The Journal of Productivity Analysis, 7, 139–159 (1996)

Hypothesis Tests Using Data Envelopment Analysis

RAJIV D. BANKER†
Carlson School of Management, University of Minnesota, Minneapolis, MN 55455
RBANKER@CSOM.UMN.EDU

Abstract

A substantial body of recent work has opened the way to exploring the statistical properties of DEA estimators of production frontiers and related efficiency measures. The purpose of this paper is to survey several possibilities that have been pursued, and to present them in a unified framework. These include the development of statistics to test hypotheses about the characteristics of the production frontier, such as returns to scale, input substitutability, and model specification, and also about variation in efficiencies relative to the production frontier.

Keywords. Data envelopment analysis, hypothesis tests, production frontier estimation

1. Introduction

The data envelopment analysis (DEA) methodology was developed in the management science tradition (Charnes, Cooper, and Rhodes 1978, 1981) with a focus on computing the relative efficiency of different decision-making units (DMUs). The DEA approach specifies the production set only in terms of properties such as convexity and monotonicity, without imposing any parametric structure on it (Banker, Charnes, and Cooper 1984). Early work in DEA did not specify any structure for the distribution of deviations from the production frontier and, therefore, did not explore the statistical properties of its estimators. Instead, it was limited to developing different mathematical programs to model different types of production relations (e.g., Banker and Maindiratta (1986), Banker and Morey (1986). Even the paper in *Econometrica* by Banker and Maindiratta (1988) focused only on checking whether a set of observations on production inuts and outputs could be reconciled with economic efficiency as in Varian's (1984) algebraic test for consistency with the weak axiom of cost minimization and whether any of the observations exhibited technical, allocative, or scale inefficiencies. The DEA approach, therefore, was labeled by some as nonstatistical (Schmidt 1985; Gong and Sickles 1992).

Recent work in DEA, however, has opened the way to exploring the statistical properties of its estimators of production functions. Banker (1993) shows that DEA provides a consistent estimator of arbitrary monotone and concave production functions when (the one-sided) deviations from the production function are regarded as stochastic variations in the technical inefficiency of individual observations. This estimator also maximizes the likelihood function if the probability density function for the inefficiency is monotone decreasing, a more robust characterization than Schmidt's (1976) corresponding result for Aigner and Chu's (1968) estimators of parametrically specified production frontiers. In addition, the empirical distribution of the DEA estimator of the technical inefficiency of individual observations asymptotically recovers its true distribution. These results are particularly useful because the DEA inefficiency estimator can be used to construct statistical tests for evaluating hypotheses about differences in inefficiency distributions for two or more groups of observations. Monte Carlo experiments reported in Banker and Chang (1994) document superior results with these DEA-based tests relative to conventional tests based on corrected ordinary least squares (COLS) estimation of parametrically specified production functions (Olson, Schmidt, and Waldman (1980)).

The purpose of this paper is to survey several possibilities that have been pursued in recent years based on the asymptotic distribution of the DEA estimator. These include the development of statistics to test hypotheses about the characteristics of the production frontier, such as returns to scale, input substitutability, and model specification. The materials in this paper are not original; I have prepared this paper by collating selected text and results from several recent papers (Banker 1989; Banker and Chang 1993, 1994, 1995; Banker, Chang, and Sinha 1994; Banker, Devaraj, and Sinha 1995). However, by presenting different DEA-based tests of hypotheses in a unified framework, I hope this paper will prove to be a useful reference to researchers in efficiency measurement and production frontier estimation.

The remainder of this paper has the following structure. In Section 2, I describe the statistical foundation provided in Banker (1993) for the DEA estimator of DMU efficiency. In Section 3, I summarize the results of three simulation studies evaluting the performance of Banker's (1993) tests of differences in efficiencies of two groups of DMUs. In Section 4, I describe the construction of statistics to test returns to scale based on the DEA inefficiency estimator, and the results of some simulation experiments to compare their performance to conventional tests of returns to scale. I present DEA tests of the input substitutability hypothesis in Section 5, along with associated simulation results. I describe some tests of model specification based on the DEA estimator in Section 6, and simulation results comparing these tests with conventional parametric methods. Finally, I present concluding remarks in Section 7.

