Selection of efficiency evaluation models

RAJIV D. BANKER *University of Minnesota*

Abstract. The purpose of this paper is to emphasize the importance of selecting the efficiency evaluation model that best represents the known and postulated relation between inputs and outputs. By ignoring the input/output relationship, Mehrez, Brown, and Khouja (MBK) misspecify the efficiency evaluation model and misinterpret the distortions resulting from their misspecification as being paradoxical. This paper shows that the use of the appropriate data envelopment analysis (DEA) model reflecting the matched output/input (MOI) technology postulated by MBK will always provide efficiency estimates that satisfy their MOI axiom. Statistical consistency of the DEA estimator also implies that even with MBK's use of a misspecified model, the frequency of violation of their MOI axiom is likely to be small for large samples.

Résumé. L'auteur insiste sur l'importance du choix du modèle de l'évaluation du rendement qui représente le mieux la relation connue et postulée entre les intrants et les extrants. En ignorant la relation intrant-extrant, Mehrez, Brown et Khouja (MBK) définissent mal le modèle d'évaluation du rendement et jugent, à tort, les distorsions résultant de cette définition comme étant paradoxales. L'auteur démontre que le recours à l'analyse intégrale appropriée des données s'inspirant de la technologie de concordance extrant-intrant postulée par MBK produit toujours des estimations du rendement qui respectent l'axiome de concordance. La cohérence statistique des estimations issues de l'analyse intégrale des données donne également à penser que même lorsqu'on utilise un modèle mal défini comme le font MBK, la fréquence des dérogations à l'axiome de concordance extrant-intrant tend à être faible pour les gros échantillons.

Introduction

Data envelopment analysis (DEA) is an efficiency evaluation and production function estimation methodology used by researchers and practitioners for over a decade. As with other econometric and management science techniques, we find cases in which the method is misused or wrongly applied. Egregious among such erroneous applications are uses of these methods that ignore the assumptions underlying the model. The paper by Mehrez, Brown, and Khouja (MBK) (pp. 329–342) represents such a case because it is founded on the flawed assumption that "DEA applies to *any* set of organizations *without regard* to the input/output relationships within each organization" (emphasis added).

Contemporary Accounting Research, Vol. 9, No. 1 (Fall 1992) pp. 343–355 ©CAAA

DEA is a very flexible methodology because it comprises a large (and growing) number of models, each designed for the production correspondence postulated to represent the conversion of inputs into outputs in an organization. In any application, it is necessary for the researcher to devote considerable time and effort to understand the production process and choose the appropriate model. Simply observing "the inputs and outputs and nothing else" and crunching the data without further thought is irresponsible. It is well-known that misspecifying a linear relation as multiplicative (or vice versa), omitting an explanatory variable, or estimating a system of simultaneous equations as independent equations, for instance, may result in biased estimates of the parameters and/or their variances in traditional econometric analysis. In a similar vein, specifying a constant return to scale model when scale economies or diseconomies likely prevail, or postulating substitutability between inputs when the production process is separable (or vice versa) can result in incorrect estimates of efficiency in DEA.

This paper shows that what appears to be a *paradox* to MBK is simply a *misspecification* of the efficiency evaluation model. Had they employed the DEA model appropriate for their *matched output/input* (MOI) technology, then the efficiency measures would not have violated their MOI axiom. To show how the appropriate efficiency evaluation model may be selected, I begin in the next section with a discussion of the assumptions underlying some basic DEA models. In the third section, I describe the model, and the corresponding aggregate efficiency measure, that is appropriate for the matched output/input technology. Finally, I present results of a simulation study to examine the frequency of violation of MBK's MOI axiom as the sample size is increased.

Postulates underlying basic DEA models

Efficiency evaluation models in DEA assume availability of data on inputs $x_{id} \geq 0$, $i \in I \equiv \{1, \ldots \psi\}$ and outputs $y_{\theta d} \geq 0$, $\theta \in \Theta \equiv \{1, \ldots \eta\}$, for each decision making unit (DMU) $d \in D \equiv \{1, \ldots \xi\}$.[1] A production (possibility) set T is defined to be $T \equiv \{(\mathbf{x}, \mathbf{y}) | \text{outputs } \mathbf{y} \equiv (y_1, \ldots y_\eta) \text{ can be produced from inputs } \mathbf{x} \equiv (x_1, \ldots x_\psi)\}$. A production possibility $(\mathbf{x}, \mathbf{y})$ in a set T is efficient if there does not exist any $(\mathbf{x}', \mathbf{y}') \in T$, with $\mathbf{x}' \leq \mathbf{x}$, $\mathbf{y}' \geq \mathbf{y}$ and a strict inequality for at least one input or output. An efficiency measure for a DMU d based on Shephard's (1970) distance function is

$$\rho_d = \inf\{\rho | (\rho \mathbf{x}_d, \mathbf{y}_d) \in T\}. \tag{1}$$

DMU d is inefficient if $\rho_d < 1$.[2] Farrell (1957) and Banker, Charnes, and Cooper (1984) also suggest efficiency measures based on Shephard's distance function.

