

# NONPARAMETRIC EFFICIENCY ESTIMATION IN STOCHASTIC ENVIRONMENTS

LAURENS CHERCHYE,  
Catholic University Leuven

TIMO KUOSMANEN,  
Helsinki School of Economics

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

## ABSTRACT

This paper develops a new nonparametric model for efficiency estimation. In contrast to Data Envelopment Analysis (DEA), it does not impose debatable production assumptions like free disposability and convexity, and it does not assume that the data are measured without error. The estimators are asymptotically unbiased and have an asymptotic variance that is comparable to that of stochastic frontier estimators (provided the latter use a correct specification of the functional form for the production relationships). In addition, the estimators can be computed using a simple enumeration algorithm.

**Key Words: efficient frontier analysis, nonparametric analysis, Data Envelopment Analysis (DEA), convexity, disposability, stochastic disturbances.**

## 1. INTRODUCTION

*Efficient frontier analysis*, henceforth EFA, is a general methodology for analyzing production relationships and productive efficiency. Since technology information typically is incomplete, EFA constructs *empirical production frontiers* on the basis of a data set of comparable *decision-making units*, henceforth DMUs, and measures the efficiency of those DMUs relative to the frontier.

Two parallel lines in EFA have been developed, viz. the parametric *Stochastic Frontier Analysis*, henceforth SFA, and the nonparametric *Data Envelopment Analysis* (DEA) (see e.g. Fried *et al.* (1993) for an excellent introduction to both approaches). SFA typically rests on explicit specifications of the functional form of the frontier and the statistical distribution of the inefficiency terms. These specifications can be problematic, since production theory does not imply specific functional forms or statistical distributions, and because reliable empirical specification tests are not available in many cases<sup>1</sup>. By contrast, DEA does not require

---

<sup>1</sup> However, flexible functional forms may help to remedy or reduce the specification errors (e.g. Diewert and Wales (1987), and Pulley and Braunstein (1992)). For example, the well-known translog estimation function can give a second-order approximation to an arbitrary twice-differentiable function. In addition, if panel data are available, the specification of a particular inefficiency distribution can be

the specification of an explicit functional form for the frontier or an explicit statistical distribution for the inefficiency terms. But notwithstanding this nonparametric orientation, DEA is not a panacea. Specifically, we discern two problems: *restrictive production assumptions* and *stochastic disturbances*. Most DEA models assume that the production possibilities have particular properties, most notably free disposability and convexity. Unfortunately, it is in many cases difficult to find a valid justification for these assumptions. For example, congestion can violate free disposability (see e.g. Färe and Grosskopf, 1983), and economies of scale and specialization can violate convexity (see e.g. Farrell, 1959). Therefore, these assumptions exclude many interesting economic phenomena at forehand, and can introduce serious specification error into the estimation. Another well-recognized problem stems from the assumption that the variables are specified correctly and measured accurately. This assumption can be violated by stochastic disturbances, e.g. caused by measurement error, omitted variables and outliers. By relying on debatable production assumptions (free disposability and convexity) and by assuming away disturbances, DEA contrasts with the tradition of nonparametric regression techniques (see e.g. Härdle, 1990, and Yatchew, 1998), which rely on minimal maintained assumptions and which do account for stochastic disturbances.

DEA models are available that address some, but (unfortunately) not all of the aforementioned limitations. The free disposability assumption is eliminated in the convex hull model by Charnes *et al.* (1985). To circumvent the convexity assumption, a number of alternative models have been proposed in the DEA literature. These include the Free Disposable Hull (FDH) model (Deprins *et al.* 1984), which drops convexity altogether, and relaxed models by e.g. Petersen (1990) and Bogetoft (1996), Post (1997a), and Bogetoft *et al.* (1998), which assume convexity in input and/or output space only. In addition, there have been attempts to account for stochastic disturbances in DEA, most notably by using stochastic programming techniques (e.g. Land *et al.* (1994), Olesen and Petersen (1995), and Cooper *et al.* (1996, 1998)). However, to the best of our knowledge, the statistical properties of the associated efficiency estimates are not yet documented. In addition, it is not clear how to relax the maintained production assumptions in those models. As a conclusion, there currently is no DEA model that does not require disposability and convexity assumptions and that can be demonstrated to effectively account for noise.

In this paper we intend to simultaneously address all the aforementioned limitations of standard DEA. Specifically, we try to develop a model that does not impose free disposability, convexity, or other kinds of restrictive assumptions, and that does allow for stochastic disturbances. Such a model would make it possible for the data to truly 'speak for themselves' rather than being forced to use the idiom of some imposed specification or some imposed set of assumptions. We focus primarily on the estimation of efficiency. Other uses of EFA not considered here include estimation of elasticities of scale and substitution, selecting target or benchmarking points for inefficient DMUs, forecasting effects of adjusting input/output mix or production scale, among other uses<sup>2</sup>.

---

avoided by assuming a particular time series pattern for inefficiency (e.g. Schmidt and Sickles (1984), and Cornwell, Schmidt and Sickles (1990)).

<sup>2</sup> In this respect, the approach developed in this paper should not be confused with the mean-variance DEA approach (Post, 1997b). That approach has a comparable mathematical programming structure,

We do not apply existing nonparametric regression techniques to the problem of efficiency measurement. Rather, we derive from scratch an entirely new nonparametric technique, especially tailored to the problem of efficiency estimation. In this respect, it is interesting to compare our approach with the Kneip and Simar (1996) approach of using kernel estimation, an existing nonparametric regression technique. Interestingly, the asymptotic variance for the efficiency estimates of our approach can be substantially lower than that of the kernel approach, especially if the signal-to-noise ratio is low. In fact, the asymptotic variance of our approach is comparable to that of the estimators used in SFA (provided the latter use a correct specification). In addition, our approach is computationally more attractive, as the efficiency estimates can be solved using a simple enumeration algorithm. This is an important feature, because computational intensity is recognized as one of the key reasons why nonparametric regression methods have not yet infiltrated the applied literature to the degree that one might expect (see e.g. Yatchew, 1998).

The remainder of this paper is organized as follows. Section 2 recaptures the standard DEA model. Section 3 introduces our new nonparametric model. Section 4 discusses an efficient computational algorithm and Section 5 discusses the asymptotic properties. Finally, Section 6 offers concluding remarks and suggestions for further research.

## 2. DATA ENVELOPMENT ANALYSIS

The DEA methodology comprises a wide variety of mathematical programming models. These models differ with respect to the assumptions imposed on the production possibilities and the employed efficiency measure. In this study, we focus on the model used in Banker (1993) which laid the statistical foundation for DEA. Following Banker, we focus on *multiple-input single-output* problems, which allows for a comparison with SFA models. Apart from multiple-input single-output problems, this approach can be used for single-input multiple-output problems and for multiple-input multiple-output problems that can be simplified using cost or revenue functions.

For each DMU in the data set  $D = \{1, \dots, n\}$ , we consider observations on a single output  $y_j \in \mathfrak{R}_+$  and multiple inputs  $x_j = (x_1 \dots x_m) \in S$ , where  $S$  is a convex subset of  $\mathfrak{R}_+^m$ . Theoretically, the efficiency of the DMUs is defined relative to the efficient production frontier  $f : S \rightarrow \mathfrak{R}_+$ . The frontier represents the maximum amount of output that can be produced by given amounts of input. We will assume throughout the paper that the frontier is *smooth*, i.e.  $\lim_{\varepsilon \rightarrow 0^+} |f(x + \varepsilon) - f(x)| = 0$  for all  $x$  in the interior of  $S$ .