2. Statistical Foundation of the DEA Estimator

Let $Y_j \equiv (y_{1j}, \ldots y_{Rj}) > 0$ and $X_j \equiv (x_{1j}, \ldots x_{Ij}) > 0$, $j = 1, \ldots N$, be the observed (R-dimensional) output and (I-dimensional) input vectors in a sample of N observations generated from the underlying production possibility set $P = \{(X, Y)|$ outputs Y can be produced from inputs $X\}$. Input quantities x_{ij}, $i = 1, \ldots I$, are random variables with positive probabilities throughout their domain sets $(x_i^L, x_i^H) \subseteq \Re^+$. Output mix proportion variables m_{rj}, $r = 1, \ldots R$, $\Sigma_{r=1}^R m_{rj} = 1$, are also random variables with positive probabilities throughout their domain sets $(m_r^L, m_r^H) \subseteq (0, 1) \subseteq \Re^+$, so that $m_{rj}/y_{rj} = m_{sj}/y_{sj}$ for all $r, s = 1, \ldots R$, and $j = 1, \ldots N$.[1] The inefficiency θ_j of an observation $(X_j, Y_j) \in P$, measured radially by the reciprocal of Shephard's (1970) distance function, is given by $\theta_j \equiv \theta(X_j, Y_j) = \sup\{\theta|(X_j, \theta Y_j) \in P\}$. The inefficiency is a random variable comprising both technical factors such as capital equipment, labor skill, and managerial ability, and random shocks that result in defects, damage, and breakdowns in any production process (Aigner and Chu 1968). Such specification permits random factors to influence the realization of production inefficiencies for different observations, but it precludes measurement errors in inputs and outputs (Førsund, Lovell and Schmidt, 1980; Greene, 1980).

Based on Banker (1993)[2], I specify the following structure for the production set P and the probability density function $f(\theta)$ for the inefficiency θ:

Postulate 1. (*Monotonicity*) If $(X_1, Y_1) \in P$, $X_2 \geq X_1$, and $Y_2 \leq Y_1$ then $(X_2, Y_2) \in P$.

Postulate 2. (*Convexity*) If $(X_1, Y_1) \in P$ and $(X_2, Y_2) \in P$ then $(\lambda_1 X_1 + \lambda_2 X_2, \lambda_1 Y_1 + \lambda_2 Y_2) \in P$ for all $\lambda_1, \lambda_2 \geq 0$ such that $\lambda_1 + \lambda_2 = 1$.

Postulate 3. (*Envelopment*) If $\theta < 1$ then $f(\theta) = 0$.

Postulate 4. (*Likelihood of efficient performance*) If $\delta > 0$ then $\int_1^{1+\delta} f(\theta)d\theta > 0$.

The DEA estimator of θ is obtained as $\hat{\theta}_j^B$ by solving the linear programming formulation of the so-called BCC model (Banker, Charnes, and Cooper (1984) given below in (1):

$$\hat{\theta}_j^B = \max \theta \tag{1}$$

$$\sum_{k=1}^{N} \lambda_k X_k \leq X_j \tag{2}$$

$$\sum_{k=1}^{N} \lambda_k Y_k \geq \theta\, Y_j \tag{3}$$

$$\sum_{k=1}^{N} \lambda_k = 1 \tag{4}$$

$$\theta, \lambda_k \geq 0. \tag{5}$$

This estimator is statistically consistent, and its asymptotic empirical distribution $\hat{F}^B(\theta)$ recovers the true distribution of θ under the maintained assumptions embodied in the above four postulates (Banker 1993, Theorems 5 and 6).[3] Korostelev, Simar, and Tsybakov (1995) show that the DEA estimator in the single-output I-input case converges at the rate $N^{-2/(I+2)}$ and that no other estimator can converge with a faster rate. The DEA estimator also maximizes likelihood if, in addition, the following postulate is satisfied (Banker 1993, Theorem 2):

Postulate 5. (*Decreasing probability density*) If $\theta_2 \geq \theta_1$, then $f(\theta_2) \leq f(\theta_1)$.