1 This notation is used to maintain consistency with the MBK paper, although it is not consistent with the prior DEA literature.

2 Clearly, $\rho_d \leq 1$. DMU d is not necessarily efficient if $\rho_d = 1$ as it does not preclude the existence of slacks s_{id}^-, $i \in I$, and $s_{\theta d}^+$, $\theta \in \Theta$, with at least one of s_{id}^-, $s_{\theta d}^+$ strictly positive and $(x_d - s_d^-, y_d + s_d^+) \in T$. See Banker, Charnes, and Cooper (1984).

The constant returns to scale matched output/input (CRS–MOI) technology considered by MBK can be represented as

$$T_M \equiv \{(\mathbf{x}, \mathbf{y}) | y_i \leq m_i x_i,\ m_i > 0,\ i \in \mathbf{I}\}. \tag{2}$$

Here, $m_i > 0$ is the maximum output possible from one unit of input i. It is evident that the efficiency measure based on Shephard's distance function is given in this case by

$$\rho_d = \max_{i \in I}\{y_{id}/m_i x_{id}\}. \tag{3}$$

MBK postulate that any aggregate efficiency measure must satisfy their matched output/input axiom (MOI axiom):

If $y_{ij}/x_{ij} > y_{ik}/x_{ik}$ *for every* $i \in I$ *for two DMUs* $j,\ k \in D$, *then* $\rho_j > \rho_k$.

It can be verified quite easily that the measure in (3) satisfies the MOI axiom.

In empirical applications, however, m_i is not known and must be estimated from observed data based on prior knowledge of the nature (form) of the production correspondence. DEA provides such an empirically oriented methodology for estimating production efficiency from observed data. Contrary to MBK's assertion that "no ... information like how the inputs are combined to produce the outputs ... is needed to get (DEA) efficiency measures," DEA models, in fact, make explicit assumptions about the underlying technology that must be considered by a researcher before choosing a particular model for an application.

The DEA model presented by Banker, Charnes, and Cooper (BCC, 1984) imposes considerably less structure on the production set than the CRS–MOI technology implicitly assumed by MBK. In particular, BCC specify the following postulates for the estimated production set T:

Postulate 1: Convexity of production set
If $(\mathbf{x}_1, \mathbf{y}_1)$ and $(\mathbf{x}_2, \mathbf{y}_2) \in T$, then $(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2, \lambda \mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) \in T$ for all $\lambda \geq 0$.

Postulate 2: Monotonicity of production set
If $(\mathbf{x}_1, \mathbf{y}_1) \in T$ and $\mathbf{x}_2 \geq \mathbf{x}_1, \mathbf{y}_2 \leq \mathbf{y}_1$, then $(\mathbf{x}_2, \mathbf{y}_2) \in T$.

Postulate 3: Envelopment of observations
Each observed $(\mathbf{x}_d, \mathbf{y}_d) \in T$ for all $d \in D$.

Postulate 4: Minimum extrapolation
If a set T' satisfies the other specified postulates for the production set, then $T \subseteq T'$.

The first two postulates embody commonly invoked regularity properties for production sets. The third postulate ensures that the estimated production set includes all actual observations. The last postulate reflects the assumption that the earlier postulates specify *all* of the available knowledge about the input-output correspondence. The estimated set is the smallest set among *all* possible production sets that satisfy the properties embodied in the earlier postulates. If any additional characteristics are desired for the production set, then they must be specified explicitly to define the restricted population from which the smallest set is determined, else there is no assurance that this smallest set will exhibit the additional characteristics.

The unique production set T estimated on the basis of the above four DEA postulates[3] is then given by $\hat{T} = \{(\mathbf{x}, \mathbf{y}) | \hat{\rho}(\mathbf{x}, \mathbf{y}) \leq 1\}$ where

$$\hat{\rho}(\mathbf{x}, \mathbf{y}) = \min \rho \tag{4.0}$$

subject to

$$\Sigma_{d \in D} \lambda_d \mathbf{x}_d \leq \rho \mathbf{x} \tag{4.1}$$

$$\Sigma_{d \in D} \lambda_d \mathbf{y}_d \geq \mathbf{y} \tag{4.2}$$

$$\Sigma_{d \in D} \lambda_d = 1 \tag{4.3}$$

$$\rho, \lambda_d \geq 0. \tag{4.4}$$

The above linear program in (4) is referred to as the BCC model in the DEA literature. Here, $\hat{\rho}(\mathbf{x}, \mathbf{y})$ is the estimated Shephard's distance function value and may be employed as a basis for an efficiency measure.[4]

The BCC approach is deterministic, the estimated efficiency measure is equal to the true value if all four postulates are satisfied. In contrast, Banker (1992) presents a statistical approach that treats Shephard's distance (efficiency) measure for individual observations as a random variable. The links between DEA and traditional econometric estimation methods become transparent if a probability density function $f(\rho)$ is specified for the efficiency measure ρ, so that $f(\rho) \geq 0$ for $0 \leq \rho \leq 1$ and $\int_0^1 f(\rho) d\rho = 1$. Postulates 3 and 4 are then replaced by the following postulate.

