

Methodological Advances in DEA: A survey and an application for the Dutch electricity sector

Laurens Cherchye

*Fund for Scientific Research – Flanders (FWO – Vlaanderen; Belgium),
Catholic University of Leuven, Campus Kortrijk and Center for Economic
Studies, E. Sabbelaan 53, B-8500 Kortrijk, Belgium*

Thierry Post*

*Erasmus University Rotterdam, P. O. Box 1738, 3000 DR Rotterdam,
The Netherlands*

We survey the methodological advances in DEA over the last 25 years and discuss the necessary conditions for a sound empirical application. We hope this survey will contribute to the further dissemination of DEA, the knowledge of its relative strengths and weaknesses, and the tools currently available for exploiting its full potential. Our main points are illustrated by the case of the DEA study used by the regulatory office of the Dutch electricity sector (Dienst Toezicht Elektriciteitswet; Dte) for setting price caps.

Key Words and Phrases: Data Envelopment Analysis, empirical production analysis, Dutch electricity sector.

1 Introduction

Production analysis is of key interest to economic research in academia, the business community and government institutions. Economic theory represents production as a constrained optimization problem; producers optimize their objectives subject to constraints imposed by the production technology. Empirical production analysis tests whether observed behavior is consistent with optimizing behavior and quantifies deviations from optimization (i.e. inefficiencies), and in addition tries to reconstruct the technology, to test hypothesis about the technology, and to forecast future firm behavior.

Data Envelopment Analysis (DEA; CHARNES *et al.*, 1978) gives a systematic methodology for analyzing productive efficiency. In the relatively short span of 25 years, DEA has established itself as a popular analytical research instrument and practical decision-support tool. An increasing number of applications are evidence

Laurens.Cherchye@econ.kuleuven.ac.be

*GTPost@few.eur.nl

of its popularity among researchers in Economics, Econometrics and Operations Research/Management Science, as well as practitioners in the business community and in government institutions. One interesting recent application is the DEA analysis used by the regulatory office of the Dutch electricity sector (Dienst Toezicht Elektriciteitswet; Dte) for setting price caps.

DEA has a strong nonparametric flavor; it is able to estimate efficiency with minimal prior assumptions about the production technology. This is an attractive feature, because economic theory generally does not put forward strong hypotheses about the technology, and in addition reliable specification tests in many cases are not available. In this respect, DEA has a comparative advantage relative to approaches that do require a particular parametric specification of the technology, including the regression-based Stochastic Frontier Analysis (SFA; see e.g. KUMBHAKAR and LOVELL 2000). Apart from the production assumptions, DEA also imposes minimal prior assumptions about firm behavior, as reflected in the sampling distribution of the observations. In this respect, DEA differs strongly from the 'nonparametric approach to production analysis' (NPA), which originated from the work by AFRIAT (1972), HANOCH and ROTHCHILD (1972), DIEWERT and PARKAN (1983) and VARIAN (1984). NPA assumes that firms behave according to a model of optimizing behavior (e.g. profit maximization or cost minimization). NPA tests whether the data are consistent with optimization and, if so, NPA can then produce an empirical production set, test hypotheses about the technology, and forecast future firm behavior. Consistency with optimization or efficiency is a precondition for NPA. By contrast, DEA allows for non-optimizing behavior or inefficiency, and it in fact focuses on estimating the degree of efficiency. In addition, DEA does so without strong prior assumptions about the sampling distribution. By contrast, SFA typically assumes a particular statistical distribution for the inefficiency terms. This is an unattractive feature because economic theory (including theories that allow for inefficiencies, e.g. by accounting for agency problems) is not sufficiently strong to justify a particular statistical distribution.

Despite the nonparametric orientation, the original CHARNES *et al.* (CCR; 1978) model does impose a series of assumptions that are restrictive in many research environments. Also, it was not clear initially how to test these assumptions and how to alter these assumptions. A number of extensions has been developed since the original CCR study. These extensions have turned DEA into a powerful and flexible analytical tool, capable of dealing with a wide variety of different problems. This study surveys these methodological developments.

We stress at the outset that standard textbooks such as COOPER *et al.* (2000) and THANASSOULIS (2001) provide extensive surveys of the standard DEA models and the extensions that are currently available. Our ambition is not to provide yet another all-embracing overview of the state-of-the-art of DEA. Rather, we focus on general guidelines for meaningful application of DEA. We present the developments that we think are most significant for improving statistical goodness and/or economic meaning. We stress that 'statistical goodness' and 'economic meaning' are

not fully objective criteria, because they are conditional upon our subjective opinions about the structure of real-life research environments and the objectives of the analysis. These subjective criteria inevitably involve a bias away from some work that others may see as significant and towards work that we see as significant (including our own work).

Hundreds of research papers have been written on the methodological aspects of DEA, and it is not simple to find a tractable structure that covers all significant contributions. In this paper, we structure the discussion along the lines of maintained assumptions. As is probably true for all methodologies, an ‘optimal’ model does not exist. Most models make perfectly good sense in the context of research environments that satisfy the maintained assumptions. For this reason, we structure our discussion along the lines of the maintained assumptions. We distinguish three different categories of assumptions:

1. assumptions about the data generating process;
2. assumptions about the objectives of the firm;
3. assumptions about the production technology.

The remainder of this paper is structured as follows. Section 2 introduces the elementary input-oriented model by CHARNES *et al.* (1978) or CCR-I model. Section 3 discusses advances in modeling the data generating process. Section 4 discusses advances in modeling the objectives of the firm. Section 5 discusses advances in modeling the production technology. In each section, we illustrate our main points by means of the case of the DEA study by the Dte. Finally, Section 6 summarizes our findings and sets out a number of general guidelines for a sound DEA application.

2 Preliminaries: the CCR-I model

To analyze producer behavior, we need a convenient way to summarize the production possibilities of the firm, i.e. which input–output combinations are technologically feasible. Let $x \equiv (x_1 \cdots x_m)^T \in \mathbb{R}_+^m$ denote a (non-zero) input vector and $y_j \equiv (y_1 \cdots y_s)^T \in \mathbb{R}_+^s$ a (non-zero) output vector. (Throughout the text, we will use \mathbb{R}^m for an m -dimensional Euclidean space, and $\mathbb{R}_+^m(\mathbb{R}_{++}^m)$ denotes the (strictly) positive orthant.) The production possibility set is generally defined as

$$T \equiv \{(x, y) \in \mathbb{R}_+^{m+s} | x \text{ can produce } y\}. \quad (1)$$

This set may alternatively be represented by the input requirement sets $I(y) \equiv \{x \in \mathbb{R}_+^m : (x, y) \in T\}$, $y \in \mathbb{R}_+^s$, or by the output producible sets $O(x) \equiv \{y \in \mathbb{R}_+^s : (x, y) \in T\}$, $x \in \mathbb{R}_+^m$.

Theoretically, producer behavior may be represented as a constrained optimization problem; the producer chooses an input–output vector (x, y) from the set T that optimizes an economic objective function. A frequently employed objective is cost

minimization at given output quantities y and input prices $w \in \mathbb{R}_+^m$. Associated with this objective is the traditional measure of cost efficiency:

$$\xi(x, y|T, w) \equiv \min_{x' \in I(y)} \left(\frac{wx'}{wx} \right). \quad (2)$$

This measure equals minimal cost over actual cost for given w and y . Cost efficiency is achieved if $\xi(x, y|T, w) = 1$; lower values indicate cost inefficiency.

We illustrate this measure in Figure 1, which depicts the input requirement set $I(y)$ for a particular output vector y in a two-input situation. In the following we consider four observed input vectors of $I(y)$ (labeled 1, 2, 3 and 4). Let relative prices correspond to the slope of the iso-cost line cc' ; under these prices the input vector m is cost minimizing over $I(y)$. Cost efficiency of any other element of $I(y)$ is measured relative to m . E.g., for observation 3, measure (2) equals $03'/03$, i.e. the relative radial distance between the iso-cost line through m and the iso-cost line through 3, dd' .

In many cases, prices cannot be measured accurately enough to make good use of economic efficiency measurement. For example, accounting data can give a poor approximation for economic prices (i.e. marginal opportunity costs), because of debatable valuation and depreciation schemes. In such cases, technical efficiency measures can be useful surrogate measures. For example, the DEBREU (1951)–FARRELL (1957) input efficiency measure gives a direct upper bound for cost efficiency by computing cost efficiency at the ‘most favorable’ prices:

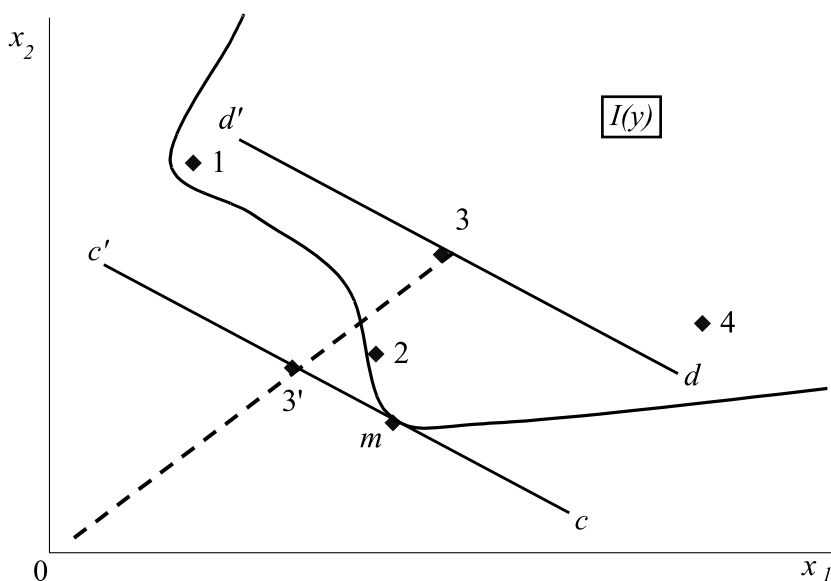


Fig. 1. Cost efficiency.

$$\theta(x, y|T) \equiv \max_{w \in \mathbb{R}_+^m : wx=1} \xi(x, y|T, w). \quad (3)$$

The Debreu–Farrell measure $\theta(x, y|T)$ thus maximizes $\xi(x, y|T, w)$ with respect to the price vector $w \in \mathbb{R}_+^m$. The restriction $wx = 1$ normalizes prices so as to exclude the (trivial) solution with all prices equal to zero, i.e. $w = 0$.

Using linear duality theory, this measure can equivalently be expressed as the minimum fraction of the input bundle that can produce the output bundle:

$$\theta(x, y|T) = \min\{\theta : (\theta x, y) \in T^+\}. \quad (4)$$

Note that the dual model does not measure efficiency relative to the original production set T , but rather relative to the expanded production set with convex and monotone input sets:

$$T^+ \equiv \{(\lambda x + (1 - \lambda)x' + s, y) : (x, y), (x', y) \in T, s \in \mathbb{R}_+^m, \lambda \in [0, 1]\}.$$

We illustrate the Debreu–Farrell input measure by recapturing our earlier example. Specifically, for observation 3, the measure equals the maximum radial contraction of 3 within the set $I^+(y)$, i.e. the convex and monotone counterpart of $I(y)$. This relative input contraction equals $03''/03$, which clearly provides an upper bound for the cost efficiency measure (i.e. $03''/03 > 03'/03$; compare with (3)).

Apart from prices, the production set is typically not fully known in practice and hence (economic or technical) efficiency cannot be measured directly. Rather, technology information is typically limited to input–output observations for a set of n comparable firms, say $S \equiv \{(x_j, y_j)\}_{j=1}^n$. To estimate efficiency from empirical data, DEA builds an empirical production set (EPS) from the data plus a set of maintained assumptions about S and T . Efficiency measures as obtained relative to the EPS are then interpreted as empirical estimators of the true efficiencies. Following the ‘minimum extrapolation principle’ (MEP; see BANKER *et al.*, 1984), the EPS is the smallest set in ‘input–output space \mathbb{R}_+^{m+s} that is consistent with the maintained assumptions. The basic CCR-I model imposes the following assumptions:

Data envelopment (DE):

$$S \subseteq T. \quad (5)$$

Strong disposability (SD):

$$T = m(T), \quad (6)$$

with $m(T) \equiv T + \mathbb{R}_+^m \times \mathbb{R}_-^s$ for the monotone hull.

Graph convexity (GC):

$$T = co(T), \quad (7)$$

with $co(T) \equiv \{(\lambda x + (1 - \lambda)x', \lambda y + (1 - \lambda)yx) : (x, y), (x', y') \in T, \lambda \in [0, 1]\}$ for the convex hull.

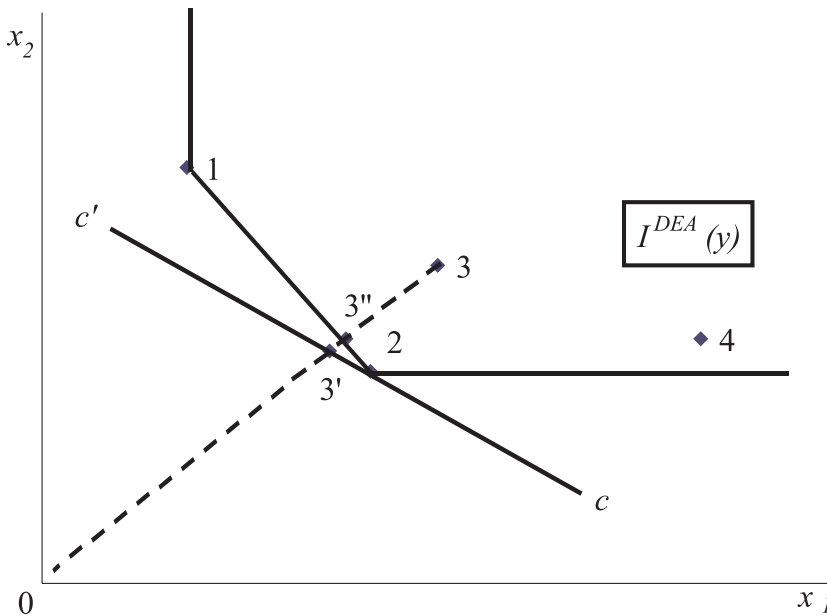


Fig. 3. DEA estimates for cost efficiency and Debreu–Farrell technical input efficiency.

produce the same output. Therefore, RU does not have an effect on the shape of $I^{DEA}(y)$, and SD and GC only affect the production set representation in input space.) Cost efficiency with respect to the set $I^{DEA}(y)$ for the observed input vector 3 equals $03'/03''$ and Debreu–Farrell input efficiency equals $03''/03$, which provide upper bounds for the cost and Debreu–Farrell efficiency measures computed from Figure 1.

The DEA approach is very useful for estimating efficiency in a conservative fashion. Specifically, if all maintained assumptions are correct, then T envelops the EPS and hence efficiency relative to the EPS is a conservative bound for true efficiency. For example, if DE, GC, SD and RU are satisfied, then $c(co(m(S))) \subseteq T$ and hence $\theta(x, y|c(co(m(S))))$ gives an upper bound to $\theta(x, y|T)$ (see the above example). In addition, the estimates can be demonstrated to be statistically consistent for a wide range of sampling distributions (e.g. BANKER, 1993, KNEIP *et al.*, 1998, and GIJBELS *et al.*, 1999). Unfortunately, the rate of convergence is low, especially if the number of input-output variables is high. Fortunately, for many application areas of current interest, large data sets are available. In addition, many extensions to the elementary models can improve the rate of convergence. We discuss these issues in greater detail in the following section.

Dte application

The regulatory office of the Dutch electricity setting (Dte) has used DEA for setting price caps for the years 2001–2003. We will use this application throughout this paper to illustrate our main points.

We start with a brief description of the Dutch electricity sector, and its different components: production, transmission and distribution. Electricity *production* in the Netherlands has traditionally been dominated by four major power producing companies: EPON, UNA, EPZ, and EZH account for approximately 60 million GWh, which is roughly 60 percent of the total power output (roughly 100 million GWh). The remaining 40 percent are accounted for by co-generated power produced by large industrial users. Domestic production is supplemented by imports from Belgium, France, Germany and Norway. Imports currently account for roughly 10 percent of the total electricity supply. Electricity *transmission* is split between the national transmission operator (TenneT) for the national 220/380 kV network high-voltage network and regional operators for transmission up to 150 kV. Finally, 18 regional electricity *distribution* companies distribute electricity to roughly 7 million electricity consumers. These distribution companies are vertically integrated with the regional network operators. We will refer to these integrated companies as Electricity Distribution units (EDUs).

The Dutch electricity market is currently in a process of transformation to a liberalized market for production and distribution. In 2000, electricity production was fully liberalized (and the foreign companies PreussenElektra, Electrabel and Reliant Energy acquired EZH, EPON, and UNA respectively). However, the market for consumer distribution still involved regional monopolies regulated by a system of price caps. For the period 2001–2003, the price caps are set on the basis of the outcomes of a DEA analysis. The analysis uses a year 2000 cross-section of data of 18 EDUs for the input–output variables listed in Table 1. (For a full description of the variables and the EDUs in our model, refer to the homepage of the regulatory office for the Dutch electricity sector (<http://www.dte.nl>).)

Dte uses the classical CCR-I model (9), i.e. efficiency is measured as cost efficiency relative to the conical convex monotone hull of the observations. Observe that Debreu–Farrell input efficiency and cost efficiency in this particular case are equivalent, because operational costs enter as the only input and input prices are assumed constant across EDUs. Table 2 gives the estimated efficiency scores.

The relationship between the efficiency scores and the price caps is almost one-to-one: over the three-year period 2001–2003, each EDU basically has to reduce its

Table 1. Input–output variables.

Inputs	Outputs
Controllable operational costs	GWh distributed
	Number of large customers
	Number of small customers
	Peak demand > 110 kV
	Peak demand < 110 kV
	Length of network
	Number of transformers

Table 2. Results.

EDU	Efficiency
COGAS	100
DELTA	100
Delftland	93.8
Essent Noord	100
REMU	70.5
ENECO	79.5
ENET	86.7
EZK	41.5
Essent Friesland	64.6
Inframosane	71.3
Essent Limburg	100
EMH	100
ONS	68.0
Essent Brabant	100
RENDON	100
Weert	100
Westland	100
NUON	74.4

prices by the degree of inefficiency (with a maximum of 8 percent per year), after correcting for productivity growth and inflation.