Banker (1993) also suggests statistical tests of differences in the inefficiency of two different types of DMUs, say types T1 and T2. Let G1 $\equiv \{1, \ldots N\} \cap$ T1 and G2 $\equiv \{1, \ldots N\} \cap$ T2, G1 $\cap$ G2 $= \phi$, be the groups of these two types of DMUs in the sample, and let N_1 and N_2 be the number of observations in the two groups G1 and G2 respectively. Then Banker (1993) presents the following three test procedures:

(i) If we maintain the assumption that the inefficiences θ_j follow the exponential distribution for both types of DMUs with means $1 + \sigma_1$ and $1 + \sigma_2$, respectively, then both $\Sigma_{j\epsilon G1}\ 2(\hat{\theta}_j^B - 1)/\sigma_1$ and $\Sigma_{j\epsilon G2}\ 2(\hat{\theta}_j^B - 1)/\sigma_2$ asymptotically follow the chi-square distribution with $2N_1$ and $2N_2$ degrees of freedom, respectively. Therefore, under the null hypothesis H0: $\sigma_1 = \sigma_2 = \sigma$ (indicating that both types of DMUs have the same inefficiency distribution), we can test the null hypothesis H0 against the alternative hypothesis H1: $\sigma_1 > \sigma_2$ (indicating that the first type is on average less efficient than the second type), using the test statistic given by $T_{EX} = [\Sigma_{j\epsilon G1}\ (\hat{\theta}_j^B - 1)/N_1]/[\Sigma_{j\epsilon G2}\ (\hat{\theta}_j^B - 1)/N_2]$ evaluated relative to the F-distribution with $(2N_1, 2N_2)$ degrees of freedom.

(ii) If we maintain the assumption that the random variables $\psi_j = \theta_j - 1$ are distributed as $|N(0, \sigma_i)|$, $i = 1, 2$, for the two types of DMUs then both $\Sigma_{j\epsilon G1}\ (\hat{\theta}_j^B - 1)^2/\sigma_1$ and $\Sigma_{j\epsilon G2}\ (\hat{\theta}_j^B - 1)^2/\sigma_2$ asymptotically follow the chi-square distribution with N_1 and N_2 degrees of freedom, respectively. Therefore, under the null hypothesis H0: $\sigma_1 = \sigma_2 = \sigma$, we can test the null hypothesis H0 against the alternative hypothesis H1: $\sigma_1 > \sigma_2$ using the test statistic given by $T_{HN} = [\Sigma_{j\epsilon G1}\ (\hat{\theta}_j^B - 1)^2/N_1]/[\Sigma_{j\epsilon G2}\ (\hat{\theta}_j^B - 1)^2/N_2]$ evaluated relative to the F-distribution with (N_1, N_2) degrees of freedom.

(iii) If no such assumptions are maintained about the probability distribution, then we can employ a nonparametric Smirnov's two-sample type test procedure based on the maximum vertical distance of the empirical distribution $\hat{F}^B_{G1}(\hat{\theta}_j^B)$ of DMUs in group G1 above the empirical distribution $\hat{F}^B_{G2}(\hat{\theta}_j^B)$ of DMUs in group G2. In this case, the test statistic is given by $T_{SM} \equiv \max\{\hat{F}^B_{G1}(\hat{\theta}_j^B) - \hat{F}^B_{G2}(\hat{\theta}_j^B) | j = 1, \ldots, N\}$.

The principle underlying the construction of the test-statistics in (i) and (ii) above can be extended readily to a few other cases where different assumptions are maintained about the inefficiency distributions. For instance, if we maintain the assumption that for some transformation function $R(\cdot)$, $R(\theta_j)$ are distributed as normal, or half-normal, then we can use the test statistic $T_R = [\Sigma_{j\epsilon G1}\ (R(\theta_j))^2/N_1]/[\Sigma_{j\epsilon G2}\ (R(\theta_j))^2/N_2]$ to test the null hypothesis.

That the inefficiencies θ_j are independently distributed is a maintained assumption, and the distribution of the DEA estimator of inefficiency $\hat{\theta}_j^B$ asymptotically recovers the true distribution. Therefore, $\hat{\theta}_j^B$ are asymptotically independent of each other. However, for any finite sample, the DEA estimates of inefficiencies of different observations need not be independently distributed. More importantly, for any finite sample, the distribution of the DEA estimator $\hat{\theta}_j^B$ need not follow the underlying true distribution of θ. In fact, Banker (1993) shows that the DEA estimator is biased for finite samples. Therefore, $\Sigma_{j\epsilon G1}\ 2(\hat{\theta}_j^B - 1)/\sigma_i$ or $\Sigma_{j\epsilon G1}\ (\hat{\theta}_j^B - 1)^2/\sigma_i$ need not follow the chi-square distribution, these two statistics for the two groups G1 and G2 need not be independently distributed, and their ratio need not follow the F-distribution for finite samples. Thus, it is important to evaluate the performance of these new tests with systematic simulation studies (Banker 1993, p. 1272).

3. Simulation Studies of Tests of Efficiency Differences

This section presents results of two simulation studies reported earlier in Banker and Chang (1994, 1995), and a third study by Kittelsen (1995). Banker and Chang's first study was designed to compare Banker's (1993) tests with traditional methods used in DEA and with an appropriately specified COLS model. This study also evaluated the results of a new test using the COLS estimators that is based on Banker's (1993) Theorem 6 about the asymptotic recovery of the true inefficiency distribution in frontier estimation. The second study evaluates the performance of these various tests under conditions designed to be adverse to the new DEA-based tests. Specifically, it considers settings that reflect the two most common criticisms of the DEA methodology: (1) Possible pesence of measurement (or specification) errors; and (2) explicit incorporation of optimal allocations such as those embodied in Shephard's lemma. In addition, the setting is made even more adverse to the new DEA-based tests by simulating true inefficiency values θ_j from probability density functions that vanish at $\theta = 1$ instead of having a mode at $\theta = 1$ as assumed in Banker's first two tests. Also discussed in this section is a study by Kittelsen (1995) which is similar to the first Banker and Chang (1994) study except that it only evaluates the occurrence of Type II errors.

In the first study, Banker and Chang (1994) generated 800 different data sets varying the specification of three factors: production technology, inefficiency distribution, and sample size. They considered five different production technologies, four different inefficiency distributions, and four different sample sizes, and replicated each of the 80 (=5 * 4 * 4) settings 10 times to yield 800 (=80 * 10) data sets. Banker and Chang (1994) specified the following five different single output, multiple-input production technologies described in terms of Cobb-Douglas type production frontiers $q = h(X)$, where q represented the maximum output that can be produced from the input vector X:

Two-input, one-output technologies:

Linear homogeneous production frontier: $q = 10x_1^{0.6}x_2^{0.4}$ (6)

Concave production frontier: $q = 10x_1^{0.4}x_2^{0.3}$ (7)

Convex production frontier: $q = 10x_1^{0.8}x_2^{0.5}$ (8)

Five-input, one-output technologies:

Concave production frontier: $q = 10x_1^{0.2}x_2^{0.2}x_3^{0.1}x_4^{0.1}x_5^{0.1}$ (9)

Convex production frontier: $q = 10x_1^{0.3}x_2^{0.3}x_3^{0.3}x_4^{0.2}x_5^{0.2}$ (10)

Each sample generated for this study had two groups of DMUs. The first group of DMUs comprised 40% of the observations in each sample, and the second group comprised the remaining 60% of the observations. Four different inefficiency distributions were considered. The first distribution combined two half-normal distributions that reflect differences in the inefficiencies of the two groups of DMUs. The random variable $\psi_j = \theta_j - 1 \geq 0$ was simulated from the half-normal distribution $|N(0, 1)|$ for the first group of DMUs. The random variable ψ_j for the second group of DMUs was generated from the half-normal distribution $|N(0, 0.78)|$. The second distribution combined two exponential distributions with different mean inefficiencies for the two groups. The third distribution had both groups following the same half-normal distribution, and the fourth distribution had both groups with the same exponential distribution.

Four different sample sizes: 50, 100, 150, and 200, were considered. The values X_j of the input vector for each observation j were simulated from independent uniform distributions on the interval [5, 15]. The corresponding value of the output $y_j = q_j/\theta_j$ for each observation j was obtained using the production frontier described in (6) to (10) above and the simulated value of the inefficiency term $\theta_j = \psi_j + 1$.

