with an identical efficiency measure for technical (TE), structural (STE) and scale efficiency (SCE) components equals for both models respectively 29, 42 and 18, and 13, 41 and 23. For the technical (TE) and structural efficiency (STE) components all these observations are efficient. For the scale efficiency measure (SCE) 15 and 7 observations in both specifications have an identical, though not unity ratio.[21]

Second, the similarities in rankings are evaluated by Spearman rank correlations between the respective components of both orientations. These correlation coefficients, also reported in Table 4, show two major findings. First, correlations in the passengers model are much lower than in the vehicle kilometres case. Second, correlation results are lowest for scale efficiency (SCE) and congestion (STE) components, depending on the model specification. Unity correlations for overall technical efficiency (OTE) merely reflect the fact that both orientations yield, by definition, identical efficiency measures.

5.3. Decomposition Results over Size Classes

Next, findings at the sample level are complemented with a detailed analysis over output size classes. Table 5 is structured very similar to Table 3. To facilitate comparisons, the size classes of the passengers model follow the vehicle kilometres classification. The sample without outliers in the vehicle kilometres case contains no more big operators, resulting in the three largest, but empty size classes being eliminated from the table.

Table 5 yields the following major conclusions. First, the ordering of sources of overall technical efficiency varies considerably over the size classes. There are, for example, size classes for which scale efficiency (SCE) is the major source of poor performance instead of technical efficiency (TE) at the sample level. Second, measurement orientation clearly matters for specific size classes. For the smallest class in the vehicle kilometres specification without outliers, for instance, output measurement indicates that scale efficiency is the major source while input-based decomposition points to technical efficiency being most important. The orientation of measurement also affects the returns to scale determination for specific observations. Finally, the pattern of returns to scale over the size classes starts with the majority of observations being subject to IRS and ends with a major importance of DRS.

---

21. Their input and output efficiency measures relative to the short run technology ($L^{sd-vrs}(y)$) are identical since they are projected on a CRS part of this technology. In terms of Figure 3, they are situated in region CC or along the line segments AB or CD.

TABLE 5

Decompositions of Overall Technical Efficiency: Size Classes

	OTE_o OTE_i	SCE_o SCE_i	TE_o TE_i	STE_o STE_i	IRS_o IRS_i	CRS_o CRS_i	DRS_o DRS_i
Vehicle kilometres							
Full sample							
<100	0.667(1)	0.949(1)	0.742(9)	0.962(10)	41	1	9
(51)		0.931(1)	0.756(10)	0.964(11)	41	1	9
100-200	0.681(0)	0.993(0)	0.713(1)	0.964(0)	16	0	7
(23)		0.994(0)	0.707(1)	0.971(0)	19	0	4
200-300	0.729(2)	0.927(2)	0.790(2)	0.991(5)	1	2	11
(14)		0.979(2)	0.757(2)	0.981(2)	4	2	8
300-400	0.732(0)	0.893(0)	0.870(0)	0.949(0)	0	0	4
(4)		0.931(0)	0.851(0)	0.935(0)	1	0	3
400-500	0.791(0)	0.871(0)	0.923(1)	0.984(0)	0	0	3
(3)		0.893(0)	0.911(1)	0.974(0)	0	0	3
500>	0.661(1)	0.735(1)	0.911(7)	0.989(16)	0	1	18
(19)		0.764(1)	0.885(7)	0.985(14)	0	1	18
Adjusted sample							
<100	0.767(3)	0.885(3)	0.899(21)	0.969(34)	12	3	35
(50)		0.897(4)	0.891(21)	0.965(34)	25	3	22
100-200	0.698(0)	0.755(0)	0.937(7)	0.985(10)	0	0	20
(20)		0.765(0)	0.929(7)	0.983(10)	6	0	14
200-300	0.745(0)	0.745(0)	1.000(1)	1.000(1)	0	0	1
(1)		0.745(0)	1.000(1)	1.000(1)	0	0	1
Passengers							
Full sample							
<100	0.603(2)	0.883(2)	0.745(12)	0.941(23)	48	2	1
(51)		0.804(2)	0.812(14)	0.937(24)	48	2	1
100-200	0.707(1)	0.992(1)	0.732(2)	0.979(5)	4	1	18
(23)		0.992(1)	0.733(2)	0.976(11)	13	1	9
200-300	0.705(1)	0.957(1)	0.746(1)	0.987(4)	1	1	12
(14)		0.985(1)	0.730(1)	0.983(3)	2	1	11
300-400	0.738(0)	0.937(0)	0.808(0)	0.976(1)	0	0	4
(4)		0.958(0)	0.791(0)	0.980(0)	0	0	4
400-500	0.674(0)	0.928(0)	0.821(1)	0.904(1)	0	0	3
(3)		0.943(0)	0.815(1)	0.899(0)	0	0	3
500>	0.759(1)	0.874(1)	0.934(9)	0.936(6)	0	1	18
(19)		0.886(1)	0.897(8)	0.962(8)	0	1	18
Adjusted sample							
<100	0.646(1)	0.870(1)	0.804(12)	0.950(23)	37	1	0
(38)		0.817(1)	0.843(13)	0.956(23)	37	1	0
100-200	0.752(0)	0.993(0)	0.775(1)	0.982(9)	12	0	2
(14)		0.986(1)	0.779(1)	0.982(9)	12	0	2
200-300	0.767(2)	0.995(2)	0.789(2)	0.977(4)	1	2	5
(8)		0.994(2)	0.786(2)	0.982(5)	5	2	1
300-400	0.817(0)	0.996(0)	0.834(0)	0.980(1)	0	0	2
(2)		0.998(0)	0.824(0)	0.991(1)	0	0	2
400-500	0.731(0)	0.995(0)	0.827(1)	0.908(1)	1	0	2
(3)		0.995(0)	0.827(1)	0.908(2)	2	0	1
500>	0.820(3)	0.932(3)	0.935(9)	0.948(4)	0	3	10
(13)		0.945(3)	0.906(7)	0.964(7)	1	3	9