---

but it deals with the effects of errors-in-variables on *performance benchmarking* rather than *efficiency estimation*, and in addition it builds on a substantially different set of maintained production and distribution assumptions.

Observed output can deviate from the frontier, because of inefficiency terms  $u_j \in \mathfrak{R}^3$ . The relationship between observed output and frontier output can be formalized as follows:

$$(1) \quad y_j = f(x_j) - u_j \quad j \in D.$$

Unfortunately, production theory can not be applied directly in practice, because the true frontier cannot be observed. Therefore, the inefficiency terms are estimated using empirical data. DEA measures efficiency by comparing observed performance with *observed best practice reference units*. These reference units are *fictitious DMUs*, constructed as convex combinations of the DMUs in the data set. Specifically, the model maximizes, for every evaluated unit separately, the distance of the observed output from the output level of a reference unit that consumes amounts of inputs that are smaller than or equal to that of the evaluated unit. More formally, using  $\lambda = (\lambda_1 \cdots \lambda_n)^T \in \mathfrak{R}_+^n$  to denote the fractions of the DMUs in the reference unit, the inefficiency of the evaluated unit (DMU<sub>k</sub>) is estimated by solving the following maximization problem:

$$(2) \quad \hat{u}_k = \max_{\lambda \in \mathfrak{R}_+^n} \left\{ \lambda_j y_j - y_k \mid \lambda_j x_j \leq x_k; \quad \lambda_j = 1 \right\}_{j \in D}.$$

Banker (1993) demonstrated that this estimator is statistically consistent (i.e., roughly speaking, asymptotically unbiased and with a vanishing variance) if certain production and distribution assumptions hold. Specifically, the production frontier must be non-decreasing and concave, or equivalently the production set (i.e. the hypograph of the frontier) must be free disposable and convex. In addition, the inputs and the inefficiency terms are considered as random variables with independent and identical distributions with probability density functions  $g(x)$  and  $h(u)$  respectively. The inputs and inefficiency terms must be mutually independent; i.e. the joint probability density  $l(x, u)$  is given by  $l(x, u) = g(x)h(u)$ . The inputs must have a strictly positive density for the entire domain  $S$ , i.e.  $g(x) > 0 \quad \forall x \in S$ , and the inefficiency terms must have a strictly positive density at zero, but no density at negative values, i.e.  $h(0) > 0, h(u) = 0 \quad \forall u < 0$ . If these distribution and production assumptions hold, then the DEA estimator (2) is statistically consistent.

These assumptions do not include an explicit functional form for production frontier  $f(x)$  and the distribution function of the inefficiency terms  $h(u)$ . This gives DEA a comparative advantage relative to SFA, which typically do require an explicit specification. However, some of these assumptions can be restrictive in particular research environments, e.g. because they exclude economic phenomena like congestion and economies of scale and specialization. Stochastic disturbances are another problem. Recall, that  $(y_j, x_j), j = 1, \dots, n$ , represent the production data. This

---

<sup>3</sup> Following both the stochastic frontier approach and Banker (1993), we choose to model additive technical inefficiency terms. This should not be confused with Farrell's (1957) output measure of technical inefficiency, i.e.  $y_j / f(x_j)$ , which is typically employed in DEA.

is not a harmless issue of notation. Equating the data set with the true values of the inputs and outputs effectively assumes that the input-output values are correctly specified. That assumption may be violated if stochastic disturbances such as measurement error, omitted variables, and outliers perturb the observations. Standard DEA can be extremely sensitive for such disturbances, because it selects ill-diversified extreme points as reference units, which are extremely sensitive for perturbations.

### 3. AN ALTERNATIVE NONPARAMETRIC APPROACH

In this study, we introduce stochastic disturbances analogously to SFA. More specifically, in the relationship between observed output and frontier output, we assume a residual that is composed of two parts - an inefficiency term  $u_j \in \mathfrak{R}$  and a stochastic disturbance term  $v_j \in \mathfrak{R}$ , i.e.:

$$(3) \quad y_j = f(x_j) - u_j + v_j \quad j \in D.$$

In contrast to SFA, we will not specify the distribution functions for the inefficiency terms and the disturbances, so as to preserve the nonparametric nature of DEA. Unfortunately, the deconvolution of inefficiency terms and disturbances is highly dependent on these probability distributions. This is especially true for estimating the level of mean inefficiency  $E(u_j) = \mu \quad \forall j \in D$ . Therefore, it is difficult to find good estimates for the absolute value of the inefficiency terms, without making strong distribution assumptions. However, for many research purposes, measuring inefficiency relative to mean inefficiency is just as good information as measuring inefficiency in absolute terms. Such purposes include the efficiency ranking of a sample of DMUs, explaining the variation of inefficiency terms, and testing hypotheses on inefficiency differences. Therefore, we focus on estimating inefficiency relative to the mean rather than in absolute terms, i.e.:

$$(4) \quad w_j = u_j - \mu \quad j \in D.$$

We adhere to the standard distribution assumptions for the inputs and inefficiency terms discussed in Section 2. Furthermore, we assume that the disturbances are identically and independently distributed with probability density function  $k(v)$ . That distribution has a zero mean and variance  $\sigma_v^2, 0 < \sigma_v^2 < \infty$ . Furthermore, we assume that the disturbances are independent of the inputs and the inefficiency terms, i.e. the joint probability distribution is given by  $r(x, u, v) = g(x)h(u)k(v)$ .

As discussed in Section 2, standard DEA selects reference units that maximize observed output. Since output observations can include stochastic disturbances, that selection method increases the likelihood of selecting a reference unit with a positive disturbance. By contrast, below we will develop a selection method that is independent of the disturbances. In addition, that selection method favors reference units that are well diversified, and that are located in the proximity of the evaluated

unit in input space. Before we turn to that method, we first analyze some properties of independently selected reference units.

Consider the output difference relative to an arbitrary convex combination of observations as an estimator for relative inefficiency:

$$(5) \quad \hat{w}_k(\lambda) = \frac{\sum_{j \in D} \lambda_j y_j}{\sum_{j \in D} \lambda_j} - y_k \quad \lambda_j = 1; \lambda_j \geq 0 \quad \forall j \in D.$$

The approach developed below will adhere to the use of convex combinations of observations. However, the use of convex combinations does not imply that we assume that the production set is convex. Rather, our approach replaces the assumption of global convexity with a more general assumption of convexity over a neighborhood of the evaluated DMU. As progressively more DMUs are introduced in the sample, the neighborhood shrinks and the local convexity assumption becomes less restrictive. The approach gives asymptotically unbiased estimates for all smooth technologies, whether convex or non-convex.

By substituting (3), and using (4), we find that the estimator involves the following estimation error:

$$(6) \quad \hat{w}_k(\lambda) - w_k = \frac{\sum_{j \in D} \lambda_j f(x_j) - f(x_k)}{\sum_{j \in D} \lambda_j} - \frac{\sum_{j \in D} \lambda_j u_j}{\sum_{j \in D} \lambda_j} + \frac{\sum_{j \in D} \lambda_j v_j - v_k}{\sum_{j \in D} \lambda_j} + \mu.$$

If the reference unit is selected independently of output, the above distribution assumptions yield the following mean-squared error:

$$(7) \quad E\left((\hat{w}_k(\lambda) - w_k)^2\right) = \left( \frac{\sum_{j \in D} \lambda_j f(x_j) - f(x_k)}{\sum_{j \in D} \lambda_j} \right)^2 + \frac{\sum_{j \in D} \lambda_j^2 \sigma_u^2}{\sum_{j \in D} \lambda_j} + \left( \frac{\sum_{j \in D} \lambda_j^2 - 2\lambda_k + 1}{\sum_{j \in D} \lambda_j} \right) \sigma_v^2,$$

which can be decomposed into bias

$$(8) \quad E(\hat{w}_k(\lambda) - w_k) = \frac{\sum_{j \in D} \lambda_j f(x_j) - f(x_k)}{\sum_{j \in D} \lambda_j}$$

and error variance

$$(9) \quad V(\hat{w}_k(\lambda) - w_k) = \frac{\sum_{j \in D} \lambda_j^2 \sigma_u^2}{\sum_{j \in D} \lambda_j} + \left( \frac{\sum_{j \in D} \lambda_j^2 - 2\lambda_k + 1}{\sum_{j \in D} \lambda_j} \right) \sigma_v^2.$$

Here,  $\sigma_u^2, 0 < \sigma_u^2 < \infty$ , denotes the variance of the inefficiency term.

Not surprisingly, these expressions suggest that the output difference relative to an independently selected reference unit generally gives a biased estimator with a non-zero error variance. However, bias and variance vary from one reference unit to the other. More specifically, variance is related to the *degree of diversification* of the reference unit and the *degree of self-comparison*, and bias is related to the *distance in input space* of the reference unit from the evaluated unit.

Interestingly, the error variance does not depend on the unknown production frontier. Rather, the variance depends on the degree of diversification of the reference unit (i.e.  $1/\sum_{j \in D} \lambda_j^2$ ) and the degree of self-comparison, i.e. the extent to which the evaluated unit is included in the reference unit (i.e.  $\lambda_k$ ). Interestingly, increasing the degree of self-comparison (and hence lowering the degree of diversification) can *reduce* variance, because it increases the correlation between the disturbances of the reference unit and that of the evaluated unit. The statistical goodness of our approach depends heavily on accounting for this correlation, as is elaborated in Section 5.

By contrast, bias does depend on the frontier. Therefore, we are skeptical about the possibility to quantify the bias based on our minimal set of maintained assumptions. Still, the above expression implies that an unbiased estimator is obtained by restricting the distance in input space between the referencing units and the evaluated unit to zero, i.e. if  $(x_{ij} - x_{ik}) = 0 \forall j \in D : \lambda_j > 0, i = 1, \dots, m$ , then  $\hat{w}_k(\lambda)$  is unbiased. In addition, for smooth frontiers, bias generally is lower the closer the referencing units are located to the evaluated DMU, at least in the proximity of the evaluated DMU. To exploit this insight, we propose to use a distance measure  $\xi_j$  that goes to zero as the input distance goes to zero, i.e.  $\lim_{x_j \rightarrow x_k} \xi_j = 0$ , so that  $\hat{w}_k(\lambda)$  is unbiased if  $\xi_j = 0 \forall j \in D : \lambda_j > 0$ . One such measure is the squared vector norm of the inputs standardized by their sample standard deviations

$$s_i = \sqrt{\frac{1}{n-1} \sum_{j \in D} x_{ij}^2 - \frac{1}{(n-1)^2} \left( \sum_{j \in D} x_{ij} \right)^2} \quad i = 1, \dots, m, \text{ i.e.:}$$

$$(10) \quad \xi_j = \sum_{i=1}^m \left( \frac{x_{ij} - x_{ik}}{s_i} \right)^2,$$

Alternative distance measures can be used. For example, the squared Mahalanobis distance could be used here, so as to account for interdependencies between inputs. The asymptotic properties derived below apply for any distance measure that satisfies the property  $\lim_{x_j \rightarrow x_k} \xi_j = 0$ . However, the distance measure may affect the small sample

performance. The problem of selecting the appropriate distance measure is somewhat similar to selecting the appropriate weighing function or kernel in kernel regression, as the kernel translates the distance of an observation into the weight assigned to that observation. We are sceptical about the possibilities of formally deriving 'optimal' parameter values without imposing additional structure and compromising the intended nonparametric orientation. Instead, we propose a more heuristic route. In practice, the distance measure might be chosen in a data-adaptive way by trying

different values and selecting the one that minimizes out-of-sample prediction error, a technique known as cross-validation. Clarck (1975) first proposed this approach for selecting the appropriate smoothing parameter or bandwidth in kernel estimation, and Härdle and Marron (1985) and Härdle *et al.* (1988) discuss its properties. Alternatively, the model could be solved for different distance measures, so as to assess the sensitivity of the results with respect to different measures.

The above insights can be used to develop a 'good' selection method for reference units. More specifically, we expect that *well-diversified* weighed averages of *independently selected DMUs in the proximity* of the evaluated unit are good reference units. Under the above distribution assumptions, asymptotically, an arbitrary large number of DMUs can be found that use approximately the amount of inputs of the evaluated unit (see Section 5 below). However, in finite samples, it is generally not possible to find many observations that consume approximately that amount of input. Therefore, requiring the referencing units to consume as much input as the evaluated unit would result in self-identifiers (i.e.  $\lambda_k = 1, \lambda_j = 0 \forall j \in D, j \neq k$ ) or in ill-diversified reference units. However, allowing referencing units to consume different amounts of input may give biased estimates. Hence, we face a trade-off between bias and variance in the selection of the reference unit.

We therefore propose to select the reference unit that minimizes a weighed sum of the error variance and the weighed average for input distance measure (10), i.e.

$$(11) \quad \lambda^* = \arg \min_{\lambda \in \mathfrak{R}_+^D} \left\{ \lambda_j^2 \sigma_u^2 + \left( \lambda_j^2 - 2\lambda_k + 1 \right) \sigma_v^2 + \kappa \sum_{j \in D} \lambda_j \xi_j \mid \lambda_j = 1 \right\}.$$

Here,  $\kappa, 0 < \kappa < \infty$ , represents the weight associated with input distance. It can be interpreted as a 'smoothing parameter'. The optimal lambda's can be computed for several values of  $x_k \in S$ , producing an empirical representation of the unknown production technology<sup>4</sup>. The weight  $\kappa$  can be viewed as a smoothing parameter for that representation. If we specify a small value for  $\kappa$ , few observations will have a significant lambda and the resulting representation will appear rough ("unsmoothed"). By contrast, if it is set at a large value, many observations have a significant lambda and the representation will appear "smooth". In this respect,  $\kappa$  behaves similar to the smoothing parameter in kernel regression, i.e. the 'bandwidth' that controls for the local neighbourhood. Interestingly, the value of  $\kappa$  is irrelevant for the asymptotic properties of the estimators (see Section 5 below). However, for finite sample performance it can be relevant. As for the distance measure, cross-validation and sensitivity analysis could provide guidance for the selection of the appropriate  $\kappa$ .

Unfortunately, the variance terms  $\sigma_u^2$  and  $\sigma_v^2$  are typically not known, and have to be estimated. Using estimates  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$ , a reference unit can be selected as:

---

<sup>4</sup> However, our approach is specially tailored to the estimation of efficiency rather than the estimation of the frontier. This is reflected by the fact that the objective is to minimise the (penalised) variance of the inefficiency estimate rather than the variance of the frontier estimate.

$$(12) \quad \lambda^* = \arg \min_{\lambda \in \mathfrak{R}_+^D} \left\{ \sum_{j \in D} \lambda_j^2 \hat{\sigma}_u^2 + \left( \sum_{j \in D} \lambda_j^2 - 2\lambda_k + 1 \right) \hat{\sigma}_v^2 + \kappa \sum_{j \in D} \lambda_j \xi_j \mid \sum_{j \in D} \lambda_j = 1 \right\}.$$

Assuming that a priori estimates of the variance terms are available is somewhat problematic, as the calculation of such estimates frequently is one of purposes for using efficient frontier analysis in the first place. Fortunately, the goodness of the estimator (12) most essentially depends on the ratio of these estimates, i.e. the estimated *signal-to-noise ratio*  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v}$ , while the exact levels of the variance estimates  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$  are of less importance. Specifically, the asymptotic properties depend on the estimated signal-to-noise ratio  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v}$  only, as is elaborated in Section 5. In addition, if the smoothing parameter  $\kappa$  is selected using a cross-validation technique, the levels of  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$  don't affect the small samples properties either. This is because an increase in the values of  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$ , keeping the signal-to-noise ratio constant, will be completely offset by a proportional increase in the value of  $\kappa$  that minimises out-of-sample prediction error. In practice, the signal-to-noise ratio  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v}$  can be specified to reflect the subjective opinion of the researcher, or the estimates from an empirical estimation<sup>5</sup>. Alternatively, the above routes of cross-validation or sensitivity analysis may also be employed. Finally, following Varian (1985), one could turn around the problem by computing the critical signal-to-noise ratio required to reach a particular conclusion about the efficiency of the evaluated DMUs, and subsequently compare this critical value with the prior opinions concerning the precision with which the data have been measured. In any event, postulating a single signal-to-noise parameter to indicate how reliable the investigator believes the data to be is much less objectionable than the common practice of postulating an entire functional form for the production function and error distribution (as in SFA).

The reference units selected by model (12) depend on the input distances  $\xi_j$  and the estimated variance terms  $\hat{\sigma}_u$  and  $\hat{\sigma}_v$  only, with observations "nearby" having a higher  $\lambda_j^*$  than observations "far away", and the evaluated unit itself having a higher  $\lambda_k^*$  if the estimated signal-to-noise ratio is lower. However, the reference units do not depend on the disturbances  $v_j$ . In addition, model (12) favors well-diversified reference units. Brief, our approach selects *independently of the disturbances* a *well-diversified* reference unit *in the proximity* of the evaluated unit. This orientation leads us to expect that the difference  $\hat{w}_k(\lambda^*)$  is a good estimator for  $w_k$ . We expect goodness for a wide range of production frontiers (including congested and non-

---

<sup>5</sup> We assume here that the estimates  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$  are independent of the inefficiency terms and the disturbances. This assumption can be violated if the estimates are based on the observations in the data set. In that case the  $\lambda^*$  in (12) can depend on the inefficiency terms and disturbances, and the expressions for bias (8) and error variance (9) can be incorrect. Future research could focus on a generalisation that allows for dependencies.

convex ones), because DMUs *in the proximity* of the evaluated unit have approximately the same conditional output mean as the evaluated DMU. In addition, we expect goodness for a wide range of error distributions, because the reference units are *independent of the disturbances* and *well-diversified*. Unfortunately, we can't quantify based on our minimal set of assumptions the statistical goodness in finite samples. However, Section 5 derives a number of attractive asymptotic properties.

Our model computes a weighed average over a neighborhood, where the weights and the neighborhood are selected endogenously, so as to optimize the goodness measure (the penalized variance term). This approach falls somewhere in between of the existing nonparametric local averaging models and optimizing models (see e.g. Yatchew, 1998). It uses the elemental idea underlying local averaging techniques that, if a function is smooth, its value at a given point can be approximated reasonably well by using observations of the function at nearby points. However, in contrast to local averaging techniques, the weights and the neighborhood are not fixed at forehand, but rather are selected endogenously, so as to optimize the goodness measure. In this respect, there is some similarity between our approach and the use of e.g. kernel models with a variable bandwidth (see e.g. Härdle, 1990).

#### 4. COMPUTATIONAL ISSUES

Note that problem (12) is non-linear, and hence can not be solved using standard Linear Programming techniques. This might appear as a disadvantage of our approach relative to DEA, which does involve LP. However, using the mathematical theory of duality, a computational algorithm can be derived that substantially reduces the computational burden. More specifically, the optimal  $\lambda_j^*$  can be derived by finding the largest subset  $P \subseteq D$  with  $\kappa \xi_j$  less than or equal to:

$$(13) \quad \alpha(P) = \frac{2\hat{\sigma}_u^2 + \kappa \sum_{j \in P} \xi_j}{\text{card}P}.$$

We refer to this set as the optimal neighborhood  $P^*$  for the evaluated unit. More formally, that set corresponds to:

$$(14) \quad P^* = \left\{ j \in D : \kappa \max_{j \in P^*} \xi_j \leq \frac{2\hat{\sigma}_u^2 + \kappa \sum_{j \in P^*} \xi_j}{\text{card}P^*} \right\}.$$

The optimal lambda's can be calculated from  $\alpha(P^*)$  as:

$$(15) \quad \lambda_j^* = \begin{cases} \max\left(0, \frac{\alpha(P^*) - \kappa \xi_j}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}\right) & j \neq k \\ \frac{\alpha(P^*) + 2\hat{\sigma}_v^2}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} & j = k \end{cases}.$$

Proof A in the Appendix formally proves the equivalence of (15) and (12). Problem (15) can be solved with minimal computational burden. For example, one way to select  $P^*$  is by ranking the observations in ascending order by their input distance, select  $P=\{k\}$ , and sequentially enlarging  $P$  to include the observation next in rank, until  $\kappa \xi_j \leq \alpha(P) \quad \forall j \in P$  is violated<sup>6</sup>. A procedure in GAUSS-code for computing the estimator  $\hat{w}_j(\lambda^*)$  in this way for is provided below.

```
PROC(1) = INEFF(y,x,sv,su,k,j);

LOCAL d,m,n,i,z,l,p,a,w;

n=ROWS(x);
m=COLS(x);
d=SUMC(((x-x[j,.] )/STDC(x)').^2)');

z=SORTC(y~x~d,m+2);
y=z[.,1];
x=z[.,2:m+1];
d=z[.,m+2];

i=1;
p=1|ZEROS(n-1,1);
a=(2*su^2+k*p'*d)/SUMC(p);
DO WHILE (a .GE k*d[i]) .AND (i .LT n);
    p[i]=1;
    a=(2*su^2+k*p'*d)/SUMC(p);
    i=i+1;
ENDO;

l=MAXC(ZEROS(1,n)|((a-k*d)/(2*(sv^2+su^2)))');
l[1]=MAXC(0|(a+2*sv^2)/(2*sv^2+2*su^2));
w=l'*y-y[1];
RETP(w);

ENDP;
```

<sup>6</sup> Using a more efficient search algorithm to find the critical observation could further improve computational efficiency.

The global variables are defined as  $y=(y_1 \cdots y_n)^T$ ,  $x=(x_1 \cdots x_n)^T$ ,  $sv=\hat{\sigma}_v$ ,  $su=\hat{\sigma}_u$ ,  $k=\kappa$ , and the local variables are defined as  $d=(\xi_1 \cdots \xi_n)^T$ ,  $l=\lambda^*$ ,  $p=(p_1 \cdots p_n)^T$ , where

$$p_i = \begin{cases} 1 & i \in P \\ 0 & i \notin P \end{cases}, \alpha = \alpha(P), \text{ and } w = \hat{w}_j(\lambda^*).$$

## 5. ASYMPTOTIC PROPERTIES

Since the reference units are selected independently of the disturbances  $v_j$ , using (8) and (9), the bias and error variance of the estimator  $\hat{w}_k(\lambda^*)$  are given by:

$$(16) \quad E(\hat{w}_k(\lambda^*) - w_k) = \sum_{j \in D} \lambda_j^* f(x_j) - f(x_k),$$

and

$$(17) \quad V(\hat{w}_k(\lambda^*) - w_k) = \sum_{j \in D} \lambda_j^{*2} \sigma_u^2 + \left( \sum_{j \in D} \lambda_j^{*2} - 2\lambda_k^* + 1 \right) \sigma_v^2.$$

Unfortunately, it is generally not possible to formally quantify these statistics for finite samples, since they depend on the (unknown) shape of the frontier and the composition of the referencing units (which varies from one sample to the other). However, some asymptotic properties can be derived. As progressively more observations are introduced into the sample, the model can select more referencing units that are located in the close proximity of the evaluated unit and that have conditional output means comparable to that of the evaluated DMU. Asymptotically, the optimal reference unit is composed of 'a very large number' of 'very small fractions' of observations 'very close to' the evaluated unit, i.e.:

$$(18) \quad \begin{cases} \lim_{n \rightarrow \infty} \lambda_j^* = 0 & \xi_j \geq \delta \\ \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} \leq \lim_{n \rightarrow \infty} \lambda_k^* \leq \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} + \varepsilon \\ \lim_{n \rightarrow \infty} \lambda_j^* \leq \varepsilon & \xi_j \leq \delta, j \neq k \end{cases}$$

Here,  $\delta$  and  $\varepsilon$  represent non-Archimedean infinitesimal small positive values. Proof B in the Appendix formally proves this asymptotic property.

This result directly implies that the estimator is asymptotically unbiased (as shown formally in Proof C of the Appendix), i.e.

$$(19) \quad \text{Asy.} E(\hat{w}_k(\lambda^*) - w_k) = 0,$$

and that it has the following asymptotic error variance (as shown formally in Proof D of the Appendix):

$$(20) \quad \text{Asy. Var}(\hat{w}_k(\lambda^*) - w_k) = \sigma_v^2 \left[ 1 - \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2 + \sigma_u^2 \left[ \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2.$$

Unfortunately, as in the parametric approach, the inefficiency estimator has certain 'intrinsic' variability. The consistency of DEA estimators for the deterministic case does not carry over to our estimator in the stochastic case, because the uncertainty regarding the evaluated point can not be diversified. The variance (20) depends on the variance terms  $\sigma_u^2$  and  $\sigma_v^2$ . In addition, it depends on the goodness of the estimated signal-to-noise ratio  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v}$ . Furthermore, the variance is independent of the distance measure  $\xi_j$ , the estimated levels of the variance terms  $\hat{\sigma}_u^2$  and  $\hat{\sigma}_v^2$ , and the smoothing parameter  $\kappa$ .

The error variance is minimal, if the estimate is correct, i.e.  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v} = \frac{\sigma_u}{\sigma_v}$ . In that case, (20) reduces to:

$$(21) \quad \text{Asy. Var}(\hat{w}_k(\lambda^*) - w_k) = \sigma_v^2 \left[ 1 - \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2} \right].$$

Interestingly, the asymptotic error variance associated with the Kneip and Simar (1996) kernel estimator is  $\sigma_v^{2.7}$ . Hence, our approach can substantially improve upon the kernel estimator, especially if the signal-to-noise ratio is low. Our explanation for this potential improvement is the two-stage orientation of the Kneip and Simar procedure. It is tailored to finding a statistically good (asymptotically unbiased and efficient) frontier estimate, and derives the efficiency estimate as a 'side-product' in a second stage. By contrast, our approach is specially tailored to finding a good efficiency estimate. In this respect, our approach comes closer to the original data envelopment analysis, and the Kneip and Simar approach comes closer to the parametric approach. The kernel approach does not require a signal-to-noise ratio. Our approach does require such an estimator, as it focuses on estimating efficiency rather than the frontier and it uses this estimator to reduce the variance below the level of the kernel estimator. Specifically, if the signal-to-noise ratio is lowered, our model increases the degree of self-comparison (i.e.  $\lambda_k^*$ ), hence increasing the correlation between the disturbances of the reference unit and that of the evaluated unit, and reducing the variance of the efficiency estimate. This correlation is ignored if (as in

---

<sup>7</sup> This asymptotic variance is obtained by applying the least squares firm effect estimators (4.1) and (4.5) in Kneip and Simar (1996) to cross-section data (i.e.  $T=1$ ). Kneip and Simar considered the more general case where panel data are available. Using panel data, the uncertainty regarding the evaluated DMU can be diversified away, and consistent estimates can be obtained. Extending our approach towards the use for panel data is a challenge for future research.

the Kneip and Simar approach) efficiency is measured in a second stage relative to a frontier that is fixed in the first stage.

The asymptotic coefficient of determination between  $w_k$  and its estimator  $d_k(\lambda^*)$  can be calculated from (20):

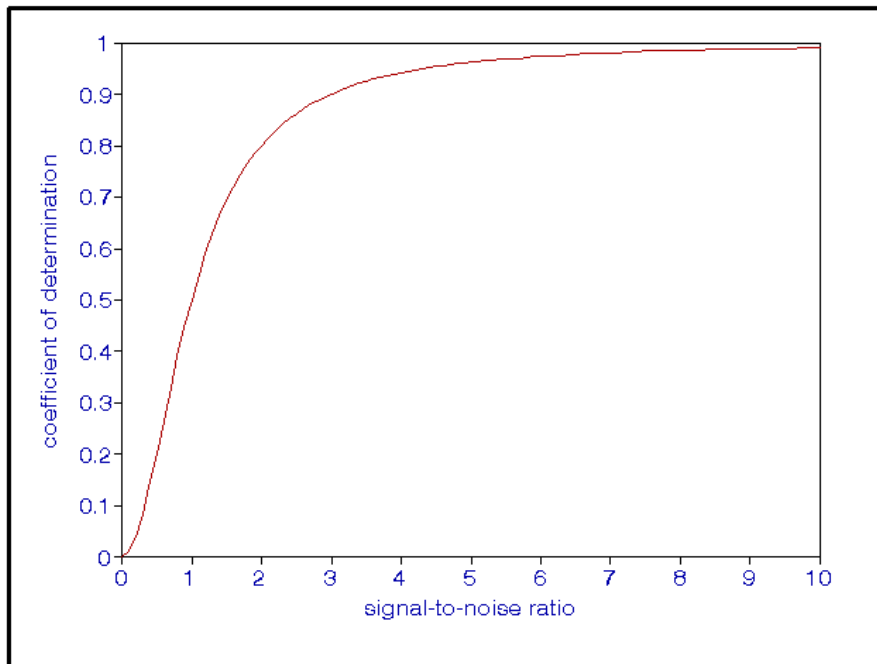
$$(22) \quad \text{Asy. } \rho^2(\hat{w}_k(\lambda^*), w_k) = 1 - \frac{\text{Asy. Var}(\hat{w}_k(\lambda^*) - w_k)}{\sigma_w^2}$$

$$= 1 - \frac{\sigma_v^2}{\sigma_u^2} \left[ 1 - \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2 - \left[ \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2.$$

This expression depends on the signal-to-noise ratio  $\frac{\sigma_u}{\sigma_v}$  and the goodness of its estimate  $\frac{\hat{\sigma}_u}{\hat{\sigma}_v}$  solely. The error variance is minimal, if the estimate is correct. In that case, (22) reduces to:

$$(23) \quad \text{Asy. } \rho^2(\hat{w}_k(\lambda^*), w_k) = 1 - \frac{\sigma_v^2}{\sigma_u^2} \left[ 1 - \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2} \right].$$

Figure 1 displays this asymptotic statistic as a function of the signal-to-noise ratio.



**Figure 1** The asymptotic coefficient of determination

The asymptotic coefficient of determination is exactly identical to the population coefficient of determination of the best linear stochastic frontier model, as derived by Waldman (1984). In addition, it is very close to that of the non-linear conditional expectation estimators proposed by Jondrow *et al.* (1982). However, as in DEA, the proposed estimator does not depend on a correct specification of the functional form of the production relationships and the distribution function of the inefficiency terms  $h(u)$ . In addition, in contrast to DEA, these results do not require free disposability and convexity, and the results do allow for stochastic disturbances.

## 6. CONCLUSIONS AND ROUTES FOR FUTURE RESEARCH

We have presented a new nonparametric model for efficient frontier analysis. In contrast to standard DEA models, it does not assume free disposability or convexity, and hence can accommodate congestion and non-convex production characteristics, such as economies of scale and specialization. In addition, our approach does not assume that the input-output data are measured accurately, and hence can accommodate stochastic disturbances, such as those caused by measurement error, omitted variables and outliers. The efficiency estimates are asymptotically unbiased and the asymptotic variance is comparable to that of the estimators used in SFA (provided the latter use a correct specification). Interestingly, this asymptotic variance can be substantially lower than that of the Kneip and Simar (1996) approach of using kernel regression. In addition, our approach is computationally much less demanding, as using a simple enumeration algorithm can solve it.

Notwithstanding these powerful statistical and computational results, this paper constitutes a mere starting point for developing a new nonparametric approach to EFA. In this respect, we see many routes for future research.

First, this paper has focussed on relaxing some debatable assumptions in the DEA methodology. However, we have maintained a number of assumptions that can be restrictive in many research environments. This calls for at least the following routes of research:

1. One of the key attractions of DEA is its flexibility in dealing with multiple-input, multiple-output technologies. By contrast, we have focused on multiple-input single-output and single-input multiple-output cases and multiple-input multiple-output cases that can be simplified using cost or revenue functions. Therefore, future research could focus on generalizing our approach towards multiple-input, multiple-output cases that cannot be simplified using cost or revenue functions.
2. We have maintained a number of statistical distribution assumptions that can be restrictive in many research environments. More specifically, we have assumed homoskedasticity and independence for the inputs, the inefficiency terms and the disturbances. These assumptions can be restrictive in many research environments. For example, *homoskedasticity* for the inefficiency term in absolute terms implies *heteroskedasticity* in percentage terms, as correctly pointed out by one of the referees for this paper. We believe the analysis below can be generalised in a straightforward way to include such effects. For example, as suggested by the aforementioned referee, heteroskedasticity can be introduced by using a log interpretation of the variables.

Second, relying on minimal production and distribution assumptions can introduce additional error in small samples. If the DMUs in the data set are considered as a random sample from a hypothetical infinite universe, computation is subject to sampling error. Sampling error occurs if few DMUs operate near the efficient frontier, and hence the sample of DMUs gives a poor representation of the universe of production possibilities. This problem is especially relevant if the technology involves many input-output variables (the *curse of dimensionality*) and if the density near the frontier is sparse. This suggests at least the following routes for future research:

3. Analyze the small sample performance using econometric analysis or simulation studies. In this respect, one could use the existing tools that allow for the analysis of the sampling distribution of DEA estimators and that allow for bias correction and the estimation of confidence intervals (e.g. Hall *et al.* (1995), Simar and Wilson (1998), and Gijbels *et al.* (1999)).
4. Sampling error is more serious if only few production assumptions are imposed. For example, it is well documented (e.g. Kneip *et al.* (1998) and Park *et al.* (1997)) that efficiency estimates from the free disposal hull model (Deprins *et al.* 1984), which assumes free disposability only, suffer more from sampling error than the estimates from the Banker *et al.* (1984) model, which assumes convexity in addition to free disposability. Therefore, it is generally desirable to include as much production information as possible. This provides a strong stimulus for developing models that do impose disposability, convexity or other production assumptions (like returns-to-scale assumptions), tests for testing those assumptions, and for analyzing the potential improvements associated with those assumptions. Still, this paper deliberately focused on using minimal production assumptions, because in many cases neither economic theory nor empirical tests can provide a valid justification for such assumptions.
5. We have focused on cross-section data exclusively. Kneip and Simar considered the more general case where panel data are available. Using panel data, the uncertainty regarding the evaluated DMU can be diversified away, and consistent estimates can be obtained. Still, using panel data is complicated by issues of data availability and the need to specify a pattern for inefficiency over time. Extending our approach towards the use for panel data is a challenge for future research. Similarly, analysing the inefficiency of groups of DMUs is a way to reduce the inherent uncertainty about firm-specific inefficiency estimates.

Third, it is not clear how the goodness of our approach is affected by the choice of the parameters that we have left unspecified, i.e. the distance measure, the variance estimators and the smoothing parameter. We have demonstrated that the asymptotic properties do not depend on these parameters (apart from the estimated signal-to-noise ratio). However, the parameters can affect small sample performance. We are sceptical about the possibilities of formally deriving 'optimal' parameter values without imposing additional structure and compromising the intended nonparametric orientation. Instead, we propose the more heuristic routes of using cross-validation techniques and sensitivity analysis. However, we have not discussed these issues in great detail, and further research is required before definite conclusions can be drawn.

Finally, this paper has focused primarily on the estimation of efficiency. Future research could focus on exploiting the insights developed in this paper for other uses of EFA, including estimation of elasticities of scale and substitution, selecting target or benchmarking points for inefficient DMUs, forecasting effects of adjusting input/output mix or production scale, among other uses.

Brief, this paper leaves many important questions unanswered. However, the powerful statistical and computational results derived in this paper should provide a strong stimulus to direct further research effort towards this technique.

#### **ACKNOWLEDGEMENTS**

We would like to express our appreciation to Paul de Boer, Jan Brinkhuisen, Teun Kloek, and Jaap Spronk, as well as the anonymous *Operations Research* referees, for their helpful comments and stimulating discussion.

#### **REFERENCES**

- Banker, R. D. (1993): 'Maximum likelihood, consistency and Data Envelopment analysis: a statistical foundation', *Management Science* 39 (10), 1265-1273.
- 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.
- Bogetoft, P. (1996): DEA on Relaxed Convexity Assumption, *Management Science* 42(3), 457-465.
- Bogetoft, P., Tama, J.M., Tind, J. (2000): Convex input and output projections of nonconvex production possibility sets, *Management Science* 46(6), 858-869.
- Brinkhuis, J. (1996): 'Introduction to duality in optimization theory', *Journal of Optimization Theory and Applications* 91 (3).
- Charnes, A., W. Cooper, B. Golany, L. Seiford, J. Stutz (1985): 'Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions', *Journal of Econometrics* 30 (1): 91-107.
- Clarck, R. (1975): 'A Calibration Curve for Radiocarbon Dates', *Antiquity* 49: 251-266.
- Cooper, W.W., Z. M. Huang and S. X. Li (1996): 'Satisficing DEA models under chance constraints'. *The Annals of Operations Research* 66: 279-295.
- Cooper, W.W., Z.M. Huang, V. Lelas, S. Xi and O.B. Olesen (1998): 'Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA'. *Journal of Productivity Analysis* 9 (1): 53-79.
- Cornwell, C., P. Schmidt and R. C. Sickles (1990): 'Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels', *Journal of Econometrics* 46: 185-200.
- Deprins, D., L. Simar and H. Tulkens (1984): 'Measuring labour efficiency in post offices', in: *The performance of public enterprises* (M. Marchand, P. Pestieu and H. Tulkens, eds.), pp. 243-267, North Holland, Amsterdam.
- Diewert, W. E. and T. J. Wales (1987): 'Flexible functional forms and global curvature conditions', *Econometrica* 25: 43-68.
- Farrell, M. J. (1957): 'The measurement of efficiency of production', *J. Roy. Statist. Soc, Ser. A*, 120: 251-281.

- Farrell, M. J. (1959): 'The convexity assumption in the theory of competitive markets', *Journal of Political Economy* 67: 377-281.
- Färe, R., and S. Grosskopf (1983): Measuring Congestion in Production, *Zeitschrift für Nationalökonomie* 43, 257-271.
- Fried, H.O., C.A.K. Lovell, and S. S. Schmidt (eds.) (1993): The Measurement of Productive Efficiency: Techniques and Applications, Oxford University Press, New York.
- 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 (445): 220-228.
- Härdle, W. (1990): '*Applied Nonparametric Regression*', Cambridge: Cambridge University Press.
- Härdle, W. and J.S. Marron (1985): 'Optimal bandwidth selection in nonparametric regression estimation', *Annals Statist.* 13: 1465-1481.
- Härdle, W., P. Hall and J. S. Marron (1988): 'How far are automatically chosen regression smoothing parameters from their optimum?', *J. Amer. Statist. Assoc.* 83: 86-99.
- Hall, P., Härdle, W. and L. Simar (1995), 'Iterated Bootstrap with Application to Frontier Models', *The Journal of Productivity Analysis* 6: 63-76.
- Jondrow, J., C. A. K. Lovell, I. S. Materov, P. Schmidt (1982): 'On the estimation of technical efficiency in the stochastic frontier production function model', *Journal of Econometrics* 19: 1377-1398.
- Kneip, A., B. U. Park and L. Simar (1998): 'A note on the convergence of nonparametric DEA efficiency measures', *Econometric Theory* 14: 783-793.
- Kneip, A. and L. Simar (1996): 'A general framework for frontier estimation with panel data', *Journal of Productivity Analysis* 7: 187-212.
- Land, K. C., A. K. Lovell and S. Thore (1994): 'Chance-Constrained Data Envelopment Analysis', *Managerial and Decision Economics* 14: 541-554.
- Olesen, O. B. and N. C. Petersen (1995): 'Chance Constrained Efficiency Evaluation', *Management Science* 41: 442- 457.
- Park, B.U., L. Simar and C. Weiner (1997): 'FDH efficiency scores from a stochastic point of view', Discussion Paper 9715, Institut de Statistique, Universtite Catholique de Louvain, under revision for *Econometric Theory*.
- Petersen, N. C. (1990): 'Data Envelopment Analysis on a relaxed set of assumptions', *Management Science* 36 (3): 305-314.
- Post, G. T. (1997a): 'Transconcave Data Envelopment Analysis', Rotterdam Institute for Business Economics Studies (RIBES) Report R9721/O, ISBN 90-5086-256-X, revised version forthcoming in *European Journal of Operational Research*.
- Post, G. T. (1997b): 'Performance benchmarking in stochastic environment using Mean-Variance Data Envelopment Analysis', Rotterdam Institute for Business Economics Studies (RIBES) Report R9701/ O, ISBN 90-5086-235-7, revised version forthcoming in *Operations Research*.
- Pulley, L. B. and Y. M. Braunstein (1992): 'A composite cost function with an application to economies of scope in banking', *Review of Economics and Statistics* 74: 221-230.
- Schmidt, P. and R. C. Sickles (1984): 'Production Frontiers and Panel Data', *Journal of Business and Economic Statistics* 2 (4): 367-374.

Simar, L. and P. Wilson (1998): 'Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models', *Management Science* 44 (1): 49-61.

Varian, H (1985): 'Nonparametric Analysis of Optimising Behaviour with Measurement Error', *Journal of Econometrics* 30 (1/2): 445-458.

Waldman, M. W. (1984): 'Properties of technical efficiency estimators in the stochastic frontier model', *Journal of Econometrics* 25: 353-364.

Yatchew, A. (1998): 'Nonparametric Regression techniques in Economics', *Journal of Economic Literature* 36: 669-721.

## APPENDIX: FORMAL PROOFS

### Proof A

The Lagrangean function associated with (12) is:

$$(A.1) \quad L(\lambda, \alpha) = \sum_{j \in D} \lambda_j^2 \hat{\sigma}_u^2 + \left( \sum_{j \in D} \lambda_j^2 - 2\lambda_k + 1 \right) \hat{\sigma}_v^2 + \kappa \sum_{j \in D} \lambda_j \xi_j - \alpha \left( \sum_{j \in D} \lambda_j - 1 \right).$$

We have from duality theory (e.g. Brinkhuis, 1996):

$$(A.2) \quad \min_{\lambda, \alpha} L(\lambda, \alpha) = \max_{\alpha} L(\lambda^*, \alpha).$$

As (12) imposes  $\lambda_j \geq 0 \forall j \in D$ , using Kuhn-Tucker conditions, we find:

$$(A.3) \quad \lambda_j^* = \begin{cases} \max \left( 0, \frac{\alpha - \kappa \xi_j}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} \right) & j \neq k \\ \max \left( 0, \frac{\alpha + 2\hat{\sigma}_v^2}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} \right) & j = k \end{cases}.$$

Therefore, using  $\sum_{j \in D} \lambda_j^* (\kappa \xi_j - \alpha) - 2\hat{\sigma}_v^2 \lambda_k^* = -2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2) \sum_{j \in D} \lambda_j^{*2}$ , we find:

$$(A.4) \quad L(\lambda^*, \alpha) = \alpha - (\hat{\sigma}_u^2 + \hat{\sigma}_v^2) \sum_{j \in D} \lambda_j^{*2} + \hat{\sigma}_v^2.$$

Consequently, using (A.2) and substituting (A.3) in (A.4), the dual of (12) is:

$$(A.5) \quad \max_{\alpha} \left[ \alpha - \frac{\max(0, \alpha + 2\hat{\sigma}_v^2)^2 + \sum_{j \in D, j \neq k} \max(0, \alpha - \kappa \xi_j)^2}{4(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} + \hat{\sigma}_v^2 \right].$$

Since  $\kappa \xi_j \leq \alpha^* \quad \lambda_j^* \geq 0 \forall j \in D$ , the solution  $\alpha^*$  is associated with the neighborhood  $P^* = \{j \in D \mid \kappa \xi_j \leq \alpha^*\}$ . In addition,  $\max(0, \alpha^* - \kappa \xi_j) = 0 \forall j \notin P^*$ . Therefore, given the neighborhood, (A.5) reduces to:

$$(A.6) \quad \max_{\alpha} \left[ \alpha - \frac{(\alpha + 2\hat{\sigma}_v^2)^2 + \sum_{j \in P^*, j \neq k} (\alpha - \kappa \xi_j)^2}{4(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} + \hat{\sigma}_v^2 \right].$$

The optimal solution to this problem is simply:

$$(A.7) \quad \alpha(P^*) = \frac{2\hat{\sigma}_u^2 + \kappa \sum_{j \in P^*} \xi_j}{\text{card}P^*}.$$

This solution is also the solution to (A.5), if  $\kappa \xi_j \leq \alpha(P^*) \forall j \in P^*$  and  $\kappa \xi_j \geq \alpha(P^*) \forall j \notin P^*$ .

Substituting (A.7) for  $\alpha$  in (A.4), and using  $\alpha(P^*) \geq 0$ , we find

$$(A.8) \quad \lambda_j^* = \begin{cases} \max\left(0, \frac{\alpha(P^*) - \kappa \xi_j}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}\right) & j \neq k \\ \frac{\alpha(P^*) + 2\hat{\sigma}_v^2}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} & j = k \end{cases} . \blacksquare$$

### Proof B

Reformulate (14) as

$$(B.1) \quad P^* = \left\{ j \in D : \left( \max_{i \in P^*} \xi_i - \xi_j \right) \leq \frac{2\hat{\sigma}_u^2}{\kappa} \right\},$$

and (15) as

$$(B.2) \quad \lambda_j^* = \begin{cases} 0 & j \notin P^* \\ \frac{\alpha(P^*) - \kappa \xi_j}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} & j \in P^* \setminus k \\ \frac{\alpha(P^*) + 2\hat{\sigma}_v^2}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} & j = k \end{cases} .$$

We build our proof on two useful results. First, since  $\left( \max_{i \in P^*} \xi_i - \xi_j \right) \geq \text{card}\{j \in D : \xi_j \leq \theta\} \left( \max_{i \in P^*} \xi_i - \theta \right)$

$\forall \theta \in \mathfrak{R}_+$ , we find:

$$(B.3) \quad P^* \subseteq \left\{ j \in D : \text{card}\{j \in D : \xi_j \leq \theta\} \left( \max_{i \in P^*} \xi_i - \theta \right) \leq \frac{2\hat{\sigma}_u^2}{\kappa} \right\} \quad \forall \theta \in \mathfrak{R}_+ .$$

Second, asymptotically, the data set contains an arbitrarily large number of units with proximity of at most the non-Archimedean infinitesimal small positive value  $\delta$ , i.e. that lie within the non-empty set  $\Xi(x_k, \delta) = \{x \in S \mid \xi(x) \leq \delta\}$ . The probability that an unit lies in this set is  $\Pr(x_k, \delta) = \frac{g(x_k) \partial x_k}{\Xi(x_k, \delta)}$ ,

where  $g(x_k)$  denotes the probability density function introduced in Section 2. Since the inputs are mutually independent, the probability of finding at least an arbitrarily large number  $\gamma$  ( $\gamma \in \mathfrak{N}_+, \gamma \leq n$ ) of such units amounts to

$$(B.4) \quad \Pr[\text{card}\{j \in D \mid \xi_j \leq \delta\} \geq \gamma] = \sum_{i=\gamma}^n \binom{n}{i} \Pr(x_k, \delta)^i (1 - \Pr(x_k, \delta))^{n-i} \quad \forall \gamma \in \mathfrak{N}_+, \gamma \leq n .$$

Since the inputs have a strictly positive density for the complete production set,  $\Pr(x_k, \delta) > 0$ , this probability converges to unity as the sample size increases, i.e.:

$$(B.5) \quad \lim_{n \rightarrow \infty} \Pr[\text{card}\{j \in D \mid \xi_j \leq \delta\} \geq \gamma] = 1 \quad \forall \gamma \in \mathfrak{N}_+, \gamma \leq n .$$

Combining (B.3) and (B.5), we find that

$$(B.6) \quad \lim_{n \rightarrow \infty} P^* \subseteq \{j \in D : \xi_j \leq \delta\},$$

and hence  $\lim_{n \rightarrow \infty} \left( \max_{j \in P^*} \xi_j \right) \leq \delta$  and  $\lim_{n \rightarrow \infty} \alpha(P^*) \leq \delta$ . Using these results and (B.2), we find:

$$(B.7) \quad \lim_{n \rightarrow \infty} \lambda^* \in \Lambda = \left\{ \lambda^* \in \mathfrak{R}_+^n \mid \lambda_j^* = 0 \quad \xi_j \geq \delta; \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} \leq \lambda_k^* \leq \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} + \varepsilon; \lambda_j^* \leq \varepsilon \quad \xi_j \leq \delta, j \neq k \right\},$$

where  $\varepsilon$  represents the non-Archimedean infinitesimal small positive value  $\varepsilon = \frac{\kappa \delta}{2(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}$ .  $\blacksquare$

### Proof C

The frontier is assumed smooth, i.e.  $\lim_{\varepsilon \rightarrow 0} |f(x + \varepsilon) - f(x)| = 0$  for all  $x$  in the interior of  $S$ . Therefore,

$\xi_j \leq \delta$  implies:

$$(C.1) \quad |f(x_j) - f(x_k)| < \varphi \quad \forall \varphi > 0$$

for all  $x_k$  in the interior of  $S$ . Therefore, property (18) implies asymptotic unbiasedness:

$$(C.2) \quad \lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{j=1}^n \lambda_j^* f(x_j) - f(x_k) \right| \leq \varphi \quad \forall \varphi > 0. \blacksquare$$

### Proof D

Consider the reference unit  $\tau \in \mathfrak{R}_+^n : \tau_k = \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)}; \tau_j = 0 \quad j \in D : j \neq k$ . Substituting  $\tau$  in (9), we find

the following variance:

$$(D.1) \quad V(\hat{w}_k(\tau) - w_k) = \sigma_v^2 \left[ 1 - \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2 + \sigma_u^2 \left[ \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2.$$

The variance associated with the reference unit  $\lambda \in \mathfrak{R}_+^n$  can be expressed as:

$$(D.2) \quad V(\hat{w}_k(\lambda) - w_k) = V(\hat{w}_k(\tau) - w_k) + \left( \sum_{j \in D} \lambda_j^2 - \tau_k^2 \right) (\sigma_v^2 + \sigma_u^2) - 2(\lambda_k - \tau_k) \sigma_v^2.$$

The minimum variance for peers from  $\Lambda$  (as defined in (B.7)) amounts to

$$(D.3) \quad \min_{\lambda \in \Lambda} V(\hat{w}_k(\lambda) - w_k) \geq V(\hat{w}_k(\tau) - w_k) - 2\sigma_v^2 \varepsilon,$$

and the maximum variance amounts to

$$(D.4) \quad \max_{\lambda \in \Lambda} V(\hat{w}_k(\lambda) - w_k) \leq V(\hat{w}_k(\tau) - w_k) + \varepsilon \left( n\varepsilon + 2 \frac{\hat{\sigma}_v^2}{(\hat{\sigma}_u^2 + \hat{\sigma}_v^2)} \right) (\sigma_v^2 + \sigma_u^2).$$

Combining (D.3) and (D.4) with property (18), we find

$$(D.5) \quad \lim_{n \rightarrow \infty} \left| V(\hat{w}_k(\lambda^*) - w_k) - \left( \sigma_v^2 \left[ 1 - \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2 + \sigma_u^2 \left[ \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_u^2} \right]^2 \right) \right| \leq \varphi \quad \forall \varphi > 0,$$

or alternatively (20).  $\blacksquare$