Two traditional test procedures, Welch's mean test and Mann-Whitney test (Siegel and Castellan 1988), that have been used to test for inefficiency differences between two groups of DMUs in the existing DEA literature (e.g., Banker, Conrad, and Strauss (1986), Grosskopf and Valdmanis (1987), and two parametric COLS specifications (loglinear and translog) of the estimated production frontier, with a dummy variable to test for differences in the mean inefficiency between the two groups of DMUs, were evaluated and compared with Banker's (1993) tests. Representative results for the concave production frontier case in (8) with half-normal distributions are reproduced here as Tables 1a and 1b.

Banker and Chang (1994) summarize their results as follows: (1) The new asymptotic DEA tests outperform COLS-based tests even when the parametric functional form employed for COLS estimation is identical to the underlying production frontier: (2) the asymptotic DEA tests are robust in that they perform well for different underlying production frontiers and inefficiency distributions: (3) Welch's mean and Mann-Whitney tests perform worse than both the new asymptotic DEA tests and the COLS-based tests: (4) the performance of COLS-based tests is improved when Banker's (1993) new test statistics are employed using COLS estimates of production inefficiencies; and (5) no marked differences in Type II error are found between the various tests with one-output two-input production technologies. But, Welch's mean test, Mann-Whitney test, and COLS-based tests perform slightly better than the asymptotic DEA tests in terms of proportion of Type II errors with one-output five-input production technologies.

The most surprising of their results was that the new asymptotic DEA tests outperformed the conventional COLS-based test even when the parametric functional form specified for COLS estimation was identical to the true underlying production frontier (see also Mensah and Li, 1993). These results prevailed for all four sample sizes and the two different dimensions of input space. These surprising results were also robust in that they prevailed even when the underlying production frontier violated the assumptions of concavity or constant returns to scale for the BCC and CCR models in DEA, and when the true inefficiency distribution was different from that postulated in Banker's (1993) DEA test statistic. In addition, the Welch and Mann-Whitney tests that have been used in several prior DEA-based research studies were found to perform the worst in their simulation study. This was true for all four sample sizes, regardless of the underlying production frontier or the inefficiency distribution used in the generation of the simulated data.

The second study by Banker and Chang (1995) considered conditions particularly favorable to the parametric cost frontier-based tests, using simulated data on multiple outputs with measurement errors, generated consistent with Shephard's lemma. Thus, their experimental setting violated the DEA assumption of no measurement error (Aigner, Lovell and Schmidt, 1977), and further it explicitly incorporated the assumption of optimal allocation of input

Table 1a. Summary of efficiency difference test results for a convex production function. (Source: Banker and Chang (1994))

	Estimation Model	Test Procedure	Sample Size			
			$N = 50$	$N = 100$	$N = 150$	$N = 200$
Banker's asymptotic DEA tests	BCC	exponential	7 (4)	4 (3)	5 (5)	7 (6)
		half-normal	7 (6)	4 (4)	9 (7)	9 (8)
	CCR	exponential	6 (1)	3 (2)	9 (4)	7 (6)
		half-normal	8 (7)	3 (3)	7 (6)	9 (8)
Tests used in prior DEA literature	BCC	Welch	5 (2)	3 (3)	2 (2)	7 (3)
		Mann-Whitney	4 (2)	4 (3)	2 (2)	7 (4)
	CCR	Welch	2 (1)	2 (2)	4 (3)	7 (5)
		Mann-Whitney	2 (1)	2 (2)	4 (2)	6 (6)
COLS-based tests	loglinear	dummy variable	3 (1)	3 (2)	6 (5)	7 (5)
	translog	dummy variable	3 (2)	1 (1)	6 (4)	8 (6)
New COLS-based tests with Banker's test statistic	loglinear	exponential	3 (1)	3 (1)	6 (5)	7 (5)
		half-normal	6 (5)	6 (5)	8 (8)	10 (9)
	translog	exponential	3 (1)	1 (1)	5 (1)	6 (5)
		half-normal	5 (4)	3 (2)	8 (5)	9 (8)

Note: The number reported in each cell is the number of iterations (out of a total of 10 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number in parentheses in each cell indicates the number of iterations for which the test statistic is significant at the 5% level.

Table 1b. Summary of efficiency difference test results for a convex production function. (Source: Banker and Chang (1994))

	Estimation Model	Test Procedure	Sample Size			
			$N = 50$	$N = 100$	$N = 150$	$N = 200$
Banker's asymptotic DEA tests	BCC	exponential	0 (0)	0 (0)	2 (1)	2 (1)
		half-normal	0 (0)	0 (0)	2 (1)	2 (2)
	CCR	exponential	0 (0)	0 (0)	1 (0)	0 (0)
		half-normal	0 (0)	1 (0)	1 (0)	1 (0)
Tests used in prior DEA literature	BCC	Welch	0 (0)	1 (1)	1 (1)	1 (1)
		Mann-Whitney	0 (0)	1 (1)	2 (1)	1 (1)
	CCR	Welch	0 (0)	1 (1)	1 (0)	1 (1)
		Mann-Whitney	0 (0)	1 (1)	2 (0)	2 (1)
COLS-based tests	loglinear	dummy variable	0 (0)	1 (1)	3 (0)	1 (1)
	translog	dummy variable	0 (0)	1 (1)	2 (1)	1 (1)
New COLS-based tests with Banker's test statistic	loglinear	exponential	0 (0)	0 (0)	0 (0)	2 (1)
		half-normal	0 (0)	0 (0)	2 (1)	2 (1)
	translog	exponential	0 (0)	0 (0)	1 (0)	0 (0)
		half-normal	0 (0)	0 (0)	2 (1)	2 (0)

Note: The number reported in each cell is the number of iterations (out of a total of 10 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number in parentheses in each cell indicates the number of iterations for which the test statistic is significant at the 5% level.

resources underlying parametric cost frontier estimation, but not considered in DEA estimation. Further, they generated the logarithm of the inefficiency variable from half-normal or exponential distributions, so that the modes of the distributions of the simulated inefficiencies θ_j were strictly greater than one, thus violating a key structure specified in Banker's (1993) tests. Their study considered an underlying production correspondence involving two inputs and two outputs, four different inefficiency distributions, two different measurement error distributions, and three different sample sizes (30, 60, and 90). There were 10 replications of each of 24 ($=4 * 2 * 3$) cells for a total of 240 data samples. Representative results for the high measurement error case (as defined in Banker, Gadh, and Gorr (1993)) appear in Table 2. In these cases, k = var(ℓn measurement error)/var(ℓn inefficiency) ranges between 0.1 and 3.4 depending on the different inefficiency distributions.

The most surprising result was that the new DEA-based tests outperformed both the parametric cost frontier-based tests and Welch's mean and Mann-Whitney tests in detecting difference in inefficiencies (Type I error) for the multiple-output case even in the presence of measurement errors. Welch's mean and Mann-Whitney tests were conservative and performed better than the new DEA-based tests and the paramtric cost frontier-based tests in terms of proportion of Type II errors with sample sizes of 30 and 60, but the new DEA-based tests performed better than the other tests for the sample size of 90.

Table 2a. Summary of statistical test results when the variance of log measurement error is 0.04. (Source: Banker and Chang (1995))

Panel A: True inefficiency is exponentially distributed with different parameters for the two groups

			Sample Size		
	Estimation Model	Test Procedure	N = 30	N = 60	N = 90
Banker's asymptotic DEA tests	BCC	exponential	5 (3)	10 (8)	7 (7)
		half-normal	6 (5)	10 (10)	8 (8)
Tests used in prior DEA literature	BCC	Welch	1 (0)	7 (4)	6 (5)
		Mann-Whitney	1 (1)	6 (5)	5 (4)
Parametric cost function-based test	translog	dummy variable	4 (4)	8 (8)	9 (9)

Panel B: True inefficiency is half-normally distributed with the different parameters for the two groups

			Sample Size		
	Estimation Model	Test Procedure	N = 30	N = 60	N = 90
Banker's asymptotic DEA tests	BCC	exponential	6 (5)	5 (5)	6 (4)
		half-normal	8 (6)	5 (5)	7 (7)
Tests used in prior DEA literature	BCC	Welch	1 (1)	5 (4)	5 (3)
		Mann-Whitney	3 (2)	2 (2)	3 (2)
Parametric cost function-based test	translog	dummy variable	4 (1)	4 (4)	5 (5)

Note: The number reported in each cell is the number of iterations (out of a total of 10 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number in parentheses in each cell indicates the number of iterations for which the test statistic is significant at the 5% level.

Table 2b. Summary of statistical test results when the variance of log measurement error is 0.04. (Source: Banker and Chang (1995))

Panel C: True inefficiency is exponentially distributed with same parameters for the two groups

	Estimation Model	Test Procedure	Sample Size		
			$N = 30$	$N = 60$	$N = 90$
Banker's asymptotic DEA tests	BCC	exponential	3 (3)	5 (4)	1 (1)
		half-normal	3 (3)	5 (5)	4 (4)
Tests used in prior DEA literature	BCC	Welch	2 (0)	4 (2)	1 (0)
		Mann-Whitney	2 (1)	3 (3)	1 (0)
Parametric cost function-based test	translog	dummy variable	2 (1)	4 (3)	3 (2)

Panel D: True inefficiency is half-normally distributed with the same parameters for the two groups

	Estimation Model	Test Procedure	Sample Size		
			$N = 30$	$N = 60$	$N = 90$
Banker's asymptotic DEA tests	BCC	exponential	3 (1)	2 (1)	3 (3)
		half-normal	3 (1)	2 (2)	4 (3)
Tests used in prior DEA literature	BCC	Welch	1 (1)	3 (1)	5 (4)
		Mann-Whitney	2 (1)	1 (1)	4 (4)
Parametric cost function-based test	translog	dummy variable	2 (1)	1 (0)	5 (5)

Note: The number reported in each cell is the number of iterations (out of a total of 10 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number in parentheses in each cell indicates the number of iterations for which the test statistic is significant at the 5% level.

Kittelsen (1995) performed a simulation study comparing the tests based on T_{EX} and T_{HN} as in (i) and (ii) above with Welch's means test. However, he considered only the case when the true underlying distribution of inefficiencies for both groups of DMUs was identical. As a result, his study evaluated only the proportion of Type II errors when the null hypothesis is falsely rejected. He performed 1,000 trials each for samples of size 20, 50, 100, 200, 500, and 1,000. His results indicated that the proportion of Type II errors for the new tests was 5-6% for samples of size 20 and reduced to 0% for samples of size 200 and greater. For all sample sizes, the occurrence of Type II errors was considerably less for these new tests than for Welch's means test. Thus, his results strongly support the conclusions drawn from the first Banker-Chang (1994) study.

Fixing the sample size at 100, Kittelsen then considered half-normal distributions with different means for the inefficiency variable, two other distributions (gamma and exponential) for the inefficiency variable, three different distributions for the output variable (normal, uniform, and lognormal), and three different production frontiers with 1, 2, or 3 inputs. The results were remarkably robust; the occurrence of Type II errors remained considerably lower for the new tests than for Welch's means test. (Kittelsen also conducted a matched pairs test that has no economic or statistical basis as there are not matched observations when testing this null hypothesis. Not surprisingly, Kittelsen's matched pair test rejected the null hypothesis (when it was true) in more than 99% of the cases!)

Kittelsen also evaluated Banker's (1993) two new tests and Welch's means test by estimating the inefficiencies separately for the two groups of DMUs after splitting the sample in each case so that the DMUs of one type are not used as referent observations for the other type. Not surprisingly, the proportion of Type II errors increased for all three tests with the smaller N available to recover the true production frontier; but the new tests continue to perform better than Welch's means test. The proportion of Type II errors increased more than the level indicated by the halving of the sample size; however, this was left unexplained in his study.

4. Tests of Returns to Scale

Existence of increasing or decreasing returns to scale is one of the most studied characteristics of production frontiers in economics because of its importance in many policy decisions. Different DEA models provide estimators for technical inefficiency under different assumptions about returