QUADRATIC DATA ENVELOPMENT ANALYSIS

TIMO KUOSMANEN
Helsinki School of Economics
kuosmane@hkkk.fi

THIERRY POST
Erasmus University Rotterdam
gtpost@few.eur.nl

ABSTRACT

Data Envelopment Analysis (DEA) provides means for piecewise linear approximation of production functions. To improve the fit to the data we propose to improve the flexibility of the frontier by extending DEA towards a more general piecewise quadratic approximation, called Quadratic Data Envelopment Analysis (QDEA). In contrast to the linear approximation, the quadratic approximation allows for economies of scale and of specialisation that violate the convexity postulate of DEA. Specifically, QDEA gives statistically consistent estimates for all production functions with bounded Hessian eigenvalues. In addition, Monte-Carlo simulations suggest that QDEA can substantially improve efficiency estimation in finite samples relative to standard DEA models.

Key Words: Data Envelopment Analysis (DEA), Efficiency Measurement, Piecewise Quadratic Approximation.

INTRODUCTION

Data Envelopment Analysis (DEA) [1] is an efficiency measurement and performance benchmarking technique that has been recognised as a valuable analytical research instrument and a practical decision support tool. One of its main advantages is that it does not require a complete specification of the functional form of production relationships, managerial preferences, and error distributions. Still, most DEA models maintain the assumption of convex production possibility set. This convexity postulate may be violated in many relevant research environments, e.g. because of indivisible inputs or outputs, or economies of scale or of specialisation [2]. In general, we have no theoretical reason to believe production possibilities in general would be convex, and various empirical studies suggest violations of convexity for real-life industries (see e.g. [3]).

The restrictive character of the convexity assumption has recently attracted increasing attention. To remedy the convexity problem, various models that allow for both convex and non-convex technologies have been proposed, most notably the Free Disposable Hull (FDH) [3,4]. Unfortunately, FDH efficiency estimates typically suffer from considerable finite sample error. Moreover, the FDH frontier is extremely 'non-smooth', and hence does not provide e.g. information on shadow prices of inputs and outputs. These features may explain why FDH applications still are quite marginal in number compared to applications of convex DEA models today (see e.g. [5] for a survey). A number of alternative approaches to compromise between DEA and FDH models have been recently presented e.g. in [6-10]. Nevertheless, we think the debate on 'deconvexifying' DEA is still open. For example, the motivation of the more relaxed assumptions imposed in these new approaches calls for further investigation (see [10]). Moreover, the computational burden associated with these approaches can turn out a considerable barrier for application.

The convexity problem essentially arises from the piecewise linear approximation employed in the standard DEA. In this paper, we propose to extend the piecewise linear approximation towards a more general piecewise quadratic approximation, called Quadratic Data Envelopment Analysis (QDEA). Focusing first on the single output technology and the Farrell output measure, we show that introducing a single second order term, the squared vector norm of the input vector, can capture non-convexities for all production functions with bounded Hessian eigenvalues. We then extend towards general multiple output technologies and alternative efficiency measures.

STANDARD DEA

The DEA methodology comprises a wide variety of mathematical programming models for performance measurement and performance benchmarking. In this paper, we take the output-oriented version of the Variable Returns-to-Scale (VRS) DEA model developed in [11] as our starting point. We focus on the multiple-input and single-output case in this section.

Suppose we observe a sample of n Decision-Making Units (DMUs) producing a single output $y_j \in \Re^1_+$ by consuming multiple inputs $x_j = \left(x_{1j} \cdots x_{mj}\right)^T \in S_X$, where S_X is a convex subset of $\Re^m_+$. These observations are represented by the output vector $y = \left(y_1 \cdots y_n\right)^T$ and the input matrix $X = \left(x_1 \cdots x_n\right)$. The production technology is represented by the frontier production function $f : S_X \to \Re^1_+$ that maps each input vector to the maximum amount of output that can be produced with this technology. The production possibility set is the hypograph of the frontier, i.e. $T = \{(y, x) \mid y \leq f(x)\}$. We assume the production function to be non-decreasing in all inputs, i.e.

(1) $\nabla f(x) \geq 0 \quad \forall x \in S_X$.

There are different ways of measuring efficiency relative to the production possibilities. In this section, we adhere to the commonly used (inverse of the) classic Farrell [12] definition of output efficiency, and we measure the efficiency of DMU '0' (for the single output technology considered here) as:

(2) $\theta_0 = \dfrac{y_0}{f(x_0)}$.

Unfortunately, the true production function f cannot be observed, and hence efficiency measure (2) cannot be directly computed. In DEA, efficiency is estimated by comparing observed performance with an empirical approximation for the true frontier. That empirical approximation is obtained as the tightest piecewise linear hull of the observations. Specifically, for each observation separately, the frontier is approximated locally by an element of the following class of functions:

(3) $K_{DEA} = \left\{ k : S_X \rightarrow \Re^1_+ \middle| k(x) = \alpha + \beta x^T ; e\alpha + \beta X^T \geq y ; \beta \in \Re^m_+ \right\}$.

This is the class of nondecreasing linear functions that envelop all observations. The functions have the linear form $k(x) = \alpha + \beta x$, with $\beta = (\beta_1 \cdots \beta_m) \in \Re^m_+$. In addition, to be consistent with the definition of a frontier production function, the approximating functions are required to envelop all observations, i.e. $(k(x_1) \cdots k(x_n))^T = e\alpha + \beta X \geq y$, where e denotes a nx1 unity vector. Finally, to comply with a non-decreasing frontier, the approximating functions are also required to be non-decreasing in all inputs, i.e. $(\nabla k(x_1)\cdots\nabla k(x_n))^T = \beta \in \Re_+^m$.

In DEA, an empirical estimate for θ is obtained by substituting the production function by the 'most favourable' element of K_{DEA}, i.e.

(4) $$\hat{\theta}_{DEA,0} = \underset{k \in K_{DEA}}{Max} \frac{y_0}{k(x_0)}$$

(5) $$= \underset{\alpha,\beta}{Max} \left\{ \frac{y_0}{\alpha + \beta x_0} \middle| e\alpha + \beta X \geq y; \beta \in \Re_+^m \right\}$$

(6) $$= \left(\underset{\alpha,\beta}{Min} \left\{ \alpha + \beta x_0 \middle| e\alpha + \beta X \leq y / y_0; \beta \in \Re_+^m \right\} \right)^{-1}.$$

The connection to the output oriented DEA VRS multiplier formulation should be evident from (6).