Size classes for vehicle kilometres: in 10,000 kilometres; for passengers: same as for vehicle kilometres. IRS = increasing returns to scale; CRS = constant returns to scale; DRS = decreasing returns to scale.

This analysis of size classes indicates that there can be a wide variation in results at the individual level. Looking solely at sample averages may yield misleading conclusions for specific operators. It is precisely one of the strengths of these decomposition techniques to provide detailed information on the performance of individual organisations. This allows formulating operator specific policy conclusions, though the potential effect of measurement orientation should induce carefulness in this matter.

5.4. **Relation to Existing Literature**

There are not that many studies reporting a similar decomposition of overall technical inefficiency. Illustrations from different economic sectors include: the BYRNES and VALDMANIS [1994] analysis of hospitals; the BYRNES *et al.* [1988] study on surface mining of coal; the FÄRE, GROSSKOPF and PASURKA [1989] investigation on electric utilities; the work of ÇAKMAK and ZAIM [1992] and FÄRE, GRABOWSKI and GROSSKOPF [1985] on agriculture; the analysis of Polish industrial sectors by KEMME and NEUFELD [1991]; and the study of FIELD [1990] on building societies. Most of them employ an output-oriented decomposition. Their results can be summarised as follows. The main source of overall technical inefficiencies (OTE) is purely technical inefficiency (TE) for BYRNES and VALDMANIS [1994]; scale inefficiency (SCE) for FÄRE, GRABOWSKI and GROSSKOPF [1985], FIELD [1990] and KEMME and NEUFELD [1991]; and congestion (STE) in the BYRNES *et al.* [1988] the ÇAKMAK and ZAIM [1992], and the FÄRE, GROSSKOPF and PASURKA [1989] cases. As to the cause of scale inefficiency, the BYRNES *et al.* [1988], the FÄRE, GRABOWSKI and GROSSKOPF [1985], the FIELD [1990] and the KEMME and NEUFELD [1991] studies on the one hand, and the FÄRE, GROSSKOPF and PASURKA [1989] work on the other hand, reveal respectively DRS and IRS for the majority of the observations. It is evident that organisations follow a wide variety of patterns as to the sources of their overall technical inefficiency. Given this diversity of reported results, it is clear that French urban transit companies are in no way emulating a peculiar pattern of inefficiencies.

Although our findings regarding efficiency and effectiveness may seem extreme, they are not unrelated to results reported in urban transport frontier studies. CHU, FIELDING and LAMAR [1992] also find, for smaller U.S. samples, the possibility of a divergence. SCHINNAR [1993: 178] cites a study of 145 public bus companies reporting a negative association between efficiency and effectiveness. TONE and SAWADA [1990: 363] describe exactly the same trade-off. COSTA [1998] considers a single operator over a small time span and finds a temporal pattern of simultaneous improvement in efficiency and effectiveness after the introduction of organisational reforms. TULKENS and WUNSCH [1994], by contrast, find a temporal pattern of improving efficiency and deteriorating effectiveness. Given that all frontier studies we are aware of (except one) yield a negative relation between efficiency and effectiveness, this apparent paradox requires further investigation. One implication for regulatory policies is that the choice between input and output monitoring may require reconsideration.

The results on returns to scale for individual operators can also be linked to traditional studies of urban transit production. BERECHMAN [1993: 137], summarising mainly average practice studies, states that “results concerning economies of scale are rather inconclusive. The empirical evidence seems to suggest that the bus industry as a whole operates under conditions of constant scale economies”. He adds that probably small firms are facing IRS, medium-sized firms enjoy small or constant scale economies, and large bus systems operate under DRS. There are also a few parametric frontier studies of urban transit reporting estimates for economies of scale (FAZIOLI, FILIPPINI and PRIONI [1993], FILIPPINI, MAGGI and PRIONI [1992], THIRY and TULKENS [1992], VITON [1986], among others). Typically, they find IRS – although VITON [1986] is an exception – indicating that urban transit is a natural monopoly.

Results reported in this study have the advantage of directly focusing on the individual operators. The size classes results are certainly in line with the BERECHMAN [1993] conjecture. However, several remaining problems prevent us from making solid conclusions regarding French operators. First, overall conclusions regarding returns to scale for the analysis without outliers rely on the output specification. Another major difficulty is that for specific cases (and size classes) the results may depend on the selected orientation of measurement. Finally, these findings are subject to the further proviso that the exclusion of network characteristics does not bias our computations of scale economies. The non-parametric VITON [1997] study, by contrast, consistently finds about 70% of observations situated in the CRS region, independent of measurement orientation. This study also confirms the rather monotonous relation between scale and size in our sample.

The economic consequences for public policy are potentially important. If IRS prevail for smaller operators, then only these urban transit companies must be considered natural monopolies, at least assuming that they supply a single output.[22] Consequently, these small companies are bound to incur losses under marginal cost pricing and appropriate regulatory regimes, e.g., subsidy schemes, must be developed. For the larger companies there may not be a natural monopoly problem at stake. However, given the above limitations, additional studies are required focusing on the determination of operator-specific returns to scale information.

6 Conclusions and Directions for Future Research

This contribution has focused on the sources of observed technically inefficient behaviour among French urban transit companies. Efficiency measurement requires frontier specifications of technology. Production technologies in this article are so-called non-parametric, deterministic reference technologies.

22. The conditions for the existence and sustainability of a natural monopoly in the multiple output case are spelled out in detail in NADIRI [1982: 483-487].

In the theoretical sections, overall efficiency has been decomposed into its sources following FÄRE, GROSSKOPF and LOVELL [1983, 1985]. This decomposition results in component-wise efficiency measures reflecting allocative, scale, structural and technical inefficiencies. Also computational details have been provided.

The empirical analysis yields several policy conclusions for the data set of French urban transit companies. First, technical inefficiency and inefficiencies due to inadequacies in scale are the major sources of poor performance, depending on sample size and output specification (see also VALDMANIS [1992]). Congestion only plays a minor role, and is mainly situated in the labour and energy inputs. The technical inefficiencies, due to managerial failures or an adverse environment, can be remedied in the short run. The determinants of technical inefficiencies found in KERSTENS [1996] may be of help in formulating improvement strategies. Second, about half of the observations – mainly small operators – are situated on an IRS part of technology. By contrast, large operators experience DRS. However, the mixed results for the analysis without outliers should make us prudent to formulate too bold policy conclusions. Therefore, whether network (scale) adjustments in the long run yield sufficient benefits given the likely minor cost reductions or revenue gains remains an open issue. Third, the performance of operators depends strongly on the output specification selected. In particular, enhancing efficiency and effectiveness may well be two different things. This finding underscores the importance of obtaining appropriate specifications of technology. It also may lead to reconsider the objectives of urban transit operators and their monitoring in regulatory policies. Fourth, sample level results may hide wide variations in results for individual companies. One of the main advantages of the methodology is that it allows formulating operator specific policy conclusions.

There are also important methodological conclusions. First, our study is the first to show that outliers may well affect decomposition results, since reference technologies are deterministic. The samples accounting for possible outlier bias yield probably the most credible results. Second, measurement orientation can have a major impact on the sources of efficiency and on the local returns to scale for individual observations, irrespective of the sample size retained. These weaknesses of the overall technical efficiency decomposition using non-parametric, deterministic technologies should make us cautious about this methodology in general and in interpreting our own results in particular.

The major drawbacks of this decomposition methodology invite further research. Sticking to non-parametric technologies, the following avenues may turn out promising. First, it may prove useful to allow for measurement error using resampling methods (see GROSSKOPF [1996] for a survey). Second, it would be desirable to have statistical test procedures when specifically inferring local scale information (see BANKER [1996] and SIMAR and WILSON [1998] for recent proposals). Finally, instead of remedying the deterministic nature of these technologies one could weaken the underlying assumptions. KERSTENS and VANDEN EECKAUT [1998], for instance, develop a more limited decomposition (without congestion component) based upon non-convex technologies and find, among other things, that local qualitative scale information depends little on measurement orientation and that technical and scale inefficiencies differ markedly between convex and non-convex technologies.

But this research may also have implications for parametric frontier specifications. First, it remains an open question to which extent also parametric decompositions of technical and scale efficiencies (see, e.g., BAUER [1990b], KUMBHAKAR, BISWAS and BAILEY [1989]) are subject to the same impact of changes in measurement orientation. Data availability and the ensuing choice of a primal or dual approach often dictate measurement orientation. Casual evidence suggests there may be a problem. ATKINSON and CORNWELL [1993], for instance, develop output and input technical efficiency measures in terms of dual cost frontiers. Their translog cost frontier results indicate substantial cardinal and ordinal differences between both types of technical efficiency. Second, a to our knowledge unexplored, alternative is to develop parametric stochastic frontier models allowing for a decomposition of overall technical efficiency into its three components. To be specific, the structural efficiency (congestion) component has so far never been estimated parametrically.

• References

ATKINSON, S.E., CORNWELL, C. (1993). – Measuring Technical Efficiency with Panel Data: A Dual Approach, *Journal of Econometrics*, 59(3), pp. 257-261.

BANKER, R. (1996). – Hypothesis Test Using Data Envelopment Analysis, *Journal of Productivity Analysis*, 7(2-3), pp. 139-159.

BANKER, R., CHARNES, A., COOPER, W. (1984). – Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science*, 30(9), pp. 1078-1092.

BANKER, R., CHANG, H., COOPER, W. (1996). – Equivalence and Implementation of Alternative Methods for Determining Returns to Scale in Data Envelopment Analysis, *European Journal of Operational Research*, 89(3), pp. 473-481.

BANNISTER, G., STOLP C. (1995). – Regional Concentration and Efficiency in Mexican Manufacturing, *European Journal of Operational Research*, 80(3), pp. 672-690.

BAUER, P. (1990a). – Recent Developments in the Econometric Estimation of Frontiers, *Journal of Econometrics*, 46(1-2), pp. 39-56.

BAUER, P. (1990b). – Decomposing TFP Growth in the Presence of Cost Inefficiency, Nonconstant Returns to Scale, and Technological Progress, *Journal of Productivity Analysis*, 1(4), pp. 287-299.

BERECHMAN, J. (1993). – *Public Transit Economics and Deregulation Policy*, Amsterdam, North-Holland.

BRADFORD, D., MALT, R., OATES, W. (1969). – The Rising Cost of Local Public Services: Some Evidence and Reflections, *National Tax Journal*, 22(1), pp. 185-202.

BYRNES, P., FÄRE, R., GROSSKOPF, S., LOVELL, C.A.K. (1988). – The Effect of Unions on Productivity: U.S. Surface Mining of Coal, *Management Science*, 34(9), pp. 1037-1053.

BYRNES, P., VALDMANIS, V. (1994). – Analyzing Technical and Allocative Efficiency of Hospitals, in: A. CHARNES, W.W. COOPER, A.Y. LEWING, L.M. SEIFORD (eds) *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston, Kluwer, pp. 129-144.

ÇAKMAK, E., ZAIM, O. (1992). – Technical Efficiency of Turkish Agriculture, *Middle East Technical University Studies in Development*, 19(3), pp. 305-316.

CHU, X., FIELDING, G., LAMAR, B. (1992). – Measuring Transit Performance Using Data Envelopment Analysis, *Transportation Research*, 26A(3), pp. 223-230.

COSTA, A. (1998). – Public Transport Efficiency and Effectiveness: Metro de Madrid, in: K. BUTTON, P. NIJKAMP, H. PRIEMUS, (eds) *Transport Networks in Europe: Concepts, Analysis and Policies*, Cheltenham, Elgar, pp. 248-264.

DERVAUX, B., KERSTENS, K., VANDEN EECKAUT, P. (1998). – Radial and Nonradial Static Efficiency Decompositions: A Focus On Congestion Measurement, *Transportation Research*, 32B(5), pp. 299-312.

FÄRE, R., GRABOWSKI, R., GROSSKOPF, S. (1985). – Technical Efficiency of Philippine Agriculture, *Applied Economics*, 17(2), pp. 205-214.

FÄRE R. GROSSKOPF, S., LOVELL, C.A.K. (1983). – The Structure of Technical Efficiency, *Scandinavian Journal of Economics*, 85(2), pp. 181-190.

FÄRE, R., GROSSKOPF, S., LOVELL, C.A.K. (1985). *The Measurement of Efficiency of Production*, Boston, Kluwer.

FÄRE, R., GROSSKOPF, S., LOVELL, C.A.K. (1994). – *Production Frontiers*, Cambridge, Cambridge University Press.

FÄRE, R., GROSSKOPF, S., PASURKA, C. (1989). – The Effect of Environmental Regulations on the Efficiency of Electric Utilities: 1969 versus 1975, *Applied Economics*, 21(2), pp. 225-235.

FÄRE, R., LOGAN, J. (1992). – The Rate of Return Regulated Version of Farrell Efficiency, *International Journal of Production Economics*, 27(1), pp. 161-165.

FÄRE, R., PRIMONT, D. (1988). – Efficiency Measures for Multiplant Firms with Limited Data, in: W. Eichhorn (ed) *Measurement in Economics*, Heidelberg, Physica-Verlag, pp. 177-186.

FARRELL, M. (1957). – The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120A(3), pp. 253-281.

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(3), pp. 305-324.

FIELD, K (1990). – Production Efficiency of British Building Societies, *Applied Economics*, 22(3), pp. 415-426.

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(3), pp. 437-455.

FØRSUND, F., HJALMARSSON, L. (1974). – On the Measurement of Productive Efficiency, *Swedish Journal of Economics*, 76(2), pp. 141-154.

FØRSUND, F., HJALMARSSON, L. (1979). – Generalised Farrell Measures of Efficiency: An Application to Milk Processing in Swedish Dairy Plants, *Economic Journal*, 89(354), pp. 294-315.

FUKUYAMA, H. (1996). – Returns to Scale and Efficiency of Credit Associations in Japan: A Nonparametric Frontier Approach, *Japan and the World Economy*, 8(3), pp. 259-277.

GANLEY, J., CUBBIN, J. (1992). – *Public Sector Efficiency Measurement: Applications of Data Envelopment Analysis*, Amsterdam, North-Holland.

GATHON, H.-J. (1989). – Indicators of Partial Productivity and Technical Efficiency in the European Urban Transit Sector, *Annals of Public and Cooperative Economics*, 60(1), pp. 43-59.

GROSSKOPF, S. (1996). – Statistical Inference and Nonparametric Efficiency: A Selective Survey, *Journal of Productivity Analysis*, 7(2-3), pp. 161-176.

JARA DÍAZ, S.R. (1982). – The Estimation of Transport Cost Functions: A Methodological Review, *Transport Reviews*, 2(3), pp. 257-278.

KEMME, D., NEUFELD, J. (1991). – The Estimation of Technical Efficiency in Polish Industry: 1961-1986, *Economic Systems*, 15(1), pp. 43-56.

KERSTENS, K. (1996). – Technical Efficiency Measurement and Explanation of French Urban Transit Companies, *Transportation Research*, 30A(6), pp. 431-452.

KERSTENS, K., VANDEN EECKAUT P. (1995). – The Economic Cost of Subsidy-Induced Technical Inefficiency: A Methodological Postscript, *International Journal of Transport Economics*, 22(2), pp. 225-229.

KERSTENS, K., VANDEN EECKAUT, P. (1998). – Distinguishing Technical and Scale Efficiency on Non-Convex and Convex Technologies: Theoretical Analysis and Empirical Illustrations, Louvain-la-Neuve, UCL (CORE Discussion Paper 9855).

KERSTENS, K., VANDEN EECKAUT P. (1999). – Estimating Returns to Scale Using Nonparametric Deterministic Technologies: A New Method Based on Goodness-of-Fit, *European Journal of Operational Research*, 113(1), pp. 206-214.

KUMBHAKAR, S.C., BISWAS, B., BAILEY, D.V. (1989). – A Study of Economic Efficiency of Utah Dairy Farmers: A System Approach, *Review of Economics and Statistics*, 71(4), pp. 595-604.

LELEU, H., DERVAUX, B. (1997). – Comparaison des différentes mesures d'efficacité technique: une application aux centres hospitaliers français, *Économie et Prévision*, (129-130), pp. 101-119.

LOVELL, C.A.K. (1993). – Production Frontiers and Productive Efficiency, in: H. Fried, C.A.K. Lovell, S. Schmidt (eds) *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford, Oxford Univ. Press, pp. 3-67.

MITRIC, S. (1988). – Organization of Urban Public Transport in France: Lessons for Developing Countries, *Transportation Research Record*, 1202, pp. 113-120.

NADIRI, M (1982). – Producers Theory, in: K.J. Arrow, M.D. Intriligator (eds), *Handbook of Mathematical Economics, Volume II*, Amsterdam, North-Holland, pp. 431-490.

OBENG, K (1994). – The Economic Cost of Subsidy-induced Technical Efficiency, *International Journal of Transport Economics*, 21(1), pp. 3-20.

PESTIAU, P., TULKENS, H. (1990). Assessing the Performance of Publich Sector Activities: Some Recent Evidence from the Productive Efficiency Viewpoint, Louvain-la-Neuve, UCL (CORE Discussion Paper 9060).

PIOT, I. (1994). – Mesure non paramétrique de l'efficacité, *Cahiers d'Économie et Sociologie Rurales*, (31), pp. 13-41.

PUCHER, J. (1988). – Urban Public Transport Subsidies in Western Europe and North America, *Transportation Quarterly*, 42(3), pp. 377-402.

SCHINNAR, A. (1993). – Tradeoffs between Efficiency and Effectiveness in Management of Public Services, in: Y. Ijiri (ed.), *Creative and Innovative Approaches to the Science of Management*, Westport (CT), Quorum Books, pp. 177-190.

SEAVER, B., TRIANTIS, K (1995). – The Impact of Outliers and Leverage Points for Efficiency Measurement and Evaluation Using High Breakdown Procedures, *Management Science*, 41(6), pp. 937-956.

SIMAR, L., WILSON, P. (1998). – Nonparametric Tests of Returns to Scale, Louvain-la-Neuve, UCL (Institut de Statistique : DP 9814).

THIRY, B., TULKENS, H. (1992). – Allowing for Inefficiency in Parametric Estimation of Production Functions for Urban Transit Firms, *Journal of Productivity Analysis*, 3(1-2), pp. 45-65.

TONE, K, SAWADA, T. (1990). – An Efficiency Analysis of Public vs. Private Bus Transportation Enterprises, in: H. Bradley (ed) *Operational Research'90*, New York, Pergamon, pp. 357-365.

TULKENS, H. (1993). – On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit, *Journal of Productivity Analysis*, 4(1-2), pp. 183-210.

TULKENS, H., WUNSCH, P. (1994). – Les Performances Économiques de la STIB au Mois le Mois, *Revue Suisse d'Économie Politique et de Statistique*, 130(4), pp. 627-646.

VALDMANIS, V. (1992). – Sensitivity Analysis for DEA Models: An Empirical Example using Public vs. NFP Hospitals, *Journal of Public Economics*, 48(2), pp. 185-205.

VITON, P. (1986). – The Question of Efficiency in Urban Bus Transportation, *Journal of Regional Science*, 26(3), pp. 499-513.

VITON, P. (1997). – Technical Efficiency in Multi-Mode Bus Transit: A Production Frontier Analysis, *Transportation Research*, 31B(1), pp. 23-39.

WILSON, P. (1993). – Detecting Outliers in Deterministic Nonparametric Frontier Models with Multiple Outputs, *Journal of Business & Economic Statistics*, 11(3), pp. 319-323.