Postulate 5: Likelihood of efficient observations
Probability density function $f(\rho) \geq 0$ for $0 \leq \rho \leq 1$, $\int_0^1 f(\rho) d\rho = 1$, is such that $\int_{1-\delta}^1 f(\rho) d\rho > 0$ for all $\delta > 0$.

The efficiency estimates provided by $\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j)$, $j \in D$, from the linear program in (4) are statistically consistent under postulates 1, 2 and 5. If postulates 5 is strengthened further by adding the following postulate 6, then $\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j)$, $j \in D$, are also maximum likelihood estimates.

Postulate 6: Increasing probability density
If $1 \geq \rho_1 \geq \rho_2 \geq 0$, then $f(\rho_1) \geq f(\rho_2)$.

Critical in the application of postulates 3 and 4, or 5 and 6, is the assumption that the only available information about the production set (i.e., the nature of

3 See Banker (1992) for a formal proof of the unique determination of the production set.
4 See Banker et al. (1984) for additional adjustments to distinguish efficient from inefficient boundary points.

the input/output relationship) is that embodied in postulates 1 and 2. In other words, the extrapolation in estimating the production set is minimized, or the likelihood maximized, subject *only* to the convexity and monotonicity postulates. If additional information (such as constant returns to scale or separability) about the production set is available or postulated, then the optimization such as that in (4) is performed subject to such additional structure.

Consider, for instance, additional knowledge that the production set exhibits constant returns to scale.

Postulate 7: Constant returns to scale
If $(\mathbf{x}, \mathbf{y}) \in T$ and $k > 0$ then $(k\mathbf{x}, k\mathbf{y}) \in T$.

Postulates 1, 2, 7, 3, and 4 determine a unique production set for which $\hat{\rho}(\mathbf{x}, \mathbf{y})$ is computed as before except that (4.0) is minimized subject only to (4.1), (4.2), and (4.4), the constraint in (4.3) is omitted; see BCC (1984). This yields the so-called CCR model in DEA, which should be employed when convexity, monotonicity, and constant returns to scale are the only stipulated properties for the production set. Note that postulate 7 is logically consistent with postulates 1, 2, and 3, but if the BCC production set based only on postulates 1, 2, 3, and 4 is estimated, where $\hat{\rho}(\mathbf{x}, \mathbf{y})$ minimizes (4.0) subject to (4.1), (4.2), (4.3), *and* (4.4), then the estimated production set need not (and usually will not) exhibit constant returns to scale. Therefore, DEA efficiency scores obtained relative to such an estimated BCC production set will not be the same as the true efficiency values if the observations are actually obtained from a constant returns to scale production set. Statistical consistency of the DEA estimator, however, assures us that the difference between the BCC and CCR efficiency values will vanish asymptotically in such a case as sample size (number of observations, ξ) is increased.

The CCR and the BCC models are the earliest, and hence the most commonly employed, models in DEA. Several other DEA models have been developed in the last decade to deal with situations such as increasing marginal product (Banker and Maindiratta, 1986), categorically scaled variables (Banker and Morey, 1986b), exogenously fixed variables (Banker and Morey, 1986a), allocative efficiency measurement (Banker, 1985; Banker and Maindiratta, 1988), random measurement errors (Banker, Datar, and Kemerer, 1991; Banker, Datar, and Rajan, 1987), and separable production functions (Banker, 1985; Banker, Das, and Datar, 1989). I shall not discuss these other models here, except to note that each is designed for specific production environments, and a researcher must select the model that is appropriate for the application at hand. The case of separable production, however, is developed further in the next section to show how the CCR model is modified to reflect the CRS–MOI technology assumed by MBK. This, in turn, enables me to show that the aggregate efficiency measure appropriate for this technology always satisfies MBK's matched output/input axiom.

Matched output/input technology

A production technology T is separable in *all* its inputs if there exist functions $g_i(\bullet): Y \to R$, $Y \subseteq R^{\eta}$, such that

$$T_s = \{(\mathbf{x}, \mathbf{y}) | x_i \geq g_i(\mathbf{y}),\ i \in I\}. \tag{5}$$

The CRS–MOI technology T_M in (2) is a special case of (5) when $g_i(\mathbf{y}) = y_i/m_i$, $m_i > 0$, for each $i \in \mathbf{I}$.

The DEA postulates are adapted to this case of *separable* production technology as follows (Banker, 1985; Banker, 1989; Banker et al., 1989).

Postulate 1′: Convexity of production function
If $\mathbf{y}_1, \mathbf{y}_2 \in Y$ then $g_i(\lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2) \leq \lambda g_i(\mathbf{y}_1) + (1-\lambda) g_i(\mathbf{y}_2)$ for all $0 \leq \lambda \leq 1$, and $i \in I$.

Postulate 2′: Monotonicity of production function
If $\mathbf{y}_1, \mathbf{y}_2 \in Y$ and $\mathbf{y}_1 \geq \mathbf{y}_2$ then $g_i(\mathbf{y}_1) \geq g_i(\mathbf{y}_2)$ for all $i \in I$.

Postulate 3′: Envelopment of observations
For all $d \in D$, $i \in I$, and $\mathbf{y}_d \in Y$, we have $x_{id} \geq g_i(\mathbf{y}_d)$.

Postulate 4′: Minimum extrapolation
If $\phi_i(\bullet)$ satisfies the other specified postulates for the production function, then $\phi_i(\mathbf{y}) \leq g_i(\mathbf{y})$ for all $\mathbf{y} \in Y$ and $i \in I$.

The unique production function $\hat{g}_i(\bullet)$ estimated on the basis of these postulates is given by the following (Banker, 1989):

$$\hat{g}_i(\mathbf{y}) = \min x_i \tag{6.0}$$

subject to

$$\Sigma_{d \in D}\ \lambda_{id} x_{id} \leq x_i \tag{6.1}$$

$$\Sigma_{d \in D}\ \lambda_{id} \mathbf{y}_d \geq \mathbf{y} \tag{6.2}$$

$$\Sigma_{d \in D}\ \lambda_{id} = 1 \tag{6.3}$$

$$x_i,\ \lambda_{id} \geq 0. \tag{6.4}$$

Defining $\hat{\rho}_i(\mathbf{x}_0, \mathbf{y}_0) = \hat{g}_i(y_0)/x_0$ for any observation $(\mathbf{x}_0, \mathbf{y}_0)$, the analogy between (4) and (6) becomes transparent. If each $g_i(\bullet)$ is postulated to be convex and monotone increasing, and $\rho_i = g_i(\mathbf{y})/x_i$, $i \in I$, are postulated to be distributed independently with $\int_{1-\delta}^{1} f_i(\rho_i) d\rho_i > 0$ for all $\delta > 0$, then the estimates of $\rho_i(\mathbf{x}_0, \mathbf{y}_0)$ obtained from (6) are statistically consistent. Further, if constant returns to scale are also imposed as in postulate 7′, then $g_i(\mathbf{y})$ is estimated by optimizing (6.0) subject only to (6.1), (6.2), and (6.4), the constraint in (6.3) is omitted.

Postulate 7′: Separable constant returns to scale production
If $k > 0$ and $\mathbf{y}$, $k\mathbf{y} \in Y$, then $g_i(k\mathbf{y}) = k g_i(\mathbf{y})$ for all $i \in I$.

The matched output/input technology of MBK assumes the following in addition to separable production and constant returns to scale.[5]

Postulate 8′: Matched output/input
For each $i \in I$, $g_i(\mathbf{y})$ is independent of y_θ for all $\theta \neq i$.

Postulates 7′ and 8′ are logically consistent with postulates 1, 2, 3, and 7 but will violate postulate 4 if the CCR efficiency measure based on the smallest set satisfying only postulates 1, 2, 3, and 7 is employed because this smallest set need not satisfy the additional assumptions embodied in the two new postulates (7′ and 8′). The appropriate DEA linear program to estimate $\hat{g}_i(\mathbf{y})$, therefore, is

$$\hat{g}_i(\mathbf{y}) = \min \mathbf{x}_i \tag{7.0}$$

subject to

$$\Sigma_{d\in D}\ \lambda_{id}x_{id} \leq x_i \tag{7.1}$$

$$\Sigma_{d\in D}\ \lambda_{id}y_{id} \geq y_i \tag{7.2}$$

$$x,\ \lambda_{id} \geq 0. \tag{7.3}$$

The optimal solution to (7) is simply $\min_{d\in D}\{y_i x_{id}/y_{id}\}$. Thus, more formally we have the following:

Proposition:
The unique function $\hat{g}_i(\mathbf{y})$ determined by postulates 1′, 2′, 7′, 8′, 3′, and 4′ is $\hat{g}_i(y_i) = \min_{d\in D}\{y_i x_{id}/y_{id}\}$, which also solves the linear program in (7).