It follows directly from [13] that the DEA efficiency estimator (4) is statistically consistent (i.e., roughly speaking, asymptotically unbiased and with a vanishing variance) for particular production and distribution structures. Specifically, the production function has to be non-decreasing (1). In addition, consistency requires the production function to be concave, or equivalently, the production possibility set enveloped by the frontier to be convex, i.e.:

(7) $$f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \geq f(x_2) \quad \forall x_1, x_2 \in S_X.$$

In addition to these production assumptions, a number of distribution assumptions are required for consistency. Specifically, the inputs and the efficiencies are considered as random variables with independent and identical distributions with probability density functions $p(x)$ and $h(\theta)$. The inputs and efficiencies are assumed mutually independent; i.e. the joint probability density $l(x,\theta)$ is given by $l(x,\theta) = p(x)h(\theta)$. The inputs are assumed to have a strictly positive density for the entire domain S_X, i.e. $p(x) > 0 \quad \forall x \in S_X$. The efficiencies are assumed to have a strictly positive density at unity, but no density outside the interval (0,1], i.e. $h(1) > 0$, $h(\theta) = 0 \quad \forall \theta \notin (0,1]$.

By relying on general production and distribution assumptions only, DEA has a comparative advantage relative to parametric estimation techniques. Unlike DEA, parametric techniques typically require a complete specification of the probability distribution for efficiency and the functional form of the production relationships. However, the parametric techniques generally do not impose the prior assumption that the production technology is convex. In many research situations, neither theoretical nor empirical justification for imposing convexity are available. Falsely maintained convexity can hence result in systematic underestimation of efficiency. Moreover, restricting attention to convex technologies solely directs attention away from potential gains from economies of scale and of specialisation. Nevertheless, in DEA applications, convexity is often assumed to hold without any theoretical justification or empirical testing.

To illustrate the potential effects of wrongly imposing convexity, consider observations $(y_A, x_A) = (38.25,40)$; $(y_B, x_B) = (133.21,80)$; and $(y_B, x_B) = (276.34,120)$ on three DMUs operating with a production technology characterised by the (non-concave) production function $f(x) = 0.05x^{1.8}$. Figure 1 displays these observations, as well as the production function and the corresponding standard DEA approximation. The latter can be represented by the following piecewise linear function:

$$(8) \qquad k(x) = \begin{cases} 0 & x \in [0,40) \\ -80.81 + 2.98x & x \in [40,120] \\ 276.34 & x \in [120,\infty) \end{cases}.$$

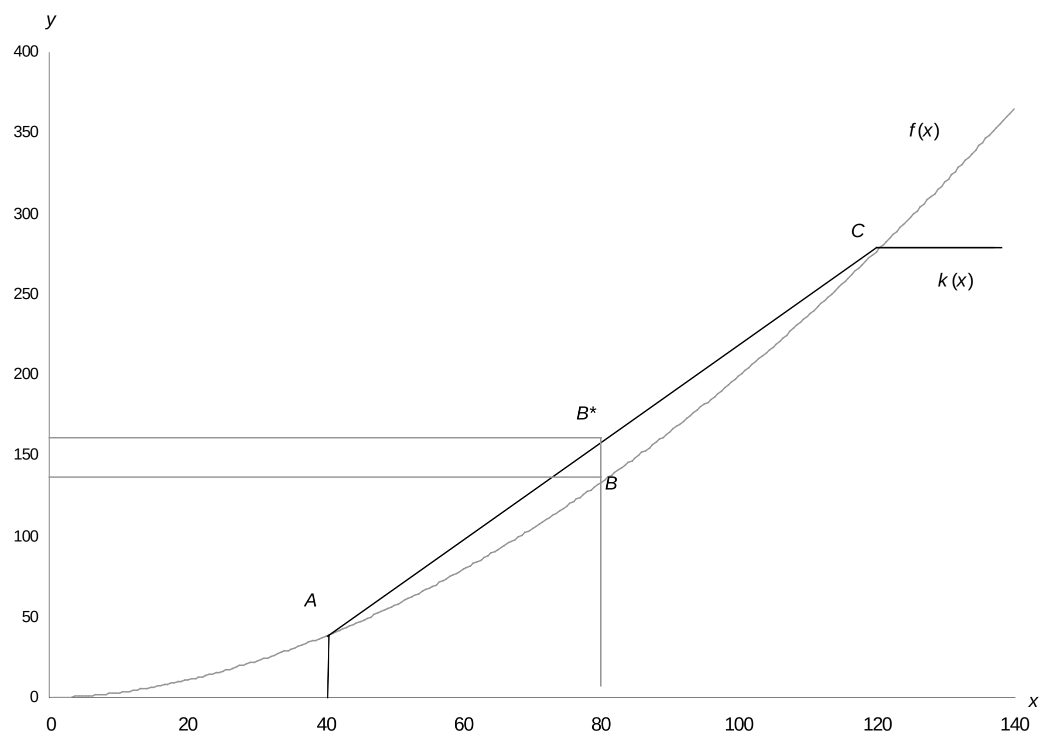

Figure 1: The convexity problem of the standard DEA model

For DMU B, the frontier output estimate of 157.31 is not feasible, because the true frontier output is 133.21, fully 15 percent below the DEA estimate. In terms of the output measure (2), a fully efficient unit receives an efficiency estimate of 0.85. Introducing additional observations can not correct this error, because adding observations corresponds to imposing additional restrictions on a maximisation problem, and consequently lowers the optimal solution (i.e. the efficiency estimate). This demonstrates how the standard DEA approach can overestimate frontier output and underestimate efficiency in cases involving non-convex production technologies.

QUADRATIC DEA

In the parametric tradition, using estimation functions with flexible functional forms can often circumvent the problem of completely specifying the functional form of the production relationships (e.g. [14]). A common approach is to introduce higher-order terms in the estimation function. For example, the well-known translog function includes second-order terms, and can therefore give a second-order Taylor series approximation to an arbitrary twice differentiable function.

In this paper, we intend to use a similar approach in DEA, by relaxing the linearity requirement for the approximating function. A potential limitation of introducing higher-order terms is the associated finite sample error. Higher-order terms effectively relax the restrictions imposed on the maximization problem, and therefore generally increase the optimal solution (i.e. the efficiency estimate). However, we will demonstrate that introducing a single second-order term suffices to account for a wide range of non-convexities without ruining the discriminatory power of the model. That term is the squared input norm, $x_0^T x_0 = \sum_{i=1}^{m} x_{i0}^2$.

More specifically, we propose to replace the class of local approximating functions K_{DEA} with the following class:

(9) $K_{QDEA} = \{k : S_X \rightarrow \Re^1_+ \mid k(x) = \alpha + \beta x + \gamma x^T x; e\alpha + \beta X + \gamma diag(X^T X) \geq y;$
$e\beta + 2\gamma X \geq 0\}$.

We refer to this approach as Quadratic Data Envelopment Analysis (QDEA). The approximating functions are required to be quadratic, i.e. $k(x) = \alpha + \beta x + \gamma x^T x$. Note that the function includes just a single additional variable relative to the approximating functions used in the standard DEA, i.e. the squared input norm. Apart from generalising the approximating function, we follow the standard DEA methodology by requiring the approximating function to envelop all observations, i.e. $(k(x_1)\cdots k(x_n))^T$ $= e\alpha + \beta X + \gamma diag(X^T X) \geq y$, where $diag(X^TX)$ denotes the squared input norm vector $(x_1^T x_1, ..., x_n^T x_n)^T$. In addition, as in the standard model, the function is required to be nondecreasing in all observations, i.e. $(\nabla k(x_1)\cdots \nabla k(x_n))^T = e\beta + 2\gamma X \geq 0$.

Substituting K_{DEA} by K_{QDEA} in efficiency estimator (4) gives the following QDEA estimator:

(10) $$\hat{\theta}_{QDEA,0} = \underset{k \in K_{QDEA}}{Max} \frac{y_0}{k(x_0)}$$

(11) $$= \underset{\alpha,\beta,\gamma}{Max} \left\{ \frac{y_0}{\alpha + \beta x_0 + \gamma x_0^T x_0} \middle| e\alpha + \beta X + \gamma diag(X^T X) \geq y; e\beta + 2\gamma X \geq 0 \right\}$$

(12) $$= \left(\underset{\alpha,\beta,\gamma}{Min} \left\{ \alpha + \beta x_0 + \gamma x_0^T x_0 \middle| e\alpha + \beta X + \gamma diag(X^T X) \leq y / y_0; e\beta + 2\gamma X \geq 0 \right\} \right)^{-1}.$$

Note that the QDEA efficiency estimator introduces just a single additional variable to the standard DEA VRS model, i.e. the squared input norm. Although the approximating function k is nonlinear, this QDEA estimator remains linear in all its parameters α, β, γ. Therefore, standard Linear Programming techniques can be employed for computing the efficiency scores.

Interestingly, this single additional variable can in many cases remedy the convexity problem associated with the standard DEA. In fact, the QDEA model can be demonstrated to give statistically consistent estimates for all non-decreasing frontiers *with bounded Hessian eigenvalues*, provided the standard distribution assumptions (see the previous section) are satisfied.

PROOF As demonstrated in Appendix A, boundedness of the Hessian implies the following inequality, using ξ to denote the maximum Hessian eigenvalue on S_X :

(P.1) $\quad f(x_1) + \nabla f(x_1)^T (x_2 - x_1) + \frac{1}{2}\xi (x_2 - x_1)^T (x_2 - x_1) \geq f(x_2) \quad \forall x_1, x_2 \in S_X$.

Therefore, the line $p(x) = (f(x_0) - \nabla f(x_0)^T x_0 + \frac{1}{2}\xi x_0^T x_0) + (\nabla f(x_0) - \xi x_0)^T x + \frac{1}{2}\xi x^T x$ is tangent to the frontier at x_0 and hence completely envelops the frontier. Moreover, its first-order derivative $p'(x) = \nabla f(x_0)$ is nonnegative, because the frontier is non-decreasing (1). Therefore,

(P.2) $\exists \alpha, \beta, \gamma : \alpha = f(x_0); \beta x_0 + \gamma x_0^T x_0 = 1; e\alpha + \beta X + \gamma diag(X^T X) \geq \left(f(x_1) \cdots f(x_n)\right)^T; e\beta + 2\gamma X \geq 0$.

Consequently, if boundedness and monotonicity hold, efficiency cannot be underestimated:

(P.3) $\quad \hat{\theta}_{QDEA,0} \geq \theta_0$.

This finding is related to a well-known result in the mathematical theory of generalized concavity. More specifically, the hypograph of any function with a bounded from above

Hessian is supported, at each point, by a quadratic polynomial. For a more detailed discussion, see [15].

In addition, the standard distribution assumptions (see the previous section) imply that the probability that a random DMU operates approximately at the frontier reference point for the evaluated unit is strictly positive, i.e., letting $\delta_1 > 0$ and $\delta_2 > 0$ denote non-Archimedean infinitesimal small values:

(P.4) $P\left[x_o - \delta_1 \leq x_j \leq x_0 + \delta_1, 1 - \delta_2 \leq \theta_j \leq 1\right] = \int_{x_o-\delta_1}^{x_0+\delta_1}\int_{1-\delta_2}^{1} p(x)h(\theta)\partial x \partial\theta > 0 \;\; \forall x_o \in S_X ,\, j = 1,\ldots,n$.

Therefore, the probability of observing at least one such DMU approximates unity as the number of observations increases, i.e.:

(P.5) $$\lim_{n\to\infty}\left[1 - \prod_{j=1}^{n} P\left[x_o - \delta_1 \leq x_j \leq x_0 + \delta_1, 1 - \delta_2 \leq \theta_j \leq 1\right]\right] = 1.$$

If such a DMU is introduced, the efficiency estimate is necessarily equal or smaller than *true* inefficiency, i.e.: $\hat{\theta}_{QDEA,0} \leq \theta_0$. Therefore, asymptotically efficiency can not be overestimated:

(P.6) $$\lim_{n\to\infty}\left\lfloor \hat{\theta}_{QDEA,0} \right\rfloor \leq \theta_0.$$

Combining (P.3) and (P.6), we find consistency:

(P.7) $$\lim_{n\to\infty}\left\lfloor \hat{\theta}_{QDEA,0} \right\rfloor = \theta_0.$$ **EOP**

Boundedness of the Hessian eigenvalues is a much less restrictive assumption than convexity, because the Hessian eigenvalues of concave functions are bounded from above by zero, as negative semidefinite matrices are characterised by non-positive eigenvalues. Therefore, QDEA can deal with a broader set of production technologies than the standard DEA. Most notably, QDEA allows for interesting properties such as economies of scale and of specialisation. For example, Appendix B demonstrates that, for a Cobb-Douglas production function, the boundedness condition allows the sum of scale parameters to vary between zero and 2. By contrast, convexity restricts the sum of parameters to be between zero and unity. QDEA also gives consistent estimates if the production function is concave (and hence standard DEA is correctly specified). However, for concave functions, QDEA efficiency estimator involves higher finite sample error than the standard DEA estimator. Similarly, if QDEA is correctly specified (i.e. boundedness holds), the efficiency estimator involves smaller finite sample error than FDH. However, FDH gives consistent estimates for a wider range of production functions, including those with unbounded Hessian eigenvalues.

