Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation

Rajiv D. Banker
Carlson School of Management, University of Minnesota, Minneapolis, Minnesota 55455

This paper provides a formal *statistical* basis for the efficiency evaluation techniques of data envelopment analysis (DEA). DEA estimators of the *best practice* monotone increasing and concave production function are shown to be also *maximum likelihood* estimators if the deviation of actual output from the efficient output is regarded as a stochastic variable with a monotone decreasing probability density function.

While the best practice frontier estimator is biased below the theoretical frontier for a finite sample size, the bias approaches zero for large samples. The DEA estimators exhibit the desirable asymptotic property of consistency, and the asymptotic distribution of the DEA estimators of inefficiency deviations is identical to the true distribution of these deviations. This result is then employed to suggest possible statistical tests of hypotheses based on asymptotic distributions.
(*Data Envelopment Analysis; Production Frontier; Nonparametric Estimation; Maximum Likelihood Estimates; Consistency; Hypothesis Tests*)

1. Introduction

Data envelopment analysis (DEA), introduced by Charnes et al. (1978, 1981), and extended by Banker (1980, 1984) and Banker et al. (1984), provides a nonparametric and extremal method for estimating production frontiers and evaluating the relative efficiency of decision making units (DMUs). In this paper, I shall develop a statistical foundation for DEA by identifying conditions under which DEA estimators are (statistically) consistent and maximize likelihood in a single output multiple input setting.

Antecedents of DEA can be traced to Farrell (1957). This same source has led to the development of a body of work in the economics literature beginning with Aigner and Chu (1968), who considered the estimation of a parametrically specified *best practice* production frontier, specifying it as $y^0 = g(\mathbf{x})$, where y^0 represents the efficient level of a single output and $\mathbf{x} \equiv (x_1, \cdots x_m)$ represents a vector of m inputs.[1] The function $g(\cdot)$ is specified parametrically as $\beta_0 + \sum_{i=1}^{m} \beta_i x_i$, or in terms of log-linear, log-quadratic or other variations that are linear in the parameters β. The deviation term ϵ_j for an observation j, $j = 1, \cdots n$, is specified as $\epsilon_j = y_j^0 - y_j = g(\mathbf{x}_j) - y_j$, where y_j is the observed output level. The parameters of the function $g(\cdot)$ are estimated using either of the following two criteria: (i) minimize $\sum_{j=1}^{n} \epsilon_j$, $\epsilon_j \geq 0$; or (ii) minimize $\sum_{j=1}^{n} \epsilon_j^2$, $\epsilon_j \geq 0$.

[1] Aigner et al (1977) and Meeusen and van den Broeck (1977) suggest a stochastic parametric production frontier estimation method. This corresponds to the nonparametric stochastic data envelopment analysis models of Banker (1990) and maximum likelihood estimation models of Banker and Maindiratta (1990). The models in all of these four papers include a symmetric random error term in addition to the term to represent DMU inefficiency, and therefore, they are not considered here in the comparison of the basic DEA model with the Aigner-Chu type best practice production frontier estimation models.

0025-1909/93/3910/1265$01.25

Schmidt (1976) showed that the Aigner-Chu models are equivalent to *maximum likelihood estimation* (MLE) models if an appropriate (probability) distribution is specified for the one-sided deviation term ϵ_j. That is, if ϵ_j are assumed to be independently and identically distributed[2] with a probability density function $f(\epsilon)$, then for some specification of the function $f(\cdot)$ the Aigner-Chu estimates of ϵ_j maximize the likelihood function $\prod_{j=1}^{n} f(\epsilon_j = g(\mathbf{x}_j) - y_j)$. Maximizing likelihood is equivalent to minimizing the sum of deviations if $f(\epsilon)$ is exponential, and to minimizing the sum of squared deviations if $f(\epsilon)$ is half-normal. Thus, the Aigner-Chu models are considered to have a *statistical* basis; see, for instance, Forsund et al. (1980).

Data envelopment analysis, on the other hand, has been regarded as a mathematical programming efficiency evaluation technique without any statistical foundation or justification. For instance, Schmidt (1985) classifies DEA as a *non-statistical* approach and states: "I am very skeptical of non-statistical efficiency measurement exercises, certainly as they are now carried out and perhaps in any way in which they could be carried out . . . I see no virtue whatever in a non-statistical approach to data."

The work of Banker (1980), and Banker et al. (1984) showed that in the single output case, for instance, DEA measures DMU efficiency relative to a monotone increasing and concave production frontier.[3] This line of research is pursued in §2 to show that DEA estimates of any ϵ_j result from a minimization operation that does not depend on estimates of other ϵ_k, $k \neq j$. This suggests that DEA can be interpreted in a manner similar to Schmidt's (1976) MLE interpretation of the Aigner-Chu models, with the principal difference being the specification of the production frontier in DEA as a nonparametric monotone increasing and concave function, instead of a parametric form linear in the parameters.

[2] In particular, it is assumed that ϵ_j are distributed independent of the choice of the input vector $\mathbf{x}_j$. See, for instance, Zellner et al. (1966). This assumption will be maintained throughout this paper

[3] Although only the Banker-Charnes-Cooper model is considered here, the analysis is directly applicable to several other DEA models including the original linear homogeneous model of Charnes et al. (1978). Extensions to the multiple output multiple input case and evidence from a Monte Carlo study are discussed in subsequent papers

Thus, in §3 of this paper I represent DEA as a MLE model:

$$\underset{f(\cdot),g(\cdot)}{\text{Maximize}} \prod_{j=1}^{n} f(\epsilon_j = g(\mathbf{x}_j) - y_j) \tag{1}$$

subject to

$$g(\cdot) \text{ is monotone increasing and concave, and} \tag{1.1}$$

$$f(\epsilon) = 0 \quad \text{for} \quad \epsilon < 0, \quad \text{that is}$$

$$\epsilon_j = g(\mathbf{x}_j) - y_j \geq 0. \tag{1.2}$$

In §3 I examine the structure that must be imposed on the probability density function $f(\epsilon)$ to make DEA equivalent to the above MLE model for estimating the production correspondence $y = g(\mathbf{x})$. DEA estimators maximize likelihood for a *broad* class of probability distributions, including both exponential and half-normal distributions considered by Schmidt (1976). Unlike the Aigner-Chu models, the DEA estimation methods are thus shown to be relatively robust to the specification of the probability distribution for the one-sided deviation term. There is a fundamental difference between the nonparametric DEA estimation and conventional parametric estimation methods. In the latter case, the production function $g(\mathbf{x})$ is specified in terms of a *finite* number of parameters (β's), and the estimation of these parameters suffices for the complete estimation of $g(\mathbf{x})$ for all values of $\mathbf{x}$. In contrast, for DEA estimation functional value $g(\mathbf{x}_j)$ for each sample observation $j = 1, \cdots n$, represents a parameter to be estimated; therefore, the number of parameters $g(\mathbf{x}_j)$ to be estimated in DEA increase as the sample size n increases. While this provides more flexibility in fitting the true monotone increasing and concave production frontier, the resultant incidental parameters problem implies that usual statistical properties of MLE do not apply directly.

Statistical properties of the DEA estimators are examined in §4 without appealing to the MLE result. I show that the frontier estimators are biased downward for a finite sample size, and I provide a lower bound for the probability of a specified level of inaccuracy in the estimates of $g(\mathbf{x})$. But, more importantly, I also show that asymptotically this bias reduces to zero; that is, the DEA estimators are *consistent*. When the true production function is monotone increasing and concave, DEA estimators are consistent if the probability of arbitrarily small deviations ϵ is strictly positive, a *weaker* condition than the monotone decreasing probability density function required for the maximum likelihood result. In particular, the asymptotic distribution of the DEA estimators of the inefficiency deviations for a given set of DMUs is identical to the true distribution of these deviations.

This last result is potentially important as it provides the statistical foundation for constructing asymptotic hypothesis tests in DEA. Section 5 describes statistical tests to evaluate hypotheses about differences in the inefficiencies of two (or more) subsets of the sample observations, when specific distributions such as exponential and half-normal are assumed for the inefficiency deviations, and also when no parametric assumptions are imposed on their distribution.

2. Monotone Increasing and Concave Production Frontiers

In this section I follow Banker (1980) and Banker et al. (1984) to formalize the relation between DEA models and the estimation of monotone increasing and concave production frontiers. I first consider n observations on a single output $y_j \geq 0$, and m inputs $x_{ij} \geq 0$, $i = 1, \cdots m$, for each of $j = 1, \cdots n$ DMUs. The production correspondence between the frontier output (y^0) and m inputs ($\mathbf{x}$), is represented as $y^0 = g(\mathbf{x})$, $\mathbf{x} \equiv (x_1, \cdots x_m) \in X$, where X is a convex subset of R^m.[4] In addition, I specify the following structure on the function $g(\cdot)$: $X \to R$.

POSTULATE 1. Monotonicity of Production Function. *If* $\mathbf{x}' \geq \mathbf{x}''$ *then* $g(\mathbf{x}') \geq g(\mathbf{x}'')$ *for all* $\mathbf{x}', \mathbf{x}'' \in X$.

POSTULATE 2. Concavity of Production Function. *If* $0 \leq \theta \leq 1$ *then* $\theta g(\mathbf{x}') + (1 - \theta) g(\mathbf{x}'') \leq g(\theta \mathbf{x}' + (1 - \theta)\mathbf{x}'')$ *for all* $\mathbf{x}', \mathbf{x}'' \in X$.

POSTULATE 3. Envelopment of Observed Data. *For each observation* $j = 1, \cdots n$, $y_j \leq g(\mathbf{x}_j)$.

[4] Each $\mathbf{x}_i$ is evidently bounded below by 0. It may be assumed in addition that $\mathbf{x}_i \leq B_i$ where B_i are some large numbers representing physical limits on the availability of resources. If X is convex and compact and g: $X \to R$ is monotone and concave as assumed in Postulates 1 and 2, then $g(\cdot)$ is continuous on X.

POSTULATE 4. Minimum Extrapolation. *If a function* $\tilde{g}(\cdot)$ *satisfies Postulates* 1, 2 *and* 3, *then* $\tilde{g}(\mathbf{x}) \geq g(x)$ *for all* $\mathbf{x} \in X$.

The unique function $y = g(\mathbf{x})$ determined for $\mathbf{x} \in X^* \equiv \{\mathbf{x} \mid \mathbf{x} \geq \sum_{j=1}^n \lambda_j \mathbf{x}_j, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0\} \subseteq X$ by the above four postulates corresponds to that estimated by DEA.[5] In particular, the values $g(\mathbf{x})$ are obtained via a linear program described in the following:

PROPOSITION 1. *If* $\hat{y}_0 = \hat{g}(\mathbf{x}_0)$, $\mathbf{x}_0 \in X^*$ *and* $\hat{g}(\cdot)$ *satisfies Postulates* 1 *to* 4, *then* $\hat{y}_0 = y_0^*$ *where*

$$y_0^* = g^*(\mathbf{x}_0) = \operatorname{Max}\left\{ y \mid y = \sum_{j=1}^n \lambda_j y_j, \quad \sum_{j=1}^n \lambda_j \mathbf{x}_j \leq \mathbf{x}_0, \quad \sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0 \right\}. \quad (2)$$

PROOF. The proof proceeds by showing that the function $y_0^* = g^*(\mathbf{x}_0)$, obtained by solving the DEA linear program in (2) for each $\mathbf{x}_0 \in X^*$, is the *unique* function satisfying Postulates 1 to 4.

It is easy to see that $g^*(\mathbf{x}_0)$ satisfies Postulate 1. If $\mathbf{x}' \geq \mathbf{x}_0$ then $y_0^* = g^*(\mathbf{x}_0)$ is evidently a feasible solution for (2), and therefore, $g^*(\mathbf{x}') = \max y \geq g^*(\mathbf{x}_0)$.

Next, let λ_j', λ_j'', $j = 1, \cdots n$, be optimal solutions to (2) for $\mathbf{x}'$ and $\mathbf{x}''$, respectively. If $\mathbf{x}_0 = \theta \mathbf{x}' + (1 - \theta)\mathbf{x}''$ then $\lambda_j^0 = \theta \lambda_j' + (1 - \theta)\lambda_j''$, $j = 1, \cdots n$, and $y = \theta g^*(\mathbf{x}') + (1 - \theta) g^*(\mathbf{x}'')$ constitute a feasible solution to (2) for $\mathbf{x}_0$. It follows therefore that $g^*(\cdot)$ satisfies Postulate 2.

Furthermore, for any observation $(y_k, \mathbf{x}_k)$, $k = 1, \cdots n$, note that $\lambda_k = 1$, $\lambda_j = 0$ for all $j \neq k$, and $y = y_k$ constitute a feasible solution to the corresponding linear program in (2). Therefore, $y_k \leq \max y = g^*(\mathbf{x}_k)$ for each observation $k = 1, \cdots n$, and hence Postulate 3 is satisfied.

To see that $g^*(\cdot)$ satisfies Postulate 4, next consider optimal λ_j^*, $j = 1, \cdots n$, solving the program in (2). Then for any function $\hat{g}$: $X \to R$ satisfying Postulates 1, 2 and 3 it follows that

$$\hat{g}\left(\sum_{j=1}^n \lambda_j^* \mathbf{x}_j \right) \geq \sum_{j=1}^n \lambda_j^* \hat{g}(\mathbf{x}_j) \geq \sum_{j=1}^n \lambda_j^* y_j = g^*(\mathbf{x}_0).$$

[5] DEA does not determine $g(\mathbf{x})$ for values of $\mathbf{x}$ in X that are not in X^*

Further, since $\mathbf{x}_0 \geq \sum_{j=1}^{n} \lambda_j^* \mathbf{x}_j$ it follows that

$$\hat{g}(\mathbf{x}_0) \geq \hat{g}\left(\sum_{j=1}^{n} \lambda_j^* \mathbf{x}_j\right) \geq g^*(\mathbf{x}_0).$$

Finally, to show uniqueness, I note that if two functions $g^*(\cdot)$ and $\hat{g}(\cdot)$ satisfy Postulate 4 then $g^*(\cdot) \geq \hat{g}(\cdot)$ and $\hat{g}(\cdot) \geq g^*(\cdot)$ for all $\mathbf{x} \in X^*$. Therefore, $\hat{g}(\cdot) \equiv g^*(\cdot)$ on X^*. □

This proposition is particularly important because it suggests why DEA estimators may maximize likelihood. Given the set of observations $(y_l, \mathbf{x}_l)$, $l = 1, \cdots n$, the production frontier value $g^*(\mathbf{x}_j)$ for any observation j, estimated using the DEA model in (2), is determined *independent* of the $g^*(\mathbf{x}_k)$ values estimated for all of the other observations $k \neq j$. That is, Proposition 1 shows that as in Postulate 4, DEA minimizes $\epsilon_j = g^*(\mathbf{x}_j) - y_j$ for each observation $j = 1, \cdots n$, subject to $g^*(\cdot)$ being monotone increasing (Postulate 1) and concave (Postulate 2), and $g^*(\mathbf{x}_l) - y_l \geq 0$ for all $l = 1, \cdots n$ (Postulate 3); this minimization for ϵ_j is *independent* of the minimization of *all* other deviation terms ϵ_k, $k \neq j$. Note that this contrasts with the Aigner-Chu model in which all the ϵ_l, $l = 1, \cdots n$, are determined simultaneously along with the parameters β_i specifying the production function $g(\cdot)$. This suggests that if we postulate a probability density function $f(\epsilon)$ for the deviations ϵ, such that $f(\cdot)$ is monotone decreasing, then the DEA estimators minimizing ϵ_j will also maximize the likelihood function $\mathcal{L} = \prod_{j=1}^{n} f(\epsilon_j = g(\mathbf{x}_j) - y_j)$. This observation motivates the following two postulates for the density function and leads to the formal analysis described in the next section:

POSTULATE 3A. Envelopment. *Efficiency deviations ϵ_j are independently and identically distributed with probability density function $f(\epsilon)$ such that $f(\epsilon) = 0$ for all $\epsilon < 0$.*

POSTULATE 4A. Monotonicity of Density Function. *If $0 \leq \epsilon' \leq \epsilon''$ then $f(\epsilon') \geq f(\epsilon'')$.*

3. Maximum Likelihood Estimation

I consider the MLE problem specified in (1) to interpret DEA estimators of production frontier values as maximum likelihood estimators under specified conditions. Note that in writing the likelihood function as $\mathcal{L} \equiv \prod_{j=1}^{n} f(\epsilon_j = g(\mathbf{x}_j) - y_j)$, it is assumed that ϵ_j are distributed identically, independent of each other and independent of $\mathbf{x}_j$, see for instance Zellner et al. (1966). The following proposition shows that the MLE problem in (1) is solved by simply obtaining the optimal solutions to the DEA models in (2) corresponding to each observation $j = 1, \cdots n$.

PROPOSITION 2. *If the probability density function $f(\epsilon)$ satisfies Postulates 3A and 4A, and $\mathbf{x}_j$ and ϵ_j are independently distributed, then the optimal solutions $y_j^* = g^*(\mathbf{x}_j)$ solving (2) for $j = 1, \cdots n$, and $\epsilon_j^* = g^*(\mathbf{x}_j) - y_j$ solve the MLE problem in (1). The piecewise-linear function estimated by solving (2) is a maximum likelihood estimate of the production function $g(\cdot)$ in (1) for all $\mathbf{x} \in X^*$.*

PROOF. The proof follows directly from Proposition 1. Any function $\hat{g}: X \to R$ solving (1), satisfies Postulates 1 and 2 by virtue of constraints in (1.1). Further, (1.2) implies $y_j \leq \hat{g}(\mathbf{x}_j)$ for $j = 1, \cdots n$, and therefore, $\hat{g}(\cdot)$ also satisfies Postulate 3. It follows from Postulate 4 then that $g^*(\mathbf{x}_j) \leq \hat{g}(\mathbf{x}_j)$ for $j = 1, \cdots n$. Therefore, for any feasible solution $\{\hat{\epsilon}_j, \hat{g}(\cdot)\}$ to (1), $\epsilon_j^* = g^*(\mathbf{x}_j) - y_j \leq \hat{g}(\mathbf{x}_j) - y_j = \hat{\epsilon}_j$ for $j = 1, \cdots n$. Since $f(\epsilon)$ is monotone decreasing, it follows that $\prod_{j=1}^{n} f(\epsilon_j^*) \geq \prod_{j=1}^{n} f(\hat{\epsilon}_j)$. That is, $g^*(\cdot)$ and $\epsilon_j^* = g^*(\mathbf{x}_j) - y_j$ constitute an optimal solution to (1) for all probability density functions $f(\epsilon)$ satisfying Postulate 4A. □

Thus, maximum likelihood estimation under constraints (1.1) and (1.2) requires solving the DEA problem in (2).[6] The intuition underlying Proposition 2 stems from Postulate 4 which emphasizes that DEA estimators of ϵ_j^* are relative to the *tightest* monotone increasing and concave production frontier enveloping the data. In other words, DEA estimators minimize the deviation $\epsilon_j^* = g^*(\mathbf{x}_j) - y_j$ for each observation $j = 1, \cdots n$. If $f(\epsilon)$ is monotone decreasing then the likelihood function is maximized by the smallest values of ϵ_j consistent with the constraints characterizing the production frontier, and therefore, the DEA estimators solve the MLE problem. The DEA estimators can thus be justified on the basis of the common statistical criterion of maximizing the likelihood of obtaining the actual sample of obser-

[6] Any other monotone increasing and concave function passing through $g^*(\mathbf{x}_j)$ for all j is also a maximum likelihood estimate

vations, if deviation of the actual output from the efficient output is considered to be a stochastic variable with a monotone decreasing probability density function. A broad class of distributions, including distributions required to justify Aigner-Chu models on the basis of the maximum likelihood criterion, exhibit monotone decreasing density functions. Thus, the employment of the maximum likelihood criterion to justify DEA estimation is more robust than a similar use of the criterion to justify parametric production frontier estimation models as in Schmidt (1976).

Monotone decreasing probability density function is identified in Proposition 2 as only a sufficient condition for DEA estimators to be maximum likelihood estimators. It is evident that for any given sample of observations the DEA estimators may maximize the likelihood function even when $f(\epsilon)$ is not monotone decreasing. However, $f(\epsilon)$ being monotone decreasing is a necessary condition in the sense of the following:

PROPOSITION 3. *If the DEA estimators $g^*(\mathbf{x})$ and $\epsilon_j^* = g^*(\mathbf{x}_j) - y_j$ solve the MLE problem in (1) for* all *possible sample observations (y_j, x_j), $j = 1, \cdots n$, and if ϵ are i.i.d. with a probability density function $f(\epsilon)$ such that $f(\epsilon'') > 0$ for some ϵ'' implies $f(\epsilon) > 0$ for all $\epsilon > 0$, $\epsilon < \epsilon''$, then $f(\epsilon)$ is monotone decreasing.*

PROOF. The proof proceeds by showing that for any $f(\epsilon)$ that is not monotone decreasing, we can construct as a counterexample, a sample of observations for which DEA estimators are not MLE.

Suppose $f(\epsilon)$ is such that for some $0 \le \epsilon' < \epsilon''$, $0 \le f(\epsilon') < f(\epsilon'')$ and $f(\epsilon) > 0$ for all ϵ such that $0 < \epsilon < \epsilon''$. Then consider the following sample of $(K + 2)$ observations from a one-input one-output production model: (y_A, x_A), (y_B, x_B) and (y_k, x_k), $k = 1, \ldots, K$ with $x_k = \lambda_k x_A + (1 - \lambda_k)x_B$, $y_k = \lambda_k y_A + (1 - \lambda_k)y_B - \epsilon'$, $\lambda_k = k/(K + 1)$, $y_A < y_B$, $x_A < x_B$. Evidently, the DEA production frontier for this sample is the line segment joining (y_A, x_A) and (y_B, x_B), and the corresponding log-likelihood function is $\log \mathcal{L} = 2 \log f(0) + K \log f(\epsilon')$. Next consider alternative production frontier values $\hat{y}_A = y_A + \epsilon'' - \epsilon'$, $\hat{y}_B = y_B + \epsilon'' - \epsilon'$ and $\hat{y}_k = y_k + \epsilon''$, $k = 1, \ldots, K$. The corresponding log-likelihood function is

$$\log \mathcal{L}' = 2 \log f(\epsilon'' - \epsilon') + K \log f(\epsilon'').$$

Therefore, for

$$K > \max \{1, 2[\log f(0) - \log f(\epsilon'' - \epsilon')]/[\log f(\epsilon'') - \log f(\epsilon')]\}$$

we have $\log \mathcal{L}' > \log \mathcal{L}$; that is DEA estimators are *not* MLE. □

4. Bias and Consistency

Results for the unidimensional extreme value estimation problem provide an interesting comparison to the production frontier estimation problem considered here. When the upper support of the distribution of a stochastic variable y is the parameter to be estimated, Tate (1959) shows that $y^* = \max \{y_j | j = 1, \cdots n\}$ is the MLE, the MLE is biased and the bias is inversely proportional to the sample size. For the single output multiple input production frontier estimation problem also, Postulate 4 implies that the DEA estimator y_0^* is not greater than the true functional value $g(\mathbf{x}_0)$. Let $\delta_0 = g(\mathbf{x}_0) - y_0^* \ge 0$ measure the "inaccuracy" in the DEA estimation of the "parameter" $g(\mathbf{x}_0)$.

PROPOSITION 4. *If ϵ are i.i.d. then for any $\Delta > 0$, $\Pr\{\delta_0 > \Delta\} \ge [1 - F(\Delta)]^n$, where $F(\Delta) = \int_0^\Delta f(\epsilon)\, d\epsilon$ and n is the sample size.*

PROOF. For a sample of size n, the probability that any realized value of ϵ_j, $j = 1, \cdots n$, is greater than Δ is simply $[1 - F(\Delta)]$. Let $\epsilon_{\min} = \min_j \{\epsilon_j | j = 1, \cdots n\}$. Then, $\Pr\{\epsilon_{\min} > \Delta\} = [1 - F(\Delta)]^n$ since ϵ_j are i.i.d.

Let λ_j^* be optimal values of λ_j in (2). Then $y_0^* = \sum_{j=1}^n \lambda_j^* y_j = \sum_{j=1}^n \lambda_j^*(g(\mathbf{x}_j) - \epsilon_j) \le g(\mathbf{x}_0) - \epsilon_{\min}$ since $g(\cdot)$ is a concave function. Therefore, $\delta_0 = g(\mathbf{x}_0) - y_0^* \ge \epsilon_{\min}$, and

$$\Pr\{\delta_0 > \Delta\} \ge \Pr\{\epsilon_{\min} > \Delta\} = [1 - F(\Delta)]^n. \quad \square$$

COROLLARY. *If $F(0) < 1$ then the DEA estimator is biased.*

PROOF. Obvious, since

$$g(\mathbf{x}_0) - E(y_0^*) = E(\delta_0) = \int_0^\infty \delta_0 d \Pr\{\delta_0\} > 0$$

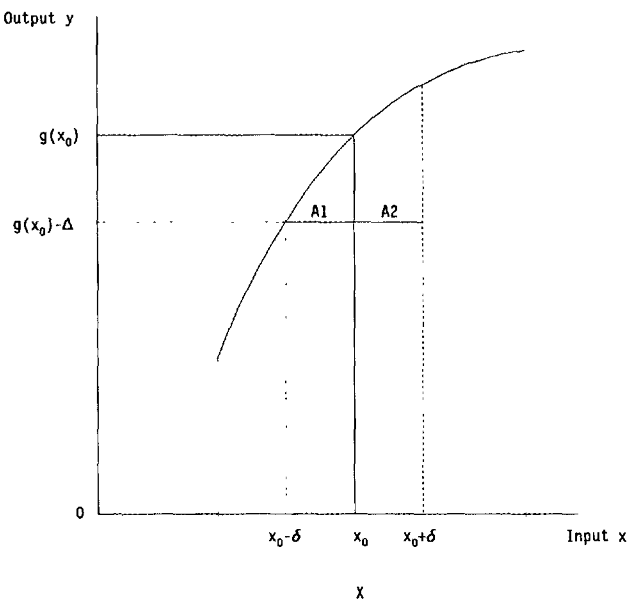
if $F(0) < 1$. □

Thus, the DEA estimators will almost always be biased. However, since these are MLE, one may suspect that they are consistent. But, in the case of the Aigner-Chu type parametric production frontier models, Schmidt (1976) points out that the statistical properties of the MLE are not obvious. The usual asymptotic properties, such as consistency, do not necessarily follow because one of the regularity conditions in the standard proofs (e.g., Amemiya (1973), Barnett (1976)) is violated. Specifically, the range of the random variable y_j in this case is $(-\infty, g^*(\mathbf{x}_j)]$, which is dependent on the estimated parameters $y_j^* = g^*(\mathbf{x}_j)$. Therefore, using Wald's (1949) consistency proof which requires less stringent regularity conditions, Greene (1980) identified $f(\epsilon = 0) = 0$ and $[\partial f(\epsilon = 0)]/\partial\epsilon = 0$ as *sufficient* conditions for the MLE to be consistent, asymptotically efficient and asymptotically normally distributed for the parametric production frontier models. However, $f(\epsilon = 0) = 0$ *and* $\epsilon \geq 0$ are evidently inconsistent with $f(\epsilon)$ being monotone decreasing. But, in any case, an additional problem with infinitely many incidental parameters, as pointed out by Neyman and Scott (1948), arises in the case of the nonparametric production frontier models because the DEA production frontier is parameterized by $y_j^* = g^*(\mathbf{x}_j)$ themselves, and their number increases with the sample size. This renders the proofs by Wald (1949) and Greene (1980) also inapplicable. An alternative direct proof (from first principles) is provided below to establish the consistency of the DEA estimators.

PROPOSITION 5. *If the production frontier $g(\mathbf{x})$ is monotone increasing and concave for $\mathbf{x} \in X$, where X is a convex and compact subset of R^m, and if $\mathbf{x}$ and ϵ are independently distributed with probability density functions $h(\cdot)$ and $f(\cdot)$ such that $h(\mathbf{x}) > 0$ for all $\mathbf{x} \in X$, $f(\epsilon) = 0$ for $\epsilon < 0$, and $F(\epsilon) = \int_{-\infty}^{\epsilon} f(\epsilon)\, d\epsilon > 0$ for all $\epsilon > 0$, then the DEA estimators $g^*(\mathbf{x})$ are weakly consistent for all $\mathbf{x}$ in the interior of X.*

PROOF. Consider first the case of a single input; see Figure 1. Since $g(\cdot)$ is continuous, for any x_0 in the interior of X and arbitrary $\Delta > 0$, there exists $\delta > 0$ such that $g(x) > g(x_0) - \Delta$ for all $x \in (x_0 - \delta, x_0 + \delta) \subseteq X$. I shall show that for sufficiently large samples (of size n say) the probability that "the difference between the true functional value $g(x_0)$ and the DEA estimator ($g_n^*(x_0)$ say) is greater than Δ" is less than any arbitrarily small positive number.

Figure 1 **Consistency of DEA Estimators**



Consider a randomly drawn observation (y, x) where $y = g(x) - \epsilon$, and x in the interior of X and $\epsilon \geq 0$ are realized values of the two independent random variables. Let

$$p_1 \equiv \Pr\{\text{Event A1}\} \equiv \Pr\{x \in (x_0 - \delta, x_0) \text{ and } y = g(x) - \epsilon > g(x_0) - \Delta\}.$$

Note that

$$\Pr\{x \in (x_0 - \delta, x_0)\} = \int_{x_0-\delta}^{x_0} h(x)\, dx > 0.$$

Furthermore, since $g(x) > g(x_0) - \Delta$ given $x \in (x_0 - \delta, x_0)$, and $F(\epsilon) > 0$ for all $\epsilon > 0$, it follows that

$$\Pr\{\epsilon < g(x) - g(x_0) + \Delta\} > 0.$$

Therefore $p_1 > 0$, as x and ϵ are independently distributed. Similarly, it can be shown that

$$p_2 \equiv \Pr\{\text{Event A2}\} \equiv \Pr\{x \in (x_0, x_0 + \delta) \text{ and } y = g(x) - \epsilon > g(x_0) - \Delta\} > 0.$$

Consider next a sample of n independent observations. Clearly, $\Pr\{$Event A1 does not occur for any of the n observations$\} = (1 - p_1)^n$ and $\Pr\{$Event A2 does not occur for any of the n observations$\} = (1 - p_2)^n$.

Suppose *both* Event A1 and Event A2 occur for at least one observation each; that is, the sample includes two observations, say (y_1, x_1) and (y_2, x_2), such that $x_1 \in (x_0 - \delta, x_0)$, $x_2 \in (x_0, x_0 + \delta)$, and both y_1 and y_2 are greater than $g(x_0) - \Delta$. In such a case, the DEA estimator $g_n^*(x_0)$, for this sample of size n, is at least as large as the minimum of y_1 and y_2. That is, $g_n^*(x_0) \geq \min\{y_1, y_2\} > g(x_0) - \Delta$, and hence $g(x_0) - g_n^*(x_0) < \Delta$.

Furthermore, since $g(\cdot)$ is monotone increasing and concave and $f(\epsilon) = 0$ for $\epsilon < 0$, it follows from Postulate 4 that the DEA estimator $g_n^*(x_0) \leq g(x_0)$ for any sample, and hence $|g(x_0) - g_n^*(x_0)| = g(x_0) - g_n^*(x_0)$. Therefore,

$$\begin{aligned}
&\Pr\{|g(x_0) - g_n^*(x_0)| > \Delta\} \\
&\quad = \Pr\{g(x_0) - g_n^*(x_0) > \Delta\} \\
&\quad \leq \Pr\{\text{Event A1 or Event A2 does not occur for all observations in the sample}\} \\
&\quad \leq (1 - p_1)^n + (1 - p_2)^n
\end{aligned}$$

and hence

$$\lim_{n\to\infty} \Pr\{|g(x_0) - g_n^*(x_0)| > \Delta\} = 0.$$

Extension to the multiple input case is obtained directly by considering an open ball with radius δ such that $g(\mathbf{x}) > g(\mathbf{x}_0) - \Delta$ for all $\mathbf{x}$ with $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, and noting that there is a positive probability that an observation $(y, \mathbf{x})$ will be such that $\mathbf{x}$ is in a specific orthant (relative to $\mathbf{x}_0$) of the open ball with $y > g(\mathbf{x}_0) - \Delta$. □

Note that Proposition 5 does not require any structure on the probability distribution $F(\epsilon)$, except that $F(\epsilon) > 0$ for all $\epsilon > 0$. In particular, the density function $f(\epsilon)$ need *not* be monotone decreasing. Furthermore, the distribution of ϵ may be a *mixture* of different distributions of ϵ for different sets of observations, identified by an attribute independent of $\mathbf{x}$. The consistency result holds also when ϵ_j are not independently distributed, provided the conditional probability of obtaining small values of ϵ continues to be strictly positive given any realized values of other ϵ_j, and that this positive probability does not vanish asymptotically. These observations will be particularly useful in constructing hypothesis tests in the next section.

5. Hypothesis Tests

Hypotheses that have commonly been of interest in efficiency analysis involve a comparison of two groups of decision making units (DMUs) to assess whether one group is more efficient than the other. For instance, Charnes et al. (1981) compare the efficiency of DMUs with and without Program Follow Through, Banker et al. (1987) compare the labor productivity before and after the implementation of a gain-sharing program, Banker et al. (1991) compare software programmer productivity for projects with and without a structured systems development methodology, and Banker et al. (1990) compare the efficiency of fast food outlets where a specific information system was installed with those that did not have the system installed. The consistency result obtained in the previous section will be employed here to suggest possible asymptotic tests in DEA to evaluate such hypotheses where one group of DMUs is compared with another.

Since the DEA estimators $g_n^*(\mathbf{x})$ are consistent and $g_n^*(\mathbf{x}) \leq g(\mathbf{x})$, it follows that for any $\Delta > 0$ and a given $(y_j, \mathbf{x}_j)$,

$$\lim_{n\to\infty} \Pr\{\epsilon_j - \epsilon_j^{*(n)} > \Delta\} = 0,$$

where $\epsilon_j^{*(n)} = g_n^*(\mathbf{x}_j) - y_j$ is the DEA estimator of the inefficiency deviation ϵ_j for an observation j in a sample of size n.

PROPOSITION 6. *If M is a specific set of m DMUs included in a sample of n observations, the asymptotic distribution of the DEA estimators $\epsilon_j^{*(n)}$, $j \in M$, is identical to the true distribution of ϵ_j.*

PROOF. For each $j \in M$,

$$\epsilon_j^{*(n)} = g_n^*(\mathbf{x}_j) - y_j \leq g(\mathbf{x}_j) - y_j = \epsilon_j.$$

Therefore, for any given constants $E_j \in R$, $\Pr\{\epsilon_j^{*(n)} \leq E_j$ for all $j \in M\} = \Pr\{\epsilon_j \leq E_j$ for all $j \in M\} + \Pr\{\epsilon_j^{*(n)} \leq E_j$ for all $j \in M$, *and* $E_j < \epsilon_j$ for some $j \in M\}$. But $\lim_{n\to\infty} \Pr\{\epsilon_j^{*(n)} \leq E_j$ for all $j \in M$, *and* E_j

$< \epsilon_j$ for some $j \in M\} = 0$ because $\lim_{n\to\infty} \Pr\{\epsilon_j^{*(n)} < \epsilon_j - \Delta\} = 0$ for any arbitrary $\Delta > 0$. Therefore,

$$\lim_{n\to\infty} \Pr\{\epsilon_j^{*(n)} \le E_j \quad \text{for all} \quad j \in M\}$$
$$= \Pr\{\epsilon_j \le E_j \quad \text{for all} \quad j \in M\}. \quad \square$$

Thus, for "large" samples the DEA estimators of the inefficiency deviations ϵ_j^* for a given set of DMUs follow the same probability distribution as the true ϵ_j. In particular, if ϵ_j are i.i.d. half-normal, then the DEA estimators ϵ_j^* will also be so distributed.[7] The Central Limit Theorem cannot be employed directly to construct hypothesis tests, however, because the parameters representing the mean and variance of ϵ are unknown. Some distributional assumptions must be made to estimate these parameters and construct statistical tests.

If an exponential distribution is assumed for ϵ as in Schmidt (1976) then the estimation of the mean and variance parameters can be avoided for the hypothesis test. Let the sample consist of n independently distributed observations, where n is large. Let m_i, $i = 1, 2$, of these observations have their inefficiency deviations distributed exponentially[8] with parameter σ_i, $i = 1, 2$. These two sets of observations will be referred to as M_i, $i = 1, 2$. Under this maintained assumption, the distribution of $\sum_{j\in M} \epsilon_j^* / \sigma_i$ is chi-square with $2m_i$ degrees of freedom (e.g., Ellis (1844)). Since the DEA estimators ϵ_j^* are asymptotically distributed as ϵ, the statistic

$$\left[\sum_{j\in M_1} \epsilon_j^* / m_1\right] \Big/ \left[\sum_{j\in M_2} \epsilon_j^* / m_2\right]$$

is asymptotically (for large n) the ratio of two chi-square variates under the null hypothesis H_0: $\sigma_1 = \sigma_2$ (with the alternative hypothesis H_1: $\sigma_1 > \sigma_2$). Therefore, a simple ratio of the average inefficiency deviations can serve here as the test statistic because the ratio of two chi-square variates with $2m_i$, $i = 1, 2$, degrees of freedom, obeys the F-distribution with $(2m_1, 2m_2)$ degrees of freedom.

If the inefficiency deviations for the observations in subset M_i, $i = 1, 2$, are assumed to follow the half-normal distribution with parameters σ_i, then the sum $\sum_{j\in M_i} (\epsilon_j / \sigma_i)^2$ is chi-square with m_i degrees of freedom. Therefore, under the null hypothesis H_0: $\sigma_1 = \sigma_2$ the statistic

$$\left[\sum_{j\in M_1} (\epsilon_j^*)^2 / m_1\right] \Big/ \left[\sum_{j\in M_2} (\epsilon_j^*)^2 / m_2\right]$$

is expected to asymptotically obey the F-distribution with (m_1, m_2) degrees of freedom. Thus, the ratio of the average of *squares* of inefficiency deviations is the appropriate test statistic instead of the ratio of the average of simply the inefficiency deviations when the exponential distribution is assumed.

In either case, the cumulative probability that inefficiency deviations are less than a specified value is greater for the M_2-type observations than that for the M_1-type observations if and only if $\sigma_1 > \sigma_2$.[9] The observations of the type in M_2 are more efficient than those in M_1 when $F_2(\epsilon) > F_1(\epsilon)$ for all $\epsilon > 0$, where F_1 and F_2 are the distributions of inefficiency deviations in M_1 and M_2, respectively. Therefore, if no parametric assumptions are maintained about the probability distribution $F(\cdot)$, then Kolmogorov-Smirnov type of nonparametric tests may be employed instead of the above parametric tests.

All of the above tests and results are applicable only to large samples (that is, n should be large), although the two sets of DMUs compared in the hypothesis tests need not be large (that is, $m_1 + m_2 \le n$). As in recent econometric work dealing with parametric production functions, if these tests are employed for relatively small samples then the results should be interpreted very cautiously, at least until systematic evidence is obtained from Monte Carlo experimentation with finite samples of varying sizes.

6. Concluding Remarks

Data envelopment analysis was developed as a mathematical programming methodology for efficiency evaluation, and as such it does not require any structure to *compute* relative efficiency scores for DMUs engaged in similar production processes. The present paper has shown, however, that by imposing suitable structure,

[7] While the ϵ^* will not be mutually independent for small samples, they approach independence as n gets large. See, for instance, Tate and Brown (1970) for a similar point regarding the Cochran Q test

[8] No such assumption is required about the distribution of the remaining $n - m_1 - m_2$ observations

[9] Estimates for σ_i may be provided by $\sum_{j\in M_i} \epsilon_j^* / m_i$ for the exponential case, and by $[\sum_{j\in M_i} (\epsilon_j^*)^2 / m_i]^{1/2}$ in the half-normal case

the efficiency estimates obtained using DEA can be provided with interesting statistical properties.

First, it was formally proved in Proposition 1 that the Banker-Charnes-Cooper postulates determine a unique production function which is the same as that estimated by DEA. Second, Propositions 2 and 3 showed that if the inefficiency deviations are treated as stochastic variables then a monotone decreasing probability density function for these deviations is a sufficient and (almost) necessary condition for DEA inefficiency estimators to maximize the likelihood of obtaining the actual sample of observations. This demonstrates the robustness of DEA relative to the Aigner-Chu parametric production frontier estimation method which is MLE only if the inefficiency distribution is exponential or half-normal.

Third, it was proved in Proposition 5 that the DEA production frontier and inefficiency estimators are weakly consistent; in large samples the expected value of the DEA estimator is almost certainly the true parameter value. This result is important as it implies in turn (as shown in Proposition 6) that for any given set of DMUs, the asymptotic distribution of the DEA inefficiency estimators is the same as their true distribution. Finally, these asymptotic properties were employed to suggest statistical tests of hypotheses using the DEA inefficiency estimators. These hypotheses tests should be particularly useful for empirical researchers seeking formal means to test whether one type of DMUs is more efficient than another.[10]

[10] This research was supported in part by NSF Grant No. SES-8709044.

References

Aigner, D. J and S. F. Chu, "On Estimating the Industry Production Function," *American Economic Rev.* (September 1968), 826–839.

——, C. A. K. Lovell and P. J. Schmidt, "Formulation and Estimation of Stochastic Frontier Production Function Models," *J. Econometrics* (July 1977), 21–37.

Amemiya, T., "Regression Analysis When the Dependent Variable Is Truncated Normal," *Econometrica* (November 1973), 997–1016.

Banker, R. D., *Studies in Cost Allocation and Efficiency Evaluation*, Unpublished Doctoral Thesis, Harvard University, Graduate School of Business Administration, 1980

——, "Estimating Most Productive Scale Size Using Data Envelopment Analysis," *European J. Operational Res* (July 1984), 35–44.

——, "Stochastic Data Envelopment Analysis," Mimeo, 1986.

——, A. Charnes and W. W. Cooper, "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis," *Management Sci* (September 1984), 1078–1092.

——, S Datar and C. Kemerer, "A Model to Evaluate Variables Impacting the Productivity of Software Maintenance Project," *Management Sci.* (January 1991), 1–18.

——, —— and M. Rajan, "Measurement of Productivity Improvements. An Empirical Analysis," *J Accounting, Auditing and Finance* (Fall 1987), 319–354

——, R. J. Kauffman and R C Morey, "Measuring Gains in Operational Efficiency from Information Technology," *J. Management Information Systems* (Fall 1990), 29–54

—— and A. Maindiratta, "Nonparametric Analysis of Technical and Allocative Efficiencies in Production," *Econometrica* (November 1988), 1315–1332

—— and ——, "Maximum Likelihood Estimation of Monotone and Concave Production Functions," *J. Productivity Analysis* (1992).

Barnett, W. A., "Maximum Likelihood and Iterated Aitken Estimation of Nonlinear Systems of Equations," *J. American Statistical Association* (1976), 354–360.

Charnes, A., W. W. Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European J. Operational Res* (1978), 429–444

——, —— and ——, "Program Evaluation and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through," *Management Sci.* (1981), 668–697.

Ellis, R. L., "On a Question in the Theory of Probabilities," *Cambridge Mathematical J.* (1844), 127–132.

Farrell, M. J, "The Measurement of Productive Efficiency," *J. Royal Statistical Society*, Series A (1957), 253–290.

Forsund, F R., C. A K Lovell and P. Schmidt, "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement," *J Econometrics* (1980), 5–25

Greene, W. H, "Maximum Likelihood Estimation of Econometric Frontier Functions," *J. Econometrics* (1980), 27–56

Meeusen, W. and J. van den Broeck, "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error," *International Economic Rev.* (June 1977), 435–444.

Neyman, J and E L. Scott, "Consistent Estimates Based on Partially Consistent Observation," *Econometrica* (1948), 1–32.

Schmidt, P., "On the Statistical Estimation of Parametric Frontier Production Functions," *Rev. Economics and Statistics* (May 1976), 238–239.

——, "On the Statistical Estimation of Parametric Frontier Production Functions: Rejoinder," *Rev. Economics and Statistics* (August 1978), 481–482.

——, "Frontier Production Functions," *Econometric Rev.* (1985).

Tate, R F., "Unbiased Estimation: Functions of Location and Scale Parameters," *Annals of Mathematical Statistics* (1959)

Tate, M W. and S. M. Brown, "Note on the Cochran Q Test," *J. American Statistical Association* (1970), 155–160.

Wald, A., "Note on the Consistency of the Maximum Likelihood Estimate," *Annals of Mathematical Statistics* (1949).

Zellner, A, J. Kmenta and J. Dreze, "Specification and Estimation of Cobb-Douglas Production Function Models," *Econometrica* (October 1966), 784–795.

Accepted by Richard M. Burton, received December 21, 1987. This paper has been with the author 30 months for 2 revisions.