Proof:
Postulates 7′ and 8′ imply $g_i(\mathbf{y}) = m_i y_i$, $m_i > 0$, $i \in I$. Postulate 3′ then implies $m_i \leq x_{id}/y_{id}$ for all $d \in D$. If $m_i < \min_{d\in D}\{x_{id}/y_{id}\}$ then there exists $\epsilon > 0$ such that $m_i + \epsilon \leq \min_{d\in D}\{x_{id}/y_{id}\}$. Therefore, a function $\phi_i(\mathbf{y}) = (m_i + \epsilon)\ y_i$ satisfies postulates 1′, 2′, 3′, 7′, and 8′, and yet $(m_i + \epsilon)y_i > m_i y_i$, thus violating postulate 4′. It follows, therefore, that $g_i(y_i) = m_i y_i$, $m_i = \min_{d\in D}\{x_{id}/y_{id}\}$ satisfies postulates 1′, 2′, 3′, 4′, 7′, and 8′. Postulate 4′ further ensures that it is the unique function satisfying these postulates.

Next, it is easy to verify that $\min_{d\in D}\{y_i x_{id}/y_{id})$ is a feasible solution to the linear program in (7) by setting $\lambda_{ic} = y_i/y_{ic}$, $\lambda_{id} = 0$ for $d \neq c$ where c is an observation ($\in D$) for which x_{id}/y_{id} is minimized, that is, $x_{ic}/y_{ic} = \min_{d\in D}\{x_{id}/y_{id}\}$.

Further, if $\{x_i^*, \lambda_{id}^*;\ d \in D\}$ is an optimal solution to (7), then define $\lambda_{ic}^0 = \lambda_{ic}^* + \sum_{d\neq c}(\lambda_{id}^* y_{id}/y_{ic})$ and $\lambda_{id}^0 = 0$ for $d \neq c$. Evidently,

5 In fact, postulates 7′ and 8′ together imply, and therefore make redundant, each of postulates 1′ and 2′.

$$\Sigma_{d\in D}\,\lambda^0_{id}y_{id} = \Sigma_{d\in D}\,\lambda^*_{id}y_{id} \geq y_i \text{ and}$$

$$\begin{aligned}\Sigma_{d\in D}\,\lambda^0_{id}x_{id} &= \Sigma_{d\in D}\,\lambda^*_{id}x_{ic}y_{id}/y_{ic}\\ &\leq \Sigma_{d\in D}\,\lambda^*_{id}x_{id}\\ &\leq x^*_i.\end{aligned}$$

Therefore, an optimal solution to (7) can be obtained by setting $\lambda_{id} = 0$ for all $d \neq c$. From (7.1) and (7.2) it follows that $\lambda^*_{ic} = y_i/y_{ic}$ and $x^*_i = \lambda^*_{ic}x_{ic} = y_i x_{ic}/y_{ic}$. ■ ■ ■

Thus, the appropriate DEA model reflecting the postulated matched output/input technology of MBK estimates the production set T_M as

$$\hat{T}_M = \{(\mathbf{x}, \mathbf{y})|x_i \geq y_i \min_{d\in D}\{x_{id}/y_{id}\},\ i \in I\}. \tag{8}$$

The partial efficiency measure for each observation (x_j, y_j), $j \in D$, is defined as

$$\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j) = \hat{g}_i(y_{ij})/x_{ij} = (y_{ij}/x_{ij}) \min_{d\in D}\{x_{id}/y_{id}\}. \tag{9}$$

An aggregate efficiency measure that is based on Shephard's distance measure is

$$\begin{aligned}\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j) &= \inf\{\rho|(\rho\mathbf{x}_j, \mathbf{y}_j) \in \hat{T}_M\}\\ &= \max_{i\in I}\{\hat{\rho}_i(\mathbf{x}_j, \mathbf{y}_j)\}\\ &= \max_{i\in I}\{(y_{ij}/x_{ij})\min_{d\in D}\{x_{id}/y_{id}\}\}.\end{aligned} \tag{10}$$

Clearly, if there exists some j, $k \in D$, with $y_{ij}/x_{ij} > y_{ik}/x_{ik}$ for every $i \in I$, then $\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j) = \max_{i\in I}\{(y_{ij}/x_{ij})\min_{d\in D}\{x_{id}/y_{id})\}\} > \max_{i\in I}\{(y_{ik}/x_{ik})\min_{d\in D}\{x_{id}/y_{id}\}\} = \hat{\rho}(\mathbf{x}_k, \mathbf{y}_k)$. Thus, it is evident that the matched output/input axiom of MBK is satisfied by the aggregate efficiency measure based on the appropriate DEA estimate of Shephard's distance function.

The above lexicographic measure in (10), however, reflects the relative output/input ratio for only one such matched pair. An alternative aggregate efficiency measure suggested by Banker (1985) is constructed as a weighted sum of the partial efficiency measures:[6]

$$\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j) = \Sigma_{i\in I}\, w_i\hat{\rho}_i(\mathbf{x}_j, \mathbf{y}_j) \text{ where } w_i \geq 0,\ \Sigma_{i\in I}\, w_i = 1. \tag{11}$$

The weights w_i are chosen to reflect the relative importance of the individual matched output/input pairs. This aggregate efficiency measure is appropriate for any separable production technology with $\hat{\rho}_i(\mathbf{x}_j, \mathbf{y}_j) = \hat{g}_i(\mathbf{y}_j)/x_{ij}$.

It is also easy to verify that the aggregate efficiency measure in (11) satisfies MBK's matched output/input axiom. Clearly, if $y_{ij}/x_{ij} > y_{ik}/x_{ik}$ for every $i \in I$, for some j, $k \in D$, then

6 Banker (1985) considers explicitly the case of a production technology separable in outputs, but his aggregate efficiency measure is readily adapted to the separable inputs case as in Banker et al. (1989).

$$\begin{aligned}\hat{\rho}_i(\mathbf{x}_j, \mathbf{y}_i) &= (y_{ij}/x_{ij})\min_{d\in D}\{x_{id}/y_{id}\}\\ &> (y_{ik}/x_{ik})\min_{d\in D}\{x_{id}/y_{id}\}\\ &= \hat{\rho}_i(\mathbf{x}_k, \mathbf{y}_k)\end{aligned}$$

Therefore,

$$\begin{aligned}\hat{\rho}(\mathbf{x}_j, \mathbf{y}_j) &= \Sigma_{i\in I}\, w_i\hat{\rho}_i(\mathbf{x}_j, \mathbf{y}_j)\\ &> \Sigma_{i\in I}\, w_i\hat{\rho}_i(\mathbf{x}_k, \mathbf{y}_k)\\ &= \hat{\rho}(\mathbf{x}_k, \mathbf{y}_k).\end{aligned}$$

Thus, the aggregate efficiency measure based on Shephard's distance function, and that suggested by Banker (1985), are both consistent with the matched output/input axiom. In the latter case, a common set of weights w_i, $i \in I$, independent of observations $d \in D$, produces an efficiency measure satisfying the MOI axiom.

A simulation study

MBK employ in their paper the CCR efficiency measure obtained by optimizing (4.0) subject to (4.1), (4.2), and (4.4). If the underlying technology is postulated to be CRS–MOI, however, the appropriate efficiency measures are those described in (10) or (11). MBK's use of the CCR efficiency measure ignores the technology they postulate. Therefore, in general, it can lead to estimated values that are not consistent with the discarded information about the postulated CRS–MOI technology.

The statistical consistency results of Banker (1992), however, suggest that the likelihood of occurrences of such observed behavior of the estimated measures will be small for large samples. Specifically, I note that the *additional* assumptions of separability and matched output/input for MBK's postulated technology T_M in (2) are *not* logically inconsistent with the assumptions of convexity, monotonicity, and constant returns to scale (postulates 1, 2, and 7) underlying the CCR efficiency measure. Since the CCR measure provides statistically consistent estimates of efficiency under the maintained assumptions reflected in postulates 1, 2, 7, and 5, it follows that the CCR model will retrieve the true production set asymptotically even when the true underlying technology is MBK's CRS–MOI technology.

As noted in the second section, the CCR measure is based on Shephard's distance measure. Therefore, if the true technology is CRS–MOI as in (2), the true efficiency measure based on Shephard's distance measure is $\max_{i\in I}\{y_{id}/m_i x_{id}\}$ as in (3). Statistical consistency of the CCR efficiency measure implies that it will approach the true value in (3) for large samples. Based on the arguments presented in the second section, it also follows that this lexicographic measure in (3) satisfies MBK's MOI axiom. Thus, distortions in estimated efficiency ranks may be observed by MBK for small samples as they use a misspecified model, but the frequency of occurrence of such distortions is likely to be small for large samples.

I illustrate the above inferences with a simulation study.[7] Consider a two-input, two-output CRS–MOI technology where the parameters m_i, $i = 1, 2$, are normalized at one. Observations are simulated by first drawing $\mathbf{x}_{id}$ randomly from independent uniform distributions on [5, 15] for each $i = 1, 2, d \in D$ (see Banker, Charnes, Cooper, and Maindiratta, 1987). The true partial efficiencies ρ_{id}, $i = 1, 2$, are simulated by drawing $\gamma_{id} = (1 - \rho_{id})/\rho_{id}$ randomly from an exponential distribution with a parameter $\mu_i = 1$, that is, $f_i(\rho_{id}) = \exp\{-(1 - \rho_{id})/\rho_{id}\}$ for $0 < \rho_{id} \leq 1$. The (true) efficiency measure based on Shephard's distance function is $\rho_d = \max_{i \in I}\{\rho_{id}\}$. The observations y_{id}, $i = 1, 2$, are then constructed by setting $y_{id} = \rho_{id} x_{id}$, $d \in D$. Three sample sizes are selected for this simulation study; the number of observations in D are set at 50, 75, or 100. The experiment is repeated five times for each sample size.

Each experiment involves first simulating 50, 75, or 100 observations $(\mathbf{x}_d, \mathbf{y}_d)$, $d \in D$, as described above. For each simulated observation $(\mathbf{x}_k, \mathbf{y}_k)$, $k \in D$, CCR efficiency measure $\hat{\rho}_k$ is then computed by solving the linear program optimizing (4.0) subject to (4.1), (4.2), and (4.4). Partial efficiency estimates $\hat{\rho}_{ik} = (y_{ik}/x_{ik})\min_{d \in D}\{x_{id}/y_{id}\}$ and the lexicographic aggregate efficiency measure $\hat{\rho}_{ok} = \max_{i \in I} \hat{\rho}_{ik}$ as in (10) are also obtained, as discussed in the third section. Four measures of estimation accuracy for each simulated observation $k \in D$ are then constructed as follows:

$$\hat{\alpha}_k = \rho_k / \hat{\rho}_k \tag{12.1}$$

$$\hat{\alpha}_{ok} = \rho_k / \hat{\rho}_{ok} \tag{12.2}$$

$$\hat{\alpha}_{ik} = \rho_{ik} / \hat{\rho}_{ik}, \quad i = 1,\, 2. \tag{12.3}$$

For a finite sample size, DEA overestimates the efficiency relative to the true production set, as it considers only the *best practice* production set, which is a subset of the true production set. Therefore, the accuracy measures are less than or equal to one for all $k \in D$.

The mean and standard deviations for the three estimation accuracy measures over all five experiments for each sample size are reported in Table 1. Since the DEA efficiency measures are statistically consistent, accuracy measure $\hat{\alpha}_k$ for the CCR efficiency estimates is expected to increase with larger sample size. The evidence supports this expectation as $0.9559 < 0.9685 < 0.9767$. The same result holds for the accuracy measure $\hat{\alpha}_{ok}$ designed specially for the MOI technology as $0.9808 < 0.9856 < 0.9925$. Although both aggregate efficiency measures perform very well as evidenced by their high accuracy measures, note

7 Computational assistance by Hsi-hui Chang is gratefully acknowledged.

TABLE 1
Simulation study results

	Sample size of 50		Sample size of 75		Sample size of 100	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Accuracy measures:						
$\hat{\alpha}_k = \rho_k/\hat{\rho}_k$	0.9559	0.0426	0.9685	0.0287	0.9767	0.0190
$\hat{\alpha}_{0k} = \rho_k/\hat{\rho}_{0k}$	0.9808	0.0240	0.9856	0.0107	0.9925	0.0032
$\hat{\alpha}_{1k} = \rho_{1k}/\hat{\rho}_{1k}$	0.9889	0.0070	0.9843	0.0139	0.9930	0.0039
$\hat{\alpha}_{2k} = \rho_{2k}/\hat{\rho}_{2k}$	0.9705	0.0318	0.9868	0.0066	0.9919	0.0016
Proportion of type A observation pairs	0.519	0.028	0.499	0.053	0.508	0.027
Proportion of type B observation pairs	0.00767	0.00767	0.00702	0.00778	0.00388	0.00191

Estimation accuracy measures $\hat{\alpha}_k$, $\hat{\alpha}_{0k}$, $\hat{\alpha}_{1k}$, and $\hat{\alpha}_{2k}$ are defined in expressions (12.1), (12.2), and (12.3) in the text. Type A observation pairs are observations j, k for which $y_{ij}/x_{ij} > y_{ik}/x_{ik}$ for all $i \in I$. Type B observation pairs are the type A pairs for which $\rho_j < \rho_k$. Proportions are computed as ratios of the number of occurrences of type A (or type B) pairs to the total number of pairs ($= N(N-1)/2$, where N is the sample size). Expected proportion of type A observation pairs is 0.500.

that $\hat{\alpha}_k < \hat{\alpha}_{ok}$ for each sample size, because $\hat{\alpha}_{ok}$ reflects the accuracy of the efficiency measure that is appropriate for the underlying technology.

Table 1 also presents evidence on the frequency of cases for which MBK's MOI axiom is violated. For this purpose, type A observation pairs are defined to be simulated observations j, $k \in D$ such that $y_{ij}/x_{ij} > y_{ik}/x_{ik}$ for $i = 1, 2$. Type B observation pairs are defined to be the Type A pairs for which the CCR efficiency measure $\hat{\rho}_j < \hat{\rho}_k$, thus violating the MOI axiom. Table 1 reports that type B observation pairs occur in only 0.767 percent of the cases for experiments with sample size 50, reducing further to 0.702 percent and 0.388 percent for sample sizes of 75 and 100. The decreasing proportion of type B observation pairs, reducing to less than half a percent for sample size of 100, is consistent with the expectation that the likelihood of their occurrence approaches zero for sufficiently large samples.

Concluding remarks

An important responsibility of a researcher is to select carefully the estimation model that best represents the known and postulated relation between all variables. The same principle applies also to the selection of efficiency evaluation models, where a researcher must consider what characterizes the underlying production technology that represents the conversion of inputs into outputs. If the selected estimation model does not fully reflect the known characteristics of the production technology, it is not surprising to find that some efficiency estimates are not consistent with the known characteristics that are omitted. The paper by MBK documents such distortions resulting from their misspecification of the estimation model.

The rules for practitioners in selecting the appropriate DEA model are straightforward. Choose the DEA model that is based on postulates specifying precisely the information that is known or believed to exist for the output/input correspondence in the organization being studied. If any additional characteristics are known, make sure that the DEA efficiency measure is modified, or the "minimum extrapolation" postulate in DEA will be violated. That is, the smallest set will be picked from *all* possible sets satisfying the specified postulates, including sets that do not satisfy the *additional* characteristics. Validity of the estimated efficiency measures cannot be guaranteed if a misspecified model is employed. Banker's (1992) statistical consistency results, however, are reassuring for large sample studies even when the model is misspecified. If the additional characteristics omitted in the DEA model used for the study are logically consistent with the *other* postulates on which the model is based, then the errors resulting from the misspecification are likely to be small in large samples.

The present paper shows that the choice of the efficiency evaluation model that best represents MBK's postulated matched output/input technology provides aggregate efficiency estimates that *never* violate their matched output/input axiom. Therefore, what appears to be a *paradox* to them is simply distortion caused by the *misspecification* of their efficiency model. MBK's misspecified model does not lead to serious distortions in large samples, however, because of the statistical consistency of the DEA estimator. The simulation study presented here illustrates this asymptotic property. The percentage of observation pairs violating MBK's MOI axiom is found to be less than 1 percent for sample size of 50 and less than half a percent for sample size of 100. Although this result is reassuring, it should not undermine the importance of devoting sufficient time and effort to understand the output/input relationship before estimating production efficiencies.

References

Banker, R.D., "Productivity Measurement and Management Control," in *The Management of Productivity and Technology in Manufacturing*, ed. P. Kleindorfer (New York: Plenum, 1985) pp. 239–257.

———, "Econometric Estimation and Data Envelopment Analysis," *Research in Governmental and Nonprofit Accounting* (1989) pp. 231–244.

———, "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation," *Management Science* (in press, 1992).

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

——— and A. Maindiratta, "A Comparison of DEA and Translog Estimates of Production Functions Using Simulated Data from a Known Technology," *Applications of Modern Production Theory: Efficiency and Productivity*, ed. A. Dogramaci and R. Fare (Boston: Kluwer, 1987) pp. 33–58.

Banker, R.D., S. Das, and S.M. Datar, "Analysis of Cost Variances for Management Control in Hospitals," *Research in Governmental and Nonprofit Accounting* (1989) pp. 269–292.

Banker, R.D., S.M. Datar, and C. Kemerer, "A Model to Evaluate Variables Impacting the Productivity of Software Maintenance Projects," *Management Science* (January 1991) pp. 1–18.

Banker, R.D., S.M. Datar, and M. Rajan, "Measurement of Productivity Improvements: An Empirical Analysis," *Journal of Accounting, Auditing and Finance* (Fall 1987) pp. 319–354.

Banker, R.D. and A. Maindiratta, "Piecewise Loglinear Estimation of Efficient Production Surfaces," *Management Science* (January 1986) pp. 126–135.

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

Banker, R.D. and R.C. Morey, "Efficiency Analysis for Exogenously Fixed Inputs and Outputs," *Operations Research* (July/August 1986a), pp. 513–521.

———, "Data Envelopment Analysis with Categorical Inputs and Outputs," *Management Science* (December 1986b) pp. 163–167.

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

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

Mehrez, A., J.R. Brown, and M. Khouja, "Aggregate Efficiency Measures and Simpson's Paradox," *Contemporary Accounting Research* (Fall 1992) pp. 329–342.

Shephard, R.W., *The Theory of Cost and Production Functions* (Princeton, N.J.: Princeton University Press, 1970).