To illustrate the operation of the QDEA model, reconsider the example introduced in the previous section. The second-order derivative of the production function corresponds to $f''(x)=0.072x^{-0.2}$, and consequently the Hessian eigenvalue on the interval [40,120] is bounded by $\xi=f''(40)=0.0344$. Therefore, the QDEA model cannot underestimate efficiency for observation B. Figure 2 displays the true production function and the QDEA approximation $k(x)$. The latter can be represented by the following piecewise quadratic function:

$$(13) \qquad k(x)=\begin{cases} 0 & x\in[0,40) \\ 133.21\text{-}4.748x+0.059x^2 & x\in[40,80) \\ 133.21\text{-}2.386x+0.030x^2 & x\in[80,120) \\ 276.34 & x\in[120,\infty) \end{cases}.$$

For observation B, the frontier output estimate coincides with the true frontier output of 133.21.

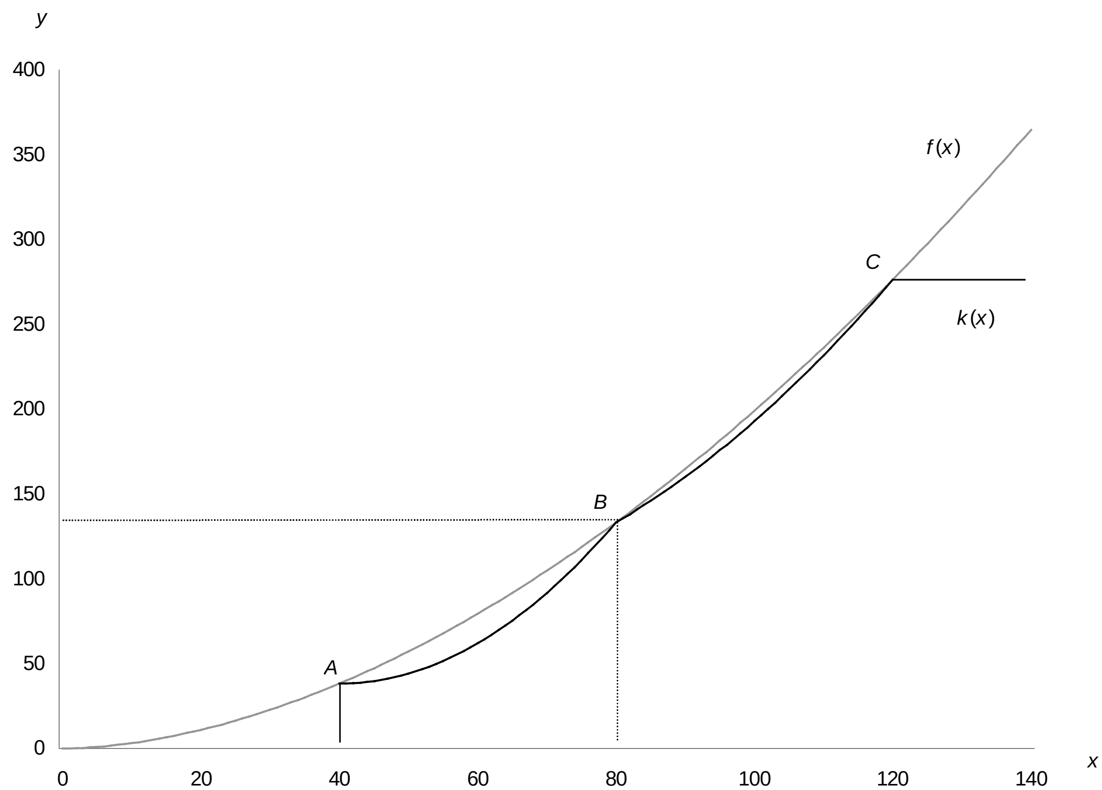

Figure 2: The QDEA model

MULTI-OUTPUT TECHNOLOGIES AND ALTERNATIVE EFFICIENCY MEASURES

Thus far, we have focused on the single output technology and Farrell's output measure. One attractive feature of standard DEA is the easy treatise of multi-product technologies. Therefore, in this section we extend the QDEA model proposed in the previous section to more general multiple output technologies. We redefine $y_j = (y_{1j} \cdots y_{qj})^T \in S_Y$, $j = 1,\ldots,n$, as an output *vector*, where S_Y is a convex subset of $\Re_+^q$. The output data are represented by the matrix $Y = \left(y_1 \cdots y_n\right)$. The multi-output production technology is represented by the implicit production function $h(y, x) = 0$, $\nabla h_x \leq 0, \nabla h_y \geq 0$.

Let us first generalise the Farrell output measure to the multi-output setting. When it seems reasonable to assume the potential violations of convexity do not carry over to the output space, we can still try to capture the non-convexities by the quadratic input norm. In the multi-output setting, this approach boils down to the following set of local approximation functions:

(14) $$K'_{QDEA} = \left\{ k' : S_y \times S_x \to \Re^1 \middle| k'(y,x) = \alpha + \beta x + \delta y + \gamma x^T x \right.$$

$$\left. \alpha + \beta X + \delta Y + \gamma diag X^T X \leq 0; \delta \geq 0; e\beta + 2\gamma \leq 0 \right\}.$$

This class involves quadratic functions of the form $k'(y,x) = \alpha + \beta x + \delta y + \gamma x^T x$. As in the single output case, the approximating function is required to envelop all observations, i.e. $k'(Y,X) = \alpha + \beta X + \delta Y + \gamma diag X^T X \leq 0$. In addition, the function is required to be nondecreasing for output and nonincreasing for input for all observations, i.e. $\delta \geq 0$ and $\left(\nabla k'_x(y_1,x_1) \cdots \nabla k'_x(y_n,x_n)\right)^T = e\beta + 2\gamma \leq 0$.

The Farrell measure (10) generalises to the multiple output setting directly as:

(15) $$\hat{\theta}_{QDEA,0} = \underset{\theta, k' \in K'_{QDEA}}{Max} \left\{ \theta \middle| k'(\theta y_0, x_0) \leq 0 \right\}$$

(16) $$= \left(\underset{\alpha,\beta,\gamma,\delta}{Min} \left\{ \alpha + \beta x_0 + \gamma x_0^T x_0 \,\middle|\, \delta y_0 = 1; e\alpha + \beta X + \delta y + \gamma diag(X^T X) \leq 0; e\beta + 2\gamma X \leq 0; \delta \geq 0 \right\} \right)^{-1}.$$

Note that (16) is linear in its parameters, and hence can be computed by standard Linear Programming techniques.

In general, however, we might find the previous assumption of convex output correspondences overly restrictive. In the general case, we can better capture the non-convexities e.g. by using the squared vector norm of the entire production vector, i.e. $(x\ y)^T(x\ y)$. This gives the following set of local approximation functions:

$$(17)\qquad K''_{QDEA} = \{k'': S_y \times S_x \to \Re^1 \mid k''(y,x) = \alpha + \beta x + \delta y + \gamma(x\ y)^T(x\ y)$$

$$\alpha + \beta X + \delta Y + \gamma diag(X\ Y)^T(X\ Y) \le 0; e\delta + 2\gamma X \ge 0; e\beta + 2\gamma Y \le 0\}.$$

This class involves quadratic functions of the form $k(y,x) = \alpha + \beta x + \delta y + \gamma(x\ y)^T(x\ y)$. As in the single output case, the approximating function is required to envelop all observations, i.e. $k''(Y,X) = \alpha + \beta X + \delta Y + \gamma diag(X\ Y)^T(X\ Y) \le 0$, where $diag(X\ Y)^T(X\ Y)$ denotes the squared vector norm vector of the production data, i.e. $((x_1\ y_1)^T(x_1\ y_1),...,(x_n\ y_n)^T(x_n\ y_n))^T$. In addition, the function is required to be nondecreasing for output and nonincreasing for input for all observations, i.e. $\left(\nabla k_y''(y_1,x_1)\cdots\nabla k_y''(y_n,x_n)\right)^T = e\delta + 2\gamma X \ge 0$ and $\left(\nabla k_x''(y_1,x_1)\cdots\nabla k_x''(y_n,x_n)\right)^T = e\beta + 2\gamma Y \le 0$.

We might also want to consider alternative orientations of measurement. We next resort to the most general efficiency measure, the *directional distance function* proposed in [16,17], inspired by Luenberger's [18] benefit function. This measure can be defined as:

$$(18)\qquad \mu_0 = \underset{\theta}{Max}\{\theta \mid h(y_0 + \theta g_y, x_0 - \theta g_x) \le 0\},$$

where (g_y, g_x) is the direction vector. Directional distance function contains several interesting special cases. For instance, $(g_y, g_x) = (0, x_0)$ gives the common Farrell input measure as $1-\mu_0$, while $(g_y, g_x) = (y_0, 0)$ yields the corresponding output measure as $(1+\mu_0)$.

Like above, an empirical efficiency estimator can be obtained by substituting the true production correspondence in the distance measure (18) by the 'most favourable' element of K_{QDEA}. This gives the empirical directional distance function:

$$(19) \qquad \hat{\mu}_{QDEA,0} = \underset{\theta, k'' \in K_{QDEA}^{''}}{Max} \left\{ \theta \middle| k''(y_0 + \theta g_y, x_0 - \theta g_x) = 0 \right\}$$

$$(20) \qquad = \underset{\theta,\alpha,\beta,\delta,\gamma}{\mathrm{Max}} \Big\{ \theta \Big| \alpha + \beta(x_0 - \theta g_x) + \delta(y_0 + \theta g_y) + \gamma \big[(x_0 - \theta g_x)\,(y_0 + \theta g_y) \big]^T \big[(x_0 - \theta g_x)\,(y_0 + \theta g_y) \big] = 0; \\ \alpha + \beta X + \delta Y + \gamma diag(X\ Y)^T (X\ Y) \leq 0; e\delta + 2\gamma X \geq 0; e\beta + 2\gamma Y \leq 0 \Big\}.$$

Unfortunately, the general QDEA directional distance function is non-linear in its parameters, and in general cannot be computed by Linear Programming. This can be computationally demanding. Still, solving a single Non-linear Optimisation problem suffices even in this most general case. Note that alternative approaches for relaxing convexity, such as those presented in [8], [9] and [10], involve recursive or iterative protocols of solving an unknown number of Mathematical Programming problems.

SIMULATING FINITE SAMPLE PERFORMANCE

Above we demonstrated that QDEA gives consistent estimators for a wide range of production functions, including non-concave functions. Still, introducing higher-order terms can increase finite sample error. However, the QDEA model minimises the finite sample error, because the non-convexities are captured by the single squared input norm term. Unfortunately, no analytic solutions are available for the finite sample properties of DEA estimators. Consequently, we are not able to formally derive conditions under which QDEA improves efficiency estimation. To analyse the finite sample performance, we performed a series of Monte-Carlo simulations.

A difficulty of using simulated data is the selection of the appropriate data generating process (henceforth DGP) and the appropriate range for varying the parameters of the DGP. If the simulations do not represent a wide range of real-life research environments, the simulation results can have little significance for research practice. Unfortunately, to the best of our knowledge, there currently is no generally accepted framework for performing simulation studies in DEA. Therefore, we adopt a heuristic approach in this study. Specifically, we use a DPG that is relatively simple, and hence limits the computational burden. Still, we think the simulation setting is realistic enough to give us a clue of the finite sample performance of QDEA.

In the simulations, all units (j=1,…,n) were assumed to face a common Cobb-Douglas production technology, that maps two inputs into a single output. The following frontier production function formally represents that technology:

$$
\begin{aligned}
& f(x_1, x_2) = \gamma_1 x_1^{\gamma_2} x_2^{\gamma_3} \\
(21) \qquad & y_j = f(x_{1j}, x_{2j})\theta_j; \\
& \ln\theta_j \sim \left|N(0,\sigma)\right|.
\end{aligned}
$$

The inputs were drawn from a bivariate lognormal distribution with (for both inputs) mean 2 and standard deviation 0.25 and coefficient of correlation 0.9. The log-normal distribution was selected to prevent negative values and moreover reflect the skewness of the size distribution typically encountered in empirical studies. The correlation was specified to reflect the high degree of correlation between inputs typically encountered in empirical studies.

Output data was computed by multiplying the value of production function with the efficiency percentage θ_j . The log-efficiencies follow a half normal distribution, i.e. they are the absolute values of a variable with mean 0 and standard deviation σ. Under this assumption, the log-efficiency has an average value of $(2/\pi)^{0.5}\sigma$ and a standard deviation of $((\pi\text{-}2)/\pi)^{0.5}\sigma$.

We compared the QDEA efficiency estimates obtained by (8) to the corresponding DEA VRS estimates obtained by (4), as well as to the FDH estimates [4]. The FDH approach compares the evaluated unit with the most productive DMU that dominates it. FDH estimates for Farrell output measure can be computed with a simple enumeration algorithm [3]. In the case of a single output that algorithm reduces to:

$$\text{(22)} \qquad \hat{\theta}_{FDH,0} = \underset{j\in\{1,\dots,n\}}{Min} \left\{ \frac{y_0}{y_j} \middle| x_j \le x_0 \right\}.$$

The FDH model can deal with non-convex technologies, because it relies on dominance relationships alone. Specifically, for all non-decreasing frontiers, the FDH efficiency estimators are statistically consistent, provided the standard distribution assumptions hold. However, they typically involve higher finite sample error than the DEA and QDEA estimators. Whereas DEA and QDEA approaches allow for comparison with virtual input-output combinations, FDH restricts comparison to observed combinations only. Therefore, FDH efficiency estimates are necessarily higher than DEA and QDEA estimates.

The starting point for the simulations was a scenario with 200 observations, a 95 percent average efficiency level and coefficient values $\gamma_1 = 10, \gamma_2 = \gamma_3 = \gamma = 0.7$. In this study, we intend to analyse the influence of the sample size, the efficiency level and non-convexity. Therefore, we systematically varied the values for *n*, σ, and γ. More specifically, we considered samples of 50, 200 and 500 observations, average efficiency levels of 99, 95 and 90 percent, and values of 0.3, 0.7 and 1.2 for γ. For all γ values, the production function is non-decreasing in both inputs. For $\gamma \le 0.5$ the frontier is concave, exhibiting non-increasing returns-to-scale, and, for $\gamma > 0.5$, it is convex, exhibiting increasing returns-to-scale. In addition, for $\gamma \le 1$, the frontier has bounded Hessian eigenvalues, but not for $\gamma > 1$. (as is demonstrated in Appendix B). Therefore, the values $\gamma = 0.3$, $\gamma = 0.7$ and $\gamma = 1.2$ involve different patterns of specification error for the different models (see Table 1).

Table 1: Specification of DEA models at different values of γ

	FDH	QDEA	DEA VRS
$\gamma = 0.3$	Correct	Correct	Correct
$\gamma = 0.7$	Correct	Correct	Erroneous
$\gamma = 1.2$	Correct	Erroneous	Erroneous

Each experiment consisted of generating a set of artificial data from the above data-generating process, and employing each of the three competing efficiency estimators to these data. Total of 100 experiments were conducted in each combinations of sample size, efficiency level and production technology. Next, the obtained efficiency estimates were used to gauge the sampling distribution properties of the competing estimators. We considered three sample statistics: bias (BIAS), standard deviation (STD), and root mean squared error (RMSE). These statistics were computed as follows:

(23) $$BIAS = \tfrac{1}{n} \sum_{j=1}^{n} (\hat{\theta}_j - \theta_j)\,;$$

(24) $$STD = \sqrt{\tfrac{1}{n} \sum_{j=1}^{n} (\hat{\theta}_j - \bar{\hat{\theta}}_j)^2}\,;$$

(25) $$RMSE = \sqrt{\tfrac{1}{n} \sum_{j=1}^{n} (\hat{\theta}_j - \theta_j)^2}\,.$$

Table 2 displays the average values for the sample statistics, computed over the 100 replications. In addition, the table indicates whether the FDH and DEA statistics differ significantly from the QDEA statistics. For this purpose, the sample means for the statistics were compared using a paired *t*-test.

In all scenarios (even those involving large samples and high efficiency levels), the null hypothesis of unbiased efficiency estimates can be rejected in the *t*-test with a significance level of 1 percent. The bias is positive if the model is correctly specified, and negative if the model is erroneously specified. Our findings suggest that errors from erroneous specification outweigh finite sample error in magnitude. This seems especially discomforting to the DEA approach, which imposes relatively strong assumptions on the production technology, while more encouraging to QDEA and FDH, which rely on weaker assumptions.

Table 2: Sample statistics for the alternative models.

			n=50			n=200			n=500		
			FDH	**QDEA**	**DEA**	**FDH**	**QDEA**	**DEA**	**FDH**	**QDEA**	**DEA**
γ=0.30	$E(\theta)$=0.99	**BIAS**	0.010-	0.006	0.006	0.009-	0.004	0.003+	0.008-	0.002	0.002+
		STD	0.007-	0.004	0.004	0.007-	0.003	0.003	0.006-	0.003	0.003
		RMSE	0.012-	0.008	0.007	0.011-	0.005	0.004+	0.009-	0.004	0.003+
	$E(\theta)$=0.95	**BIAS**	0.038-	0.021	0.016+	0.028-	0.011	0.007+	0.023-	0.008	0.004+
		STD	0.028-	0.017	0.014+	0.019-	0.008	0.007+	0.018-	0.008	0.007+
		RMSE	0.047-	0.027	0.022+	0.034-	0.013	0.010+	0.029-	0.011	0.008+
	$E(\theta)$=0.90	**BIAS**	0.071-	0.040	0.030+	0.045-	0.019	0.012+	0.035-	0.015	0.007+
		STD	0.051-	0.039	0.029+	0.032-	0.014	0.013+	0.026-	0.013	0.009+
		RMSE	0.087-	0.056	0.041+	0.056-	0.023	0.017+	0.043-	0.020	0.012+
γ=0.70	$E(\theta)$=0.99	**BIAS**	0.010-	0.003	-0.325-	0.009-	0.001	-0.255-	0.007-	0.002	-0.446-
		STD	0.007	0.006	0.175-	0.007-	0.003	0.119-	0.006-	0.004	0.170-
		RMSE	0.012-	0.007	0.369-	0.011-	0.003	0.281-	0.009-	0.005	0.477-
	$E(\theta)$=0.95	**BIAS**	0.038-	0.007	-0.291-	0.009-	0.001	-0.255-	0.022-	0.012	-0.412-
		STD	0.025	0.025	0.182-	0.007-	0.003	0.119-	0.015-	0.013	0.155-
		RMSE	0.045-	0.026	0.343-	0.011-	0.003	0.281-	0.027-	0.018	0.440-
	$E(\theta)$=0.90	**BIAS**	0.079-	0.005	-0.241-	0.047-	0.003	-0.192-	0.035-	0.020	-0.391-
		STD	0.063-	0.050	0.172-	0.033	0.032	0.111-	0.027-	0.026	0.175-
		RMSE	0.101-	0.050	0.296-	0.058-	0.032	0.221-	0.045-	0.032	0.428-
γ=1.20	$E(\theta)$=0.99	**BIAS**	0.009+	-0.324	-0.657-	0.008+	-0.415	-0.767-	0.008+	-0.439	-0.797-
		STD	0.007+	0.176	0.227-	0.006+	0.178	0.181	0.006+	0.187	0.178
		RMSE	0.011+	0.368	0.695-	0.009+	0.451	0.788-	0.010+	0.477	0.817-
	$E(\theta)$=0.95	**BIAS**	0.038+	-0.327	-0.648-	0.031+	-0.409	-0.746-	0.022+	-0.442	-0.787-
		STD	0.028+	0.176	0.215-	0.021+	0.169	0.176	0.017+	0.188	0.219-
		RMSE	0.047+	0.371	0.682-	0.038+	0.442	0.766-	0.027+	0.480	0.817-
	$E(\theta)$=0.90	**BIAS**	0.077+	-0.286	-0.567-	0.047+	-0.346	-0.657-	0.036+	-0.434	-0.728-
		STD	0.051+	0.192	0.265-	0.034+	0.154	0.214-	0.028+	0.164	0.171
		RMSE	0.092+	0.344	0.625-	0.058+	0.378	0.690-	0.046+	0.464	0.748-

Note: + (-) stands for a statistic that is significantly better (worse) than the associated QDEA statistic at a 1 percent significance level. The shaded area indicates the model with the lowest RMSE.

For the base case (n=200, $E(\theta)$=0.95, and γ=0.7), the FDH errors are centred well above zero and widely spread out. By contrast, the QDEA errors are centred near zero and have relatively little variance. Since FDH compares the evaluated unit with observed input-output combinations only, while QDEA compares with virtual combinations also, the former necessarily has higher estimates than the latter. In addition, the production frontier is non-decreasing and has bounded Hessian eigenvalues. Therefore, both models cannot underestimate efficiency, and consequently FDH involves more finite sample error than QDEA. The DEA model also suffers from high error variance. However, in contrast to the FDH it is biased below true efficiency. This reflects the fact that DEA can confuse inefficiency and non-convexity if the technology is non-convex. The QDEA model can handle the non-convex nature of the technology, because the Cobb-Douglas production function has bounded Hessian eigenvalues in the base case.

Varying sample size, efficiency level and non-convexity leads to some changes in conclusions relative to the base case. If $\gamma = 1.2$, convexity is more severely violated, and hence DEA performance deteriorates further. Also QDEA estimator becomes inconsistent in this case, but the RMSE is only about 60% of that associated with DEA in all alternative cases. In this scenario, FDH turns out superior even in small samples. On the other hand, when $\gamma = 0.3$, the frontier becomes concave. Consequently, DEA cannot underestimate efficiency, and gives consistent estimates. QDEA involves greater finite sample error than DEA due to the additional variable, i.e. quadratic input norm. Still, when we balance the finite sample error against the risk of specification error, QDEA proves competitive, especially in larger samples and at high efficiency levels.

Our findings suggest QDEA could be a valuable research tool in circumstances where convexity might be violated, but not to such an extent that would call for the FDH approach. Nevertheless, it should be understood that the magnitude of error from erroneous specification and finite sample error for the alternative models may differ in alternative research environments. Ultimately, the usefulness of the proposed approach should be proven in the real-life research practice.

CONCLUSIONS

We have demonstrated how resorting to a non-linear approximating function in DEA can account for non-convex production technologies. In contrast to most DEA models, QDEA allows for economies of scale and of specialisation. From statistical point of view, the QDEA efficiency estimators are consistent for a broader set of production technologies than the VRS DEA estimators. It should be stressed that the QDEA approach is also convenient from the computational point of view: The efficiency measures can often be computed by solving a single Linear Programming problem, and even in the worst case, by solving a single Non-linear Optimisation problem. In addition, Monte-Carlo simulations suggest that the QDEA can improve inefficiency estimation in finite samples, both relative to the DEA VRS and the less restrictive FDH specifications.

Above we have focussed on extending the range of model specifications in DEA. However, we have not addressed the important issue of how to select the best specification if the prior information on production possibilities is incomplete. In some cases, the analyst can still have some useful knowledge about the shape of the true frontier, e.g. from economic theory, practical knowledge about the industry under evaluation, or econometric estimation. Alternative possibility is to resort to empirical specification tests. For example, note that the QDEA problem (12) coincides to the DEA VRS problem (6) by imposing the restriction $\gamma = 0$. Statistical specification tests are currently available for testing this type of restrictions, see e.g. [21].

The above discussion focused on a quadratic function that only includes the single quadratic term – the squared input norm, or alternatively the squared norm of the entire production vector. We emphasise that introducing additional variables affects the discriminatory power of the model. Hence, we proposed introduce a single additional variable only, so as to minimise the finite sample error. Of course, if the assumption of bounded eigenvalues is considered too restrictive, we could introduce multiple quadratic terms with separate weighting parameters, or extend further to higher-order polynomial approximations. Future research efforts could be directed to investigating these possibilities, as well as analysing alternative non-linear approximations. Such approximations could provide consistent estimators for even broader classes of production frontiers.

APPENDIX A

In this appendix, we intend to proof inequality (P.1). Consider, for any two fixed input vectors $x_{t}, x_k \in S_X$, the univariate cross section of the production frontier $h(\tau) = f(x_k + \tau(x_t - x_k))$ as a function of $\tau \in R_1$. Applying the Mean Value Theorem [19], we obtain:

(A.1) $\exists \varphi \in [0,\tau] : h'(\tau) - h'(0) = h''(\varphi)\tau \quad \forall \tau \in R_1$.

Integrating this expression over [0, τ] yields:

(A.2) $\exists \varphi \in [0,\tau] : h(\tau) - h(0) - h'(0)\tau = \frac{1}{2} h''(\varphi)\tau^2 \quad \forall \tau \in R_1$.

Setting τ=1, we have:

(A.3) $\exists \varphi \in [0,1] : h(1) - h(0) - h'(0) = \frac{1}{2} h''(\varphi)$.

Using the chain rule for differentiation, we obtain:

(A.4) $\exists\varphi \in [0,1]: f(x_1) - f(x_2) - \nabla f(x_1)(x_2 - x_1)^T = \frac{1}{2}(x_2 - x_1)H(x_1 + \varphi(x_2 - x_1))(x_2 - x_1)^T$.

In addition, using $\chi(\varphi)$ to denote the maximum eigenvalue of Hessian matrix $H(x_k + \varphi(x_t - x_k))$, we have from matrix theory [20, p.112]:

(A.5) $(x_2 - x_1)H(x_1 + \varphi(x_2 - x_1))(x_2 - x_1)^T \le \frac{1}{2}\chi(\varphi)x_1 + \varphi(x_2 - x_1)(x_2 - x_1)^T$.

Combining (A.4) and (A.5), we find:

(A.6) $\exists\varphi \in [0,1]: f(x_1) + \nabla f(x_1)(x_2 - x_1)^T + \frac{1}{2}\chi(\varphi)(x_1 + \varphi(x_2 - x_1))(x_1 + \varphi(x_2 - x_1))^T \ge f(x_2)$

Consequently, letting ξ denote the maximum eigenvalue of Hessian matrix $H(x_k)$ for all $x_k \in S_X$, we obtain:

(A.7) $f(x_1) + \nabla f(x_1)(x_2 - x_1)^T + \frac{1}{2}\xi(x_2 - x_1)(x_2 - x_1)^T \ge f(x_2) \quad \forall x_1, x_2 \in S_X$. **EOP**

APPENDIX B

In this appendix, we intend to demonstrate that the Cobb-Douglas production function

$(C.1) \quad f(x_1, x_2) = \alpha x_1^{\beta_1} x_2^{\beta_2} \quad 0 < \beta_1, \beta_2 < 1;\ 1 < \beta_1 + \beta_2 \le 2$

has Hessian eigenvalues that are bounded from above on $S_X = \{(x_1, x_2) : x_1, x_2 > 0\}$. The proof for $0 \le \beta_1 + \beta_2 \le 1$ is omitted as obvious, as the Hessian eigenvalues of concave functions are bounded from above by zero, as negative semidefinite matrices are characterised by non-positive eigenvalues.

The Hessian matrix of (C.1) is represented by:

$$(C.2) \quad H = \alpha\begin{bmatrix} \beta_1(\beta_1 - 1)x_1^{\beta_1 - 2}x_2^{\beta_2} & \beta_1\beta_2 x_1^{\beta_1 - 1}x_2^{\beta_2 - 1} \\ \beta_1\beta_2 x_1^{\beta_1 - 1}x_2^{\beta_2 - 1} & \beta_2(\beta_2 - 1)x_1^{\beta_1}x_2^{\beta_2 - 2} \end{bmatrix}.$$

The eigenvalues of H can be found by solving the characteristic equation: $\xi^2 - trH\xi + |H| = 0$. Since the determinant and the trace are strictly negative, the solution gives eigenvalues $\xi_1 > 0 > \xi_2$. The upper eigenvalue is represented by:

$$(C.3) \quad \xi_1 = \tfrac{1}{2}(trH + \sqrt{(trH)^2 - 4|H|}) = \tfrac{1}{2}\frac{-4|H|}{\sqrt{(trH)^2 - 4|H|} - trH}$$

$$= \frac{\alpha\beta_1\beta_2(\beta_1 + \beta_2 - 1)x_1^2 x_2^2}{(\sqrt{(\beta_1(1-\beta_1)x_1^2 + \beta_2(1-\beta_2)x_2^2)^2 + 4\beta_1\beta_2(\beta_1 + \beta_2 - 1)x_1^2x_2^2} + \beta_1(1-\beta_1)x_1^2 + \beta_2(1-\beta_2)x_2^2)x_1^{2-\beta 1}x_2^{2-\beta_2}}$$

Writing $x_1 = r\cos\varphi$ and $x_2 = r\sin\varphi$, and omitting φ for shortness, one has, dividing numerator and denominator in the quotient by r^2:

$$(C.4) \quad \xi_1 = \frac{\alpha\beta_1\beta_2(\beta_1+\beta_2-1)\cos^{\beta_1}\sin^{\beta_2}}{\left(\sqrt{(\beta_1(1-\beta_1)\cos^2+\beta_2(1-\beta_2)\sin^2)^2+4\beta_1\beta_2(\beta_1+\beta_2-1)\cos^2\sin^2}+\beta_1(1-\beta_1)\cos^2+\beta_2(1-\beta_2)\sin^2\right)r^{2-\beta_1-\beta_2}}$$

In this expression, the term

$$\sqrt{(\beta_1(1-\beta_1)\cos^2+\beta_2(1-\beta_2)\sin^2)^2+4\beta_1\beta_2(\beta_1+\beta_2-1)\cos^2\sin^2}+\beta_1(1-\beta_1)\cos^2+\beta_2(1-\beta_2)\sin^2$$

is bounded away from zero. Since $\beta_1+\beta_2 \le 2$, the term $1/r^{2-\beta_1-\beta_2}$ is bounded too. Therefore, ξ_1 is bounded on $S_X = \{(x_1,x_2) : x_1, x_2 > 0\}$. **EOP**

REFERENCES

1 Charnes, A., Cooper, W. W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6), 429-444.

2 Farrell, M.J. (1959). Convexity assumption in theory of competitive markets. *Journal of Political Economy* 67, 377-391.

3 Tulkens, H. (1993). On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts, and Urban Transit, *The Journal of Productivity Analysis* 4(1), 183-210.

4 Deprins, D., L. Simar and H. Tulkens (1984). Measuring labour efficiency in post offices. In M. Marchand, P. Pestieu and H. Tulkens (eds.) *The performance of public enterprises*, 243-267. North Holland, Amsterdam.

5 Seiford, L. M. (1996). Data Envelopment Analysis: The Evolution of the State of the Art 1978-1995, *Journal Productivity Analysis* 7, 99-139.

6 Petersen, N. C. (1990). Data Envelopment Analysis on a relaxed set of assumptions, *Management Science*, vol. 36 (3), pp. 305-314.

7 Bogetoft, P. (1996). DEA on Relaxed Convexity Assumptions, *Management Science* 42(3), 457-465

8 Bogetoft, P., J.M. Tama, and J. Tind (2000). Convex Input and Output Projections of Nonconvex Production Possibility Sets. *Management Science* 46(6), 858-869.

9 Post, G. T. (1997b): Transconcave Data Envelopment Analysis, Rotterdam Institute for Business Economics Studies (RIBES) Report, R9721/ O. Forthcoming in *European Journal of Operational Research*

10 Kuosmanen T (2000). DEA with Efficiency Classification Preserving Conditional Convexity. Forthcoming in *European Journal of Operational Research*

11 Banker, R.D., Charnes, A. and Cooper W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078-1092.

12 Farrell, M. J. (1957). The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society* Series A 120 (3), 253 – 281

13 Banker, R. D. (1993). Maximum likelihood, consistency and Data Envelopment analysis: a statistical foundation, *Management Science*, October 1993, 1265-1273.

14 Diewert, W. E. and T. J. Wales (1987). Flexible functional forms and curvature conditions, *Econometrica* 55, 43-68.

15 Avriel, M., W. E. Diewert, S. Schaible, and I. Zang (1988). *Generalised Concavity*, Plenum Press, New York.

16 Chambers, R. G., Y. Chung, and R. Färe (1996). Benefit and Distance Functions. *Journal of Economic Theory* 70, 407-418

17 Chambers, R.G., Y. Chung, and R. Färe (1998). Profit, Directional Distance Functions and Nerlovian Efficiency. *Journal of Optimization Theory and Applications* 98 (2), p.351-364.

18 Luenberger, D.G. (1992). Benefit Functions and Duality. *Journal of Mathematical Economics* 21, 261-481

19 Chiang, Alpha. C. (1984). *Fundamental methods of mathematical economics*. McGraw-Hill Inc., ISBN 0-07-010813-7.

20 Ortega, J.M. (1989). *Matrix theory*. Plenum Press. ISBN 0-306-42433-9.