In the following sections, we will use this DEA application to illustrate the critical role of assumptions on the data generating process, the firm objectives and the production technology.

3 The data generating process

DEA ‘lets the data speak for itself’ as the EPS is directly constructed from the observed data (see the MEP discussed in the previous section). This nonparametric orientation comes at the price of a high sensitivity to sampling error and errors-in-variables. The original DEA studies explicitly mention only one assumption about the data generating process: data envelopment (DE; see (4)). If the maintained production assumptions are correct, then this assumption suffices to guarantee that the efficiency estimates provide conservative bounds for the true efficiency values (see Section 2). However, to obtain statistically good efficiency estimates, two further assumptions are imposed implicitly on the data generating process: (1) the observations give a good representation of the complete production technology, and (2) the observations are measured with full accuracy. Sampling error and errors-in-variables can seriously reduce the statistical goodness of the estimates. DEA evolved in an application-oriented fashion, and the statistical goodness of the estimates initially received little attention. However, more recently, these problems have been widely recognized as highly important, and a lot of research effort has

been directed at analyzing the impact of sampling error and errors-in-variables, and at accounting for these complications.

Sampling error

The original methodology assumes that the input–output vectors give a good representation of the complete production technology. For some applications, large, statistically representative data sets are available, e.g. the Longitudinal Research Database (LRD), which is a large panel data set of U.S. manufacturing plants developed by the U.S. Bureau of the Census (see e.g. BARTELSMAN and DOMS, 2000 for a discussion). High quality data are available also for financial institutions (e.g. the Bankscope data set comprises high-quality panel-data for thousands of financial institutions; see BERGER and HUMPHREY (1997), for a recent survey of applications in this area). However, many applications involve small samples, e.g. due to a lack of homogeneous reference units or the proprietary nature of the data. For example, the Dte study discussed in Section 2 uses a cross-section data set of only 18 electricity distribution companies. Small samples generally do not give a full representation of the technology. Hence, inefficient firms can be wrongly classified as efficient, or ‘true inefficiency’ can be substantially underestimated.

Knowledge of the sampling distribution can correct for small sample bias and construct confidence intervals. Two approaches exist to estimate the sampling distribution of the estimates: (1) analytical asymptotic analysis and (2) bootstrap techniques. In some cases, it is possible to analytically derive the asymptotic sampling distribution (e.g. GIJBELS *et al.*, 1999). However, there currently exist results for the single-input single-output case only, and it is not clear how to generalize these results to the general multi-input multi-output case. In addition, the approach requires the estimation of unknown distribution parameters, which introduces additional statistical noise and imprecision. Alternatively, the sampling distribution can be approximated using the bootstrap, a versatile statistical resampling technique, first introduced by EFRON (1979) and EFRON and Gong (1983). Bootstrapping involves the repeated simulation of the data generating process and the application of the original estimator to each simulated sample so that the resulting estimators mimic the sampling distribution of the original estimator. The bootstrap is a well-established tool to analyze the sensitivity of empirical estimators to sampling variation in situations where the sampling distribution is difficult or impossible to obtain analytically. The convenient LP structure of the efficiency estimators implies that it is possible in the DEA case to substitute brute computational force to overcome analytical intractability. A bootstrapping procedure that is especially tailored to DEA is developed in SIMAR and WILSON (1998).

Errors-in-variables

Apart from ignoring sampling error, the original methodology also assumes the input–output vectors are measured with full accuracy. In practice, data are almost

always contaminated by errors-in-variables. For example, much empirical research uses accounting data that can give a flawed representation of the underlying economic values, e.g. because of debatable valuation and depreciation schemes. Since efficiency analysis relies on comparison with extreme observations, the results are extremely sensitive to errors; a single outlier can substantially affect the outcomes for the entire sample. The impact of errors-in-variables is subtly different from the impact of sampling error. Efficiency estimation relies on measuring the distance of the evaluated production vector from the frontier of the empirical production set. It is possible in many cases to find a statistically good estimate for the frontier in large samples. Unfortunately, cross-section data sets contain only a single observation for the evaluated firm. For this reason, it is impossible, even in large samples, to obtain a robust efficiency estimate from cross-section data if noise is important. Fortunately, the use of panel data can reduce this problem, admittedly at the cost of imposing assumptions about the evolution of the firm-level efficiencies over time.

Various approaches have been proposed to account for errors-in-variables in DEA. We distinguish four different approaches: (1) outlier detection, (2) sensitivity analysis, (3) chance constrained programming, and (4) nonparametric regression.

There exist procedures for outlier detection that can help improve the quality of the data set prior to the analysis. For example, the WILSON (1995) procedure relies on assessing the impact of excluding observations from the data set. If the exclusion from the data set of a particular firm has a large impact on the efficiency scores of the remaining firms, few additional firms support the input–output vector of that firm. Consequently, the observation is a potential outlier and it is assigned a high priority for follow-up inspection. A careful follow-up inspection of the data could reveal whether the observation has to be adjusted, omitted or can be included. This approach is very useful for detecting errors in the efficient firms and for improving the EPS. However, it critically depends on the ability of the analyst to identify the outliers from the set of prioritized observations. Also, the approach does not detect errors for the inefficient firms; therefore, it is of limited use for improving efficiency estimation.

Various procedures have been developed for analyzing the sensitivity or robustness of the efficiency results (COOPER *et al.*, 2001) provide a survey of all currently available techniques. Sensitivity analysis checks the robustness of the results with respect to deviations of observations from their initial location in input–output space. For example, CHARNES *et al.* (1992) compute a particular ‘region of stability’, a cell such that all perturbations within the cell preserve the observation’s current classification- efficient or inefficient. These procedures are very useful for obtaining a first impression of the reliability of the results. However, they do not provide a rigorous analysis of the impact of errors. First, the techniques generally are concerned with partial perturbations of selected data entries only (generally chances to the data of the evaluated firm only). Second, the techniques do not account for the statistical distribution of the errors, and hence it is difficult to interpret the sensitivity measures in a statistically meaningful way.

The chance constrained programming approach (LAND *et al.*, 1994, OLESEN and PETERSEN, 1995, COOPER *et al.*, 1996, 1998, LI 1998) does explicitly account for the statistical distribution of the errors. POST (2001a) derives from mean-variance theory a rationale for some chance constrained programming models for the purpose of selecting performance benchmarks or targets. In his model, the rationale for chance constraints comes from the desire for well-diversified benchmarks if the performance improvements from benchmarking are related to the distance relative to the benchmark (i.e. the potential improvements). Diversification is attractive if the actual improvements are a concave function of the potential improvements (i.e. there are diminishing returns to benchmarking) or alternatively if the firm management is risk averse with respect to the actual improvements. However, to the best of our knowledge, there currently exists no evidence that the chance-constrained approach is also useful for improving the statistical goodness of the efficiency estimates. Also, the chance constrained programming models typically require prior specification of the full variance-covariance matrix of the errors, which does not seem fully consistent with the non-parametric orientation of DEA.

A statistically more sound approach is the application of nonparametric regression techniques to the problem of frontier estimation and efficiency estimation. KNEIP and SIMAR (1996) apply kernel estimation, an existing nonparametric regression technique, to estimate the EPS in nonparametric fashion and measure efficiency relative to the EPS. For a wide range of distribution structures, this approach yields asymptotically unbiased efficiency estimates, and the asymptotic variance goes to zero if additional time series observations are introduced. However, as argued by POST, CHERCHYE and KUOSMANEN (PCK; 2002), the efficiency estimates are not statistically efficient, i.e. it is possible to find estimates with a lower variance in finite time series. The evaluated firm enters twice in the analysis: the first time for determining the evaluated input-output vector and the second time for constructing the EPS. PCK derive from scratch an entirely new nonparametric technique, especially tailored to the problem of efficiency estimation. This technique does account for the correlation between the EPS and the evaluated vector, and the variance of the efficiency estimates can be substantially lower than that of the kernel approach, especially if the signal-to-noise ratio is low. In addition, the technique is computationally more attractive, as the efficiency estimates can be solved using a simple enumeration algorithm.

Dte application

We return to our Dte application to illustrate the above points. The efficiency results presented in the previous section are based on a very small cross-sectional data set and they are likely to be affected by sampling error in a nontrivial manner. Still, the Dte presents the results as accurate estimates without quantifying the reliability of the results, e.g. by standard deviations, confidence intervals, or t-statistics. To assess the sensitivity to sampling variation, we applied the SIMAR and WILSON (1998) bootstrap procedure. Table 3 gives the results as measured by the bias-corrected

Table 3. Robustness for sampling error.

EDU	Original efficiency estimate	Bias-Corrected estimate	Standard deviation
COGAS	100	94.4	7.8
DELTA	100	89.1	18.6
Delftland	93.8	90.5	7.1
Essent Noord	100	79.5	9.8
REMU	70.5	69.3	2.5
ENECO	79.5	59.6	3.2
ENET	86.7	53.0	3.3
EZK	41.5	41.1	8.0
Essent Friesland	64.6	63.4	2.6
Inframosane	71.3	70.4	3.4
Essent Limburg	100	93.5	9.9
EMH	100	95.0	6.2
ONS	68.0	66.2	4.4
Essent Brabant	100	94.3	8.6
REND0	100	91.9	10.8
Weert	100	88.7	14.5
Westland	100	92.0	8.4
NUON	74.4	72.4	3.6

efficiency estimate and the standard deviation found after 10,000 replications. The results demonstrate that the efficiency estimates are highly sensitive to sampling variation. For example, the efficiency estimate for DELTA is very unreliable; it has a bias of 10.9 percent points and a standard deviation of 18.6.

Apart from sampling error, the Dte results may be flawed by errors-in-variables. The critical variable in the analysis, operational cost, is an accounting variable. There is good reason to expect that this variable contains a substantial noise component, especially given the degrees of freedom that firms have in allocating costs to different time periods and (for multi-utilities) to different activities (gas, water, electricity and other). This is also reflected in the substantial corrections that Dte has performed on the data and the results. Still, Dte treats the data as perfectly accurate and it does not attempt to assess the sensitivity of the results to data perturbations.

VARIAN (1985) developed a technique to measure the quality of the data set in the NPA approach (which assumes inefficiencies do not occur; see the Introduction). More specifically, he measures data quality as the standard deviation of the data perturbations required to reject (at a given level of significance) the null hypothesis that all firms are equally efficient. He applied the technique to a data set of electricity producers in California. We apply the technique to the data set of Dutch EDUs. The results suggest that the quality of the data needs to be extremely high to discriminate between the different EDUs; we can reject the null hypothesis (at a significance level of 5 percent) only if the standard error of data perturbations is less than 1 percent of the standard deviation of the observations. Given the accounting problems discussed above, this level of accuracy seems highly unlikely, and we conclude that the data material is simply not sufficient to measure relative efficiency reliably.

4 The objectives of the firm

In Section 2 we used the cost efficiency measure as the performance measure, and we introduced the Debreu–Farrell input efficiency measure as a convenient upper bound approximation. In fact, most DEA studies focus on technical efficiency, and use Debreu–Farrell measures to gauge efficiency. For completeness, we note that a multitude of alternative technical efficiency measures are available in DEA (although they are generally used to a much lesser extent than Debreu–Farrell measures); e.g. Koopmans technical efficiency measures (see KOOPMANS, 1951; and CHARNES *et al.* 1985) and Russell technical efficiency measures (see FÄRE and LOVELL, 1978). Recently, CHAMBERS *et al.* (1996, 1998) introduced the directional distance function framework, which encompasses a whole range of technical efficiency measures that have an attractive dual interpretation in terms of economic efficiencies. (CHAMBERS *et al.* focus on convex production sets; CHERCHYE *et al.*, 2001a provide extensions for non-convex production sets. Section 5 discusses the rationale for relaxing convexity.)

There are at least three reasons for focusing on technical efficiency. First, it is interesting in many cases to decompose economic efficiency into components of technical efficiency and allocative efficiency (see e.g. the seminal treatment by FARRELL, 1957). Second, in many cases prices cannot be measured accurately enough to make good use of economic efficiency measurement and technical efficiency measures can serve as surrogate measures for economic efficiency. For example, the Debreu–Farrell input efficiency measure gives a direct upper bound for cost efficiency (see Section 2). Thirdly, the existing economic efficiency measures are based on the neoclassical theory of the firm under perfect competition and full certainty and need not be economically meaningful under imperfect competition or uncertainty. However, some technical efficiency measures remain meaningful for theories of the firm that do allow for imperfect competition or uncertainty (see e.g. KUOSMANEN and POST, 2002, and CHERCHYE *et al.*, 2002).

Still, it is important to bear in mind the following four considerations when using technical efficiency measures:

1. Some technical efficiency measures give conservative bounds to economic efficiency measures. However, technical efficiency measures generally do not allow for comparison between firms or ranking of firms on the basis of economic efficiency measures. This fundamental insight is ignored in many cases; there are numerous studies that compare the technical efficiency of firms (or groups of firms) in a cross-section or a time-series.

Figure 4 recaptures our earlier example to illustrate this point. Suppose we evaluate input vectors 3 and 4. In terms of economic (cost) efficiency, using the relative prices that correspond to the iso-cost line cc' , observation 3 is clearly more efficient than observation 4 (i.e. $03'/03 > 04'/04$). However, observation 4 outperforms observation 3 in terms of the Debreu–Farrell input measure (i.e. $04''/04 > 03''/03$).

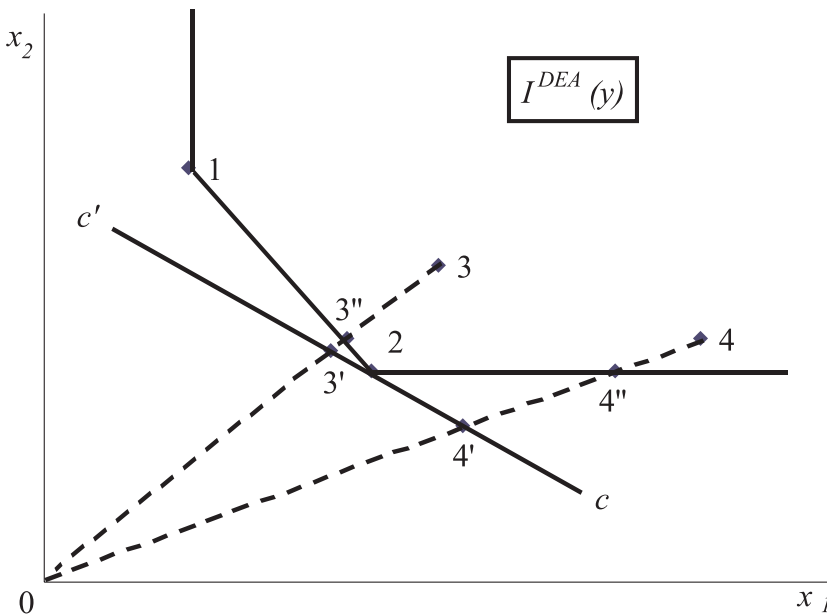


Fig. 4. Technical efficiency does not allow for cardinal measurements or ordinal ranking based on economic efficiency.

2. Technical efficiency measures can involve little power from an economic perspective. Firms may succeed in minimizing relatively inexpensive inputs and maximizing inexpensive outputs. In many cases, substituting expensive inputs by inexpensive ones and substituting inexpensive outputs by expensive ones, even in a technically inefficient way, can improve economic performance. To increase the power of the efficiency measurement tools, one can bring in additional price information, i.e. bound the set of relative prices in the (price) formulation of the technical efficiency measure (see e.g. expression (2) for Debreu–Farrell input efficiency). While fully reliable price information is usually hard to obtain, limited price information is often available. KUOSMANEN and Post (2001) provide a systematic framework for including imperfect price information into the analysis using weight restriction tools from the literature on DEA assurance regions and cone-ratio DEA models (CHARNES *et al.*, 1989; surveys of techniques to include weight-restrictions in DEA are provided by ALLEN *et al.*, 1997 and PEDRAJA-CHAPARRO *et al.*, 1997).

Related to this problem, the shadow prices of Debreu–Farrell type efficiency measures, e.g. $\arg \max_{w \in \mathbb{R}_+^m; wX=1} \zeta(x, y|T, w)$, may be poor estimates of the true (unknown) prices. In fact, many shadow prices generally equal zero. Zero shadow prices imply *slacks*; the evaluated firm may reduce outputs and/or increase inputs without changing the efficiency measure (see e.g. FÄRE *et al.*, 1994 and COOPER *et al.*, 2000 for in-depth discussion). The weight restriction tools can also help to reduce this problem. Further, a number of

alternative technical efficiency measures have been presented that correct for the slack problem; see e.g. TONE (2001) and references therein. For brevity and transparency, we abstract from a full treatment of the slack literature. Still, we stress that the economic meaning of the slack-corrected measures has been left largely unexplored in the current literature, while it is precisely the intuitive interpretation of 'economic efficiency at most favorable prices' that makes the use of Debreu–Farrell type measures attractive (see e.g. RUSSELL, 1985).

3. Many technical efficiency measures are not economically meaningful if prices are uncertain or endogenous, and the technical efficiency measures that do remain economically meaningful may involve only minimal power. For this reason, it is desirable to develop a framework for efficiency measurement specially tailored to industries where uncertainty and/or endogeneity are relevant. A whole theoretical literature has emerged on the effects of uncertainty and endogeneity on firm behavior (e.g. NEGISHI, 1961, MCCALL, 1969, SANDMO, 1971, BATRA and ULLAH, 1974, HEY, 1979, CHAVAS, 1985, APPELBAUM and KATZ, 1986, and DALAL, 1990). However, the empirical implementation of these theories is extremely difficult, because the information requirement is enormous. For example, one generally needs detailed assumptions about the market structure to choose from different theoretical models. Also, since the appropriate model depends highly on the idiosyncrasies of the industry under evaluation, it is practically impossible to find a model that is general enough to apply to a wide variety of industries. Finally, in many cases the computational burden associated with detailed theoretical models practically excludes empirical application. These complications at least partly explain why the empirical research has not 'caught up' with the theory, and predominantly represents production as a problem of optimization at exogenously fixed and fully certain prices. The parametric approach to empirical production analysis has made some important steps towards including these complications (e.g. JUST, 1974, ANTONOVITZ and ROE, 1986, APPELBAUM, 1982, 1991, APPELBAUM and KOHLI, 1997, and APPELBAUM and ULLAH, 1997). Recently, KUOSMANEN and POST (2002) and CHERCHYE *et al.* (2002) have developed tools for including price uncertainty and imperfect competition in DEA. Interestingly, these tools preserve the minimal information requirement and the computational simplicity of standard DEA tools, and hence are directly applicable for practical research problems.
4. Finally, an important practical issue is the selection of the input–output variables. It is important to reflect upon which inputs and outputs can be conceived as given (exogenous) and which inputs and outputs are controllable (endogenous); a meaningful efficiency measure should include controllable dimensions only. A related issue concerns the distinction between inputs and outputs; obviously, labeling a variable as an output while it is actually an input (or vice versa) can seriously distort the analysis. We demonstrate the

importance of this (seemingly straightforward) issue in our application in Section 6.

Dte application

We mentioned in the previous section that Debreu–Farrell efficiency and cost efficiency are equivalent in the Dte application. In addition, the cost efficiency measure that is used indeed seems a reasonable/meaningful efficiency measure, as electricity firms operate in a regulated environment where the output quantities and prices are fixed by approximation.

One debatable issue is the inclusion of the network length and the number of transformers as outputs. In empirical production analysis, physical capital is typically treated as a substitute for controllable inputs and it typically enters in short-run cost functions as fixed input rather than as output. It is possible theoretically to include physical capital variables as output if these variables are complements rather than substitutes for the controllable inputs. However, Dte fails to adequately motivate this choice for this particular case, and also the sensitivity of the results to these assumptions has not been assessed. To assess the robustness of the results, we estimated the efficiencies with physical capital treated as fixed input rather than as output (using the RUGGIERO, 1998, model). Table 4 shows the resulting efficiency scores.

Some results are very sensitive to the choice of the efficiency measure. For example, EZK has an efficiency score of 41.5 in the Dte analysis (and it has to reduce prices by 9 percent each year). However, if we include physical capital as fixed input rather than as output, then the EDU is fully efficient. This finding does not mean that EZK is truly efficient; the efficiency classification is caused by the fact that EZK has the shortest network and hence it can not be compared with a firm that uses less fixed input. However, the finding does mean that the Dte results are not robust with respect to non-trivial changes in the efficiency measure.

5 The production technology

DEA is often credited for not imposing a functional form for the production frontier. However, the maintained assumptions of SD, GC and RU in the CCR-I model discussed in Section 2 are overly restrictive in many research environments. We discuss the empirical problems associated with these assumptions below. In principle, it is possible to analyze the data without imposing additional production assumptions. In fact, the NPA approach (see Section 1) typically does not impose assumptions other than DE. A similar approach is possible in DEA, as is demonstrated in, for example, TULKENS and VAN DEN EECKAUT (1999). Unfortunately, the use of minimal assumptions generally is associated with minimal power in small samples. Hence, the model specification involves a difficult

Table 4. Robustness with respect to the efficiency measure.

EDU	Efficiency physical capital as output	Efficiency physical capital as fixed input
COGAS	100	100
DELTA	100	100
Delftland	93.8	93.8
Essent Noord	100	100
REMU	70.5	70.5
ENECO	79.5	79.5
ENET	86.7	86.7
EZK	41.5	100
Essent Friesland	64.6	65.1
Inframosane	71.3	100
Essent Limburg	100	100
EMH	100	100
ONS	68.0	100
Essent Brabant	100	100
RENDON	100	100
Weert	100	100
Westland	100	100
NUON	74.4	74.4

trade-off between specification error and power. It is therefore desirable to have a wide variety of models associated with a wide variety of different assumptions, and to develop a way to select from these models. A wide variety of different models has been developed. We discuss some of these models below, classified by the maintained assumption of the CCR-I model: SD, GC and RU. Further, various models have been proposed to account for the uncertain and dynamic nature of production (in the standard model, production is certain and static). Unfortunately, there currently are no clear guidelines for selecting from these models. As we will discuss below, economic theory and empirical specification tests are unlikely to provide effective guidance in model selection. Therefore, it is important to look for prior knowledge, e.g. from engineering knowledge of the industry under evaluation. Also, it is important to assess the economic motivation for different assumptions, and to assess the sensitivity to the model selection if prior information is limited and samples are small. Finally, the lack of guidance for model selection gives another rationale for further investing in large data sets of high quality.

Economic theory

The motivation of some assumptions lies in economic duality theory. Specifically, some assumptions follow from the objective to approximate particular economic efficiency measures using technical measures. For example, if we use the Debreu–Farrell input measure to approximate cost efficiency (see Section 2), then we can harmlessly impose disposability and convexity for the input sets, i.e. $\theta(x, y|T) = \theta(x, y|T^+)$ (compare with (3)). (The duality-based motivation for the assumptions of

SD, GC and RU is discussed below.) The duality argument is frequently misunderstood; see for example the recent exchange between THRALL (1999) and CHERCHYE *et al.* (2000). It is commonly believed that the production assumptions that can be justified by duality are actually *required* to ensure that the technical efficiency measure is economically meaningful, and dropping these assumptions is *harmful*, i.e. reduces the economic meaning. However, the contrary is true; the assumptions that can be justified by duality are *harmless* and omitting them does not affect the efficiency estimates (for example, again, $\theta(x, y|T) = \theta(x, y|T^+)$). Further, the assumptions that are harmless for one economic efficiency measure may be harmful for other measures. For example, the assumptions of SD and GC are harmless for analyzing profit efficiency, but they are harmful in the context of cost efficiency (see below). For these reasons, caution is required in using duality to justify maintained production assumptions.

Apart from duality, economic theory only forwards very weak guidelines for modeling the physical production possibilities, like the fundamental notion of *scarcity*, which in this context implies that not all input–output combinations are feasible. Another fundamental economic result is Turgot's *law of diminishing returns*. This law also does not have very strong empirical implications. The law simply states that if we hold one input constant, then the marginal productivity of other inputs decreases ultimately (i.e. for 'infinitely high' production levels). This law has minimal implications for cases where all inputs are variable and/or for finite production levels.

Specification tests

Unfortunately, there currently are no reliable empirical specification tests in DEA. Various specification tests have been proposed, based on formal statistical tests (e.g. the F-tests and the Smirnov type test by BANKER, 1993) and bootstrap approaches (SIMAR and WILSON, 2002). In the nonparametric tradition, these tests use minimal assumptions for the sampling distribution, and they are asymptotic by nature. These tests can be very useful for hypothesis testing in large samples. However, the issue of specification testing is relevant mostly for small samples. In large samples, models with minimal assumptions are preferred, because they give statistically consistent results (see Section 2) and minimize specification error. Therefore, the tests have little value added in large samples. Unfortunately, none of these tests has been demonstrated to possess acceptable size and power in small samples. Also, it seems very difficult to develop a test that does apply in small samples. Such a test would have to rely on assumptions about the sampling distribution, which reflects the behavior of the firm, including the level of inefficiency. Economic theory is useful for selecting the appropriate model of optimizing behavior and the associated efficiency measure. However, economic theory does not forward strong implications about deviations from optimizing behavior, i.e. the distribution of the inefficiency values.

In this respect, it is interesting to contrast the DEA approach with the NPA approach discussed above. In the NPA approach, VARIAN (1984) did develop nonparametric tests for production assumptions. However, those tests rely on the

maintained hypothesis of *rationalization* (see also BANKER and MAINDIRATTA, 1988), i.e. all firms are assumed to behave according to optimizing behavior. Therefore, this approach can not be used for measuring deviations from optimization or inefficiency. It seems fundamentally impossible to simultaneously estimate efficiency and to test production hypotheses. Without a maintained hypothesis about the level of efficiency, one can never disentangle violations of the evaluated production assumption from violations of optimizing behavior or inefficiencies.

Strong disposability (SD)

Monotonicity can be justified by duality theory. In many theories of the firms, the economic objective of the firm is increasing in output and decreasing in input. In the context of these models, assuming monotonicity does not affect the results of the primal model formulation, and it is required for the dual formulation. However, if the economic justification does not exist (e.g. if increasing output reduces revenues, as can be true in case of imperfect competition), or alternatively if the objective is to decompose economic efficiency into components of technical and allocative efficiency, then monotonicity assumptions can be debatable. Monotonicity excludes *congestion*, which is frequently observed, e.g. in agriculture and transportation, as pointed out for example by FÄRE and SVENSSON (1980) and FÄRE and GROSSKOPF (1983). Ways for dealing with congestion (by weakening or dropping the monotonicity axiom) have been proposed in the DEA literature (most notably by FÄRE and GROSSKOPF, 1983; FÄRE *et al.*, 1983a, 1985; and BROCKETT *et al.*, 1998). We will not discuss these in detail in this paper, but refer to CHERCHYE *et al.* (2001b) for a recent assessment of congestion analysis within DEA.

Graph convexity (GC)

Duality theory justifies GC for the purpose of measuring profit efficiency; profit is a linear function of inputs and outputs and hence imposing GC is harmless for measuring profit efficiency. However, GC is also frequently used for efficiency measures other than profit efficiency, like in the CCR-I model (9). Unfortunately, there does not seem to exist a valid motivation for convexity assumptions, apart from economic duality theory. For example, to the best of our knowledge, the “*law of diminishing marginal rates of substitution*”, as referred to by PETERSEN (1990), BOGETOFT (1996), and BOGETOFT *et al.* (2000) as a justification of convex input and output sets, is not documented in microeconomic production theory. (Further, to the best of our knowledge, there is no economically meaningful objective function that would justify this particular assumption of convex input and output sets. The most popular economic efficiency measures are cost efficiency, revenue efficiency and profit efficiency. Cost efficiency justifies convex input sets, revenue efficiency justifies convex output sets, and profit efficiency justifies GC. However, these objective functions do not justify convex input sets and convex output sets simultaneously, without justifying convexity for the entire T .)

In fact, convexity assumes away (1) indivisible inputs and outputs, (2) economies of scale, and (3) economies of specialization (=diseconomies of scope). The economic importance of these phenomena was already stressed by Farrell in his famous 1959 article “The Convexity Assumption in the Theory of Competitive Markets”, Section II (entitled ‘The importance of non-convexities’):

‘A glance at the world about us should be enough to convince us that most commodities are to some extent indivisible and that many have large indivisibilities. Similarly, whenever one refers to “economies of scale” or of “specialization”, one is pointing to concavities [=departures from convexity (CP)] in production functions. There is thus no need to argue the importance of either indivisibilities or concavities in production functions – the former are an obvious feature of the real world, and the latter have constituted a central topic in economics since the time of Adam Smith.’

FARRELL (1959, pp. 378–379)

For this reason, it is important to develop non-convex models for efficiency measures other than profit efficiency. Recent research has paid considerable attention to relaxing the overall convexity assumption (GC). For example, DEPRINS *et al.* (1984) and TULKENS (1993) dropped convexity altogether in the so-called free disposal hull (FDH) models. PETERSEN (1990) and BOGETOFT *et al.* (2000) replaced convexity of T with the somewhat milder assumption of convexity of input sets and output sets. Next, BOGETOFT (1996), CHANG (1999), DEKKER and POST (2001) and POST (2001b, 2001c) have considered convexity of either input sets or output sets. Finally, POST (2001c), and KUOSMANEN (2001) replaced convexity by the modified properties of ‘*transconvexity*’ and ‘*conditional convexity*’ respectively.

Ray unboundedness (RU)

Duality theory justifies RU for the purpose of measuring profit efficiency in the long-run, i.e. if the maximum possible profit equals zero. However, RU is frequently considered as overly restrictive for purposes other than analyzing long-run profit maximizing behavior. Many production activities exhibit increasing returns-to-scale (IRS) and/or decreasing returns-to-scale (DRS) (see e.g. FARRELL, 1957, for early accounts). FÄRE, *et al.* (1983b), BANKER *et al.* (1984), GROSSKOPF (1986) and SEIFORD and THRALL (1990), among others, have discussed implementation of alternative returns-to-scale axioms in DEA. Dropping RU is a necessary condition for developing a variable-returns-to-scale (VRS) model. However, to develop a VRS model, also GC has to be dropped, as GC is inconsistent with IRS (see e.g. PETERSEN, 1990).

Production uncertainty

In many industries, the outcomes of the production process are affected in a nontrivial way by external risk factors that are beyond the control of the firm. For example, bank performance generally depends on the uncertain influence of uncontrollable factors such as interest rates, foreign exchange rates and the business

cycle. Similarly, in agricultural and environmental production models, uncontrollable climatic and pest factors can substantially affect production. The parametric approach to production analysis has made some steps towards including production uncertainty in the analysis (e.g. CHAMBERS and QUIGGIN, 1998, 2000, POPE and JUST, 1996, 1998 and MOSCHINI, 2001). However, as far as we know, the nonparametric approach currently does not account for production uncertainty. An important difficulty is how to model uncertainty without imposing overly restrictive structure and compromising the nonparametric orientation. Still, as discussed in POST and SPRONK (2000) it is possible to include production uncertainty by combining DEA with multi-factor risk models and stochastic dominance conditions.

Production dynamics

Apart from ignoring uncertainty, the standard methodology does not acknowledge the dynamic nature of the production process. In many industries the problem of production is dynamic, in the sense that current outputs are the fruits of past inputs in addition to current inputs, and that current inputs yield future outputs in addition to current outputs. For example, in banking, substantial amounts of input can be used to acquire market share or customer goodwill, which generate output in the future in addition to the present. However, the standard methodology considers production as a problem of static optimization. Therefore, firms that appear inefficient may be efficient in reality, if a substantial part of their current inputs are used to generate future outputs. Some approaches to account for dynamics have been proposed. SENGUPTA (1995) presents a dynamic DEA model by introducing the shadow values of quasi-fixed inputs and their optimal paths into an analytic linear programming problem. FÄRE and GROSSKOPF (1996) formulate several kinds of intertemporal substitution among inputs, outputs and intermediate inputs using a network theory by which more realistic production processes across periods can be described.

A related issue is *technological change* (progress or regress) over time. This issue is particularly relevant if panel data are used and efficiency is computed for consecutive periods. For example, if production possibilities change over time, then it may be 'unfair' to compute the efficiency of a firm in one period of time by comparing it with firm observations associated with another period of time. Recent extensions of the standard DEA models allow for disentangling inefficiency and technological change; see for example, the literature on Malmquist productivity indices (e.g. FÄRE *et al.*, 1994 and the survey by FÄRE *et al.*, 1998). Still, this literature predominantly focuses on the theoretical case where the frontier is fully known, and further research is needed to study the sampling properties of the existing estimates for technology change, and possibly to develop new, statistically good estimates.

Dte application

The original Dte model uses the CCR model, which assumes SD, GC and RU. Despite their problematic nature (see above), the assumptions are not adequately motivated, and the sensitivity of the results for different assumptions is not analyzed.

Graph convexity and ray unboundedness are potentially harmful because they exclude the possibility of IRS (which is a frequently cited motivation for the mergers and acquisitions in the electricity sector). (Dynamic issues are not relevant for this application given the cross-sectional nature of the data set.)

The Dte acknowledges the possibilities of IRS and it motivates the model specification by referring to the possibilities that firms have to adjust the scale of operation by means of mergers and acquisitions. This argument implicitly refers to duality theory for analyzing long-run profit maximizing behavior in competitive industries. However, the argument is problematic for at least two reasons. First, Dte itself is the single most important obstacle for further concentration in the electricity sector; it opposes, for example, a merger of NUON and ESSENT, the two largest EDUs. Second, the argument is not consistent with the use of cost efficiency or Debreu–Farrell input efficiency as the efficiency measure. Again, cost efficiency is based on a model where output quantities and prices are not controllable, and economic duality theory cannot justify GC and RU for cost efficiency.

For these reasons, the maintained production assumptions are not adequately motivated. In addition, the Dte does not report the sensitivity of the results to different sets of production assumptions. To assess the robustness of the results, we estimated the efficiencies using the BANKER *et al.* (BCC; 1984) model (which drops RU) and the DEPRINS *et al.* (1984) FDH model (which drops both RU and GC).

The results displayed in Table 5 suggest that the efficiency estimates are very sensitive to changes in the production assumptions. If we drop RU while maintaining the other production assumptions (and use the BCC model), then the

Table 5. Robustness for different production assumptions.

EDU	Efficiency CCR	Efficiency BCC	Efficiency FDH
COGAS	100	100	100
DELTA	100	100	100
Delftland	93.8	100	100
Essent Noord	100	100	100
REMU	70.5	89.8	100
ENECO	79.5	100	100
ENET	86.7	98.8	100
EZK	41.5	100	100
Essent Friesland	64.6	67.0	100
Inframosane	71.3	100	100
Essent Limburg	100	100	100
EMH	100	100	100
ONS	68.0	69.2	100
Essent Brabant	100	100	100
REND0	100	100	100
Weert	100	100	100
Westland	100	100	100
NUON	74.4	100	100

classification of 5 EDUs (Delftland, ENECO, EZK, Inframosane, NUON) changes from inefficient to efficient. If we also drop GC (and use the FDH model), then all firms are classified as fully efficient. Again, these findings do not imply that all firms are truly efficient; the sample is much too small for that conclusion. However, the findings do demonstrate that the Dte results are not robust with respect to non-trivial changes in the maintained production assumptions. Briefly, the analysis of Dte critically depends on a number of debatable production assumptions (RU, GC). These assumptions make sense in a long-run equilibrium model for a competitive industry, but they are overly restrictive for a non-competitive and regulated industry like the electricity sector.

6 Conclusions

The original DEA models suffer from a number of limitations that reduce their practical applicability. We have discussed these limitations on the basis of the maintained assumptions regarding the nature of the production technology, the economic objective of the firm, and the quality of the data material. Fortunately, a number of methodological advances have greatly increased the flexibility of DEA to deal with a wide variety of research environments. Using these advances in many cases is a prerequisite for successful application. Our conclusions are best summarized by the following propositions for a sound application of DEA.

PROPOSITION 1 (DATA GENERATING PROCESS): *The key to obtaining statistically good efficiency estimates is the availability of high-quality data, preferably large panel data sets. The tools currently available for accounting for sampling error and errors-in-variables can help to improve the statistical goodness of the results, as well as to quantify goodness of the results.*

PROPOSITION 2A (FIRM OBJECTIVES): *Technical efficiency measures are useful proxies for economic measures. However, we do not recommend using these measures for comparison between firms or ranking of firms.*

PROPOSITION 2B (FIRM OBJECTIVES): *Using economic efficiency measures can substantially improve the power of the analysis. However, it also introduces the risk of specification error. Sometimes, prior knowledge allows for the selection of the appropriate efficiency measure. Otherwise, we recommend a thorough assessment of the sensitivity to different efficiency measures; many different models are now available.*

PROPOSITION 3 (PRODUCTION TECHNOLOGY): *Including production information can substantially improve the power of the analysis in small samples. However, it can also introduce specification error. Economic theory and empirical specification tests are*

unlikely to give effective guidance in the specification of the appropriate production assumptions. Unless there are convincing prior reasons for a particular specification (e.g. engineering knowledge of the industry under evaluation), assessment of the sensitivity of the results to different model specifications is recommended; many different models are now available.

Unfortunately, many studies ignore the limitations of the standard models as well as the recent advances, which casts serious doubt on the reliability of the outcomes. For example, our results suggest that there is good reason to doubt whether the DEA analysis used by Dte yields anything but noise. Personally, we think it is worrying that far-reaching policy decisions like setting price-caps are based on such weak analysis. This is especially worrying because more reliable results could have been obtained at minimal additional effort. It would have been relatively simple to construct a data set of better quality, preferably a large, international panel data set that includes data of all activities of the utilities (gas, water, electricity and other), observed over multiple time periods. Also, it is relatively simple to assess the sensitivity of the results with respect to key assumptions that are maintained with respect to the data generating process, the firm objectives and the production technology.

We hope this survey will contribute to the further dissemination of the knowledge of DEA, its relative strengths and weaknesses, and the tools currently available for exploiting its full potential.

Acknowledgements

The paper forms part of a research program on the fundamentals of the nonparametric methodology. A detailed description of the research program is available at <http://www.few.eur.nl/few/people/gtpost/program.htm>. We thank two anonymous referees for helpful comments and suggestions.

References

- AFRIAT, S. (1972), Efficiency estimation of production functions, *International Economic Review* **13**, 568–598.
- ALLEN, R., A. D. ATHANASSOPOULOS, R.G. DYSON and E. THANASSOULIS (1997), Weight restrictions and value judgements in DEA: evolution, development and future directions, *Annals of Operations Research* **73**, 13–34.
- ANTONOVITZ, F. and T. ROE (1986), A theoretical and empirical approach to the value of information in risky markets, *Review of Economics and Statistics* **68**, 105–114.
- APPELBAUM, E. (1982), The estimation of the degree of oligopoly power, *Journal of Econometrics* **19**, 287–299.
- APPELBAUM, E. (1991), Uncertainty and the measurement of productivity, *Journal of Productivity Analysis* **2**, 157–170.

- APPELBAUM, E. and E. KATZ (1986), Measures of risk aversion and the comparative statics of industry equilibrium, *American Economic Review* **76**, 524–529.
- APPELBAUM, E. and U. KOHLI (1997), Import price uncertainty and the distribution of income, *Review of Economics and Statistics* **79**, 620–630.
- APPELBAUM, E. and A. ULLAH (1997), Estimation of moments and production decisions under uncertainty, *Review of Economics and Statistics* **79**, 631–637.
- BANKER, R. D. (1993), Maximum likelihood, consistency and data envelopment analysis: a statistical foundation, *Management Science* **39**, 1265–1273.
- BANKER, R. D., A. CHARNES and W. W. COOPER (1984), Some models for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science* **30**, 1078–1092.
- BANKER, R. D. and A. MAINDIRATTA (1988), Nonparametric analysis of technical and allocative efficiencies in production, *Econometrica* **56**, 1315–1332.
- BARTELSMAN, E. J. and M. DOMS (2000), Understanding productivity: lessons from longitudinal microdata, *Journal of Economic Literature* **38**, 569–594.
- BATRA, R., and A. ULLAH (1974), Competitive firm and the theory of demand under price uncertainty, *Journal of Political Economy* **82**, 537–548.
- BERGER, A. N. and D. B. HUMPHREY (1997), Efficiency of financial institutions: International survey and directions for future research, *European Journal of Operational Research* **98**, 175–212.
- BOGETOFT, P. (1996), DEA on relaxed convexity assumptions, *Management Science* **42**, 457–465.
- BOGETOFT, P., J. M. TAMA and J. TIND (2000), Convex input and output projections of nonconvex production possibility sets, *Management Science* **46**, 858–869.
- BROCKETT, P. L., W. W. COOPER, H. C. SHIN and Y. WANG (1998), Inefficiency and congestion in Chinese production before and after the 1978 economic reforms, *Socio-Economic Planning Sciences* **32**, 1–20.
- CHAMBERS, R. G., Y. CHUNG and R. FÄRE (1996), Benefit and distance functions, *Journal of Economic Theory* **70**, 407–418.
- CHAMBERS, R. G., Y. CHUNG and R. FÄRE (1998), Profit, directional distance functions, and Nerlovian efficiency, *Journal of Optimization Theory and Applications* **98**, 351–364.
- CHAMBERS, R. G. and J. QUIGGIN (1998), Cost functions and duality for stochastic technologies, *American Journal of Agricultural Economics* **80**, 288–295.
- CHAMBERS, R. G. and QUIGGIN, J. (2000), *Uncertainty, production, choice and agency: the state-contingent approach*, Cambridge University Press, New York.
- CHANG, K.-P. (1999), Measuring efficiency with quasiconcave production frontiers, *European Journal of Operational Research* **115**, 497–506.
- CHARNES, A., W. W. COOPER, B. GOLANY, L. SEIFORD and J. STUTZ (1985), Foundations of data envelopment analysis for Pareto–Koopmans efficient empirical production functions, *Journal of Econometrics* **30**, 91–107.
- CHARNES, A., W. W. COOPER and E. RHODES (1978), Measuring the efficiency of decision making units, *European Journal of Operational Research* **2**, 429–444.
- CHARNES, A., W. W. COOPER, Q. L. WEI and Z. M. HUANG (1989), Cone ratio data envelopment analysis and multi-objective planning, *International Journal of Systems Science* **22**, 2057–2077.
- CHARNES, A., S. HAAG, P. JASKA and J. SEMPLE (1992), Sensitivity of efficiency classifications in data envelopment analysis, *Journal of Systems Sciences* **23**, 789–798.
- CHAVAS, J.-P. (1985), On the theory of the competitive firm under uncertainty when initial wealth is random, *Southern Economic Journal* **51**, 818–827.
- CHERCHYE, L., T. KUOSMANEN and G. T. POST (2000), What is the economic Meaning of FDH? A reply to Thrall, *Journal of Productivity Analysis* **13**, 259–263.
- CHERCHYE, L., T. KUOSMANEN and G. T. POST (2001a), FDH directional distance functions: with an application to European commercial banks, *Journal of Productivity Analysis* **15**, 201–215.

- CHERCHYE, L., T. KUOSMANEN and G. T. POST (2001b), Alternative treatments of congestion in DEA, *European Journal of Operational Research* **133**, 69–74.
- CHERCHYE, L., T. KUOSMANEN and G. T. POST (2002), Nonparametric production analysis in non-competitive environments, *International Journal of Production Economics* **80**, 279–294.
- COOPER, W. W., Z. M. HUANG and S. LI (1996), Satisficing DEA models under chance constraints, *Annals of Operations Research* **66**, 279–295.
- COOPER, W. W., Z. M. HUANG, V. LELAS, S. XI and O. B. OLESEN (1998), Chance constraint programming formulations for stochastic formulations of efficiency and dominance in DEA, *Journal of Productivity Analysis* **9**, 53–79.
- COOPER, W. W., L. M. SEIFORD and K. TONE (2000), *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*, Kluwer Academic Publishers, Dordrecht.
- COOPER, W. W., S. LI, L. M. SEIFORD, K. TONE, R. M. THRALL and J. ZHU (2001), Sensitivity and stability analysis in DEA/some recent developments, *Journal of Productivity Analysis* **15**, 217–246.
- DALAL, A. J. (1990), Symmetry restrictions in the analysis of the competitive firm under price uncertainty, *International Economic Review* **31**, 207–211.
- DEBREU, G. (1951), The coefficient of resource utilization, *Econometrica* **19**, 273–292.
- DEKKER, D. J. and G. T. POST (2001), A quasi-concave DEA model with an application for bank branch performance evaluation, *European Journal of Operational Research* **132**, 296–311.
- DEPRINS, D., L. SIMAR and H. TULKENS (1984), Measuring labor efficiency in post offices, in: M. MARCHAND, P. PESTIEAU and H. TULKENS (eds.), *The performance of public enterprises: concepts and measurement*, North-Holland, Amsterdam.
- DIEWERT, W. E. and C. PARKAN (1983), Linear programming tests of regularity conditions for production frontiers, in: W. EICHHORN, R. HENN, K. NEUMANN and R. W. SHEPHARD (eds.), *Quantitative studies on production and prices*, Physica-Verlag, Würzburg.
- EFRON, B. (1979), Bootstrap methods: another look at the jackknife, *Annals of Statistics* **7**, 1–26.
- EFRON, B. and G. GONG (1983), A leisurely look at the bootstrap, the jackknife, and cross validation, *American Statistician* **37**, 36–48.
- FÄRE, R. and S. GROSSKOPF (1983), Measuring congestion in production, *Zeitschrift für Nationalökonomie* **43**, 257–271.
- FÄRE, R., and S. GROSSKOPF (1996), *Intertemporal production frontiers: with dynamic DEA*, Kluwer, Dordrecht.
- FÄRE, R., S. GROSSKOPF and J. LOGAN (1983a), The relative efficiency of Illinois electric utilities, *Resources and Energy* **5**, 349–367.
- FÄRE, R., S. GROSSKOPF and C. A. K. LOVELL (1983b), The structure of technical efficiency, *Scandinavian Journal of Economics* **72**, 41–62.
- FÄRE, R., S. GROSSKOPF and C. A. K. LOVELL (1985), *The measurement of efficiency of production*, Kluwer Academic Publishers, Boston.
- FÄRE, R., S. GROSSKOPF and C. A. K. LOVELL (1994), *Production frontiers*, Cambridge University Press, Cambridge.
- FÄRE, R. and C. A. K. LOVELL (1978), Measuring the technical efficiency of production, *Journal of Economic Theory* **19**, 150–62.
- FÄRE, R., S. GROSSKOPF, M. NORRIS and Z. ZHANG (1994), Productivity growth, technical progress and efficiency change in industrialized countries, *American Economic Review* **84**, 66–83.
- FÄRE, R., S. GROSSKOPF and P. ROOS (1998), Malmquist productivity indices: a survey of theory and practice, in: R. FÄRE, S. GROSSKOPF and R. RUSSELL (eds.), *Index numbers: essays in honour of Sten Malmquist*, Kluwer Academic Publishers, Boston.

- FÄRE, R. and L. SVENSSON (1980), Congestion of production factors, *Econometrica* **48**, 1745–1753.
- FARRELL, M. J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society Series A* **120**, 253–281.
- FARRELL, M. J. (1959), Convexity assumption in theory of competitive markets, *Journal of Political Economy* **67**, 377–391.
- GIJBELS, I., E. MAMMEN, B. U. PARK and L. SIMAR (1999), On estimation of monotone and concave frontier functions, *Journal of the American Statistical Association* **94**, 220–228.
- GROSSKOPF, S. (1986), The role of the reference technology in measuring productive efficiency, *Economic Journal* **96**, 499–513.
- HANOCH, G. and M. ROTHSCILD (1972), Testing assumptions of production theory: a nonparametric approach, *Journal of Political Economy* **80**, 256–275.
- HEY, J. D. (1979), *Uncertainty in microeconomics*, New York University Press, New York.
- JUST, R. (1974), An investigation of the importance of risk in farmers' decisions, *American Journal of Agricultural Economics* **56**, 14–25.
- KNEIP, A. S. and L. S. SIMAR (1996), A general framework for frontier estimation with panel data, *Journal of Productivity Analysis* **7**, 182–212.
- KNEIP, A. S., B. U. PARK and L. SIMAR (1998), A note on the convergence of nonparametric DEA estimators for production efficiency scores, *Econometric Theory* **14**, 783–793.
- KOOPMANS, T. C. (1951), Analysis of production as an efficient combination of activities, in: T.C. KOOPMANS (ed.), *Activity analysis of production and allocation: proceedings of a conference*, Yale University Press, New Haven.
- KUMBHAKAR, S. C. and C. A. K. LOVELL (2000), *Stochastic frontier analysis*, Cambridge University Press, Cambridge.
- KUOSMANEN, T. (2001), DEA with efficiency classification preserving conditional convexity, *European Journal of Operational Research* **132**, 326–342.
- KUOSMANEN, T. and G. T. POST (2001), Measuring economic efficiency with incomplete price information: with an application to European commercial banks, *European Journal of Operational Research* **134**, 43–58.
- KUOSMANEN, T. and G. T. POST (2002), Nonparametric efficiency analysis under uncertainty: a first-order stochastic dominance approach, *Journal of Productivity Analysis* **17**, 183–200.
- LAND, K., C. A. K. LOVELL and S. THORE (1994), Chance-constrained data envelopment analysis, *Managerial and Decision Economics* **14**, 541–554.
- LI, S. X. (1998), Stochastic models and variable returns to scale in data envelopment analysis, *European Journal of Operational Research* **104**, 532–548.
- MCCALL, J. J. (1969), Competitive production for constant risk utility functions, *Review of Economic Studies* **34**, 417–420.
- MOSCHINI, G.-C. (2001), Production risk and the estimation of ex-ante cost functions, *Journal of Econometrics* **100**, 357–380.
- NEGISHI, T. (1961), Monopolistic competition and general equilibrium, *Review of Economic Studies* **28**, 196–201.
- OLESEN, O. B. and N. C. PETERSEN (1995), Chance constrained efficiency evaluation, *Management Science* **41**, 442–457.
- PEDRAJA-CHAPARRO, F., J. SALINAS-JIMENEZ and P. SMITH (1997), On the role of weight restrictions in data envelopment analysis, *Journal of Productivity Analysis* **8**, 215–230.
- PETERSEN, N. C. (1990), Data envelopment analysis on a relaxed set of assumptions, *Management Science* **36**, 305–314.
- POPE, R. D. and R. E. JUST (1996), Empirical implementation of ex-ante cost functions, *Journal of Econometrics* **72**, 231–249.
- POPE, R. D. and R. E. JUST (1998), Cost function estimation under risk aversion, *American Journal of Agricultural Economics* **80**, 296–302.
- POST, G. T. (2001a), Performance benchmarking in stochastic environments using mean-variance data envelopment analysis, *Operations Research* **49**, 281–292.

- POST, G. T. (2001b), Estimating non-convex production sets using transconcave DEA, *European Journal of Operational Research* **131**, 132–142.
- POST, G. T. (2001c), Transconcave data envelopment analysis, *European Journal of Operational Research* **132**, 374–389.
- POST, G. T., L. CHERCHYE and T. KUOSMANEN (2002), Non-parametric efficiency estimation in stochastic environments, *Operations Research* **50**, 645–655.
- POST, G. T. and J. SPRONK (2000), Evaluating productive performance under uncertainty: combining data envelopment analysis, mean-variance analysis, and multi-factor risk models, in: Y. SHI and M. ZELENY (eds.): *New frontiers of decision making for the information technology era*, World Scientific Publishing Co., Singapore, 249–269.
- RUGGIERO, J. (1998), Non-discretionary inputs in data envelopment analysis, *European Journal of Operational Research* **111**, 461–469.
- RUSSELL, R. (1985), Measures of technical efficiency, *Journal of Economic Theory* **35**, 109–26.
- SANDMO, A. (1971), On the theory of the competitive firm under price uncertainty, *American Economic Review* **61**, 65–73.
- SEIFORD, L. M. and R. M. THRALL (1990), Recent developments in DEA: the mathematical programming approach to frontier analysis, *Journal of Econometrics* **46**, 7–38.
- SENGUPTA, J. K. (1995), *Dynamics of data envelopment analysis, theory of systems efficiency*, Kluwer, Dordrecht.
- SIMAR, L. and P. WILSON (1998), Sensitivity analysis of efficiency scores: how to bootstrap in non-parametric frontier models, *Management Science* **44**, 49–61.
- SIMAR, L. and P. WILSON (2002), Nonparametric tests of returns to scale, *European Journal of Operational Research* **139**, 115–132.
- THANASSOULIS, E. (2001), *Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software*, Kluwer, Dordrecht.
- THRALL, R. M. (1999), What is the economic meaning of FDH?, *Journal of Productivity Analysis* **11**, 243–250.
- TONE, K. (2001), A slacks-based measure of efficiency in DEA, *European Journal of Operational Research* **130**, 498–509.
- TULKENS, H. (1993), On FDH analysis: some methodological issues and applications to retail banking, courts and urban transit, *Journal of Productivity Analysis* **4**, 183–210.
- TULKENS, H. and P. VAN DEN ECKAUT (1999), Mesurer l'efficacité: avec ou sans frontières?, in: P.-Y. BADILLO and J. C. PARADI (eds.), *La méthode DEA: analyse des performances*, Hermes Science, Paris, 75–100.
- WILSON, P. W. (1995), Detecting influential observations in data envelopment analysis, *Journal of Productivity Analysis* **6**, 27–46.
- VARIAN, H. R. (1984), The non-parametric approach to production analysis, *Econometrica* **52**, 279–297.
- VARIAN, H. R. (1985), Non-parametric tests of optimizing behavior with measurement error, *Journal of Econometrics* **30**, 445–458.

Received: September 2001. Revised: November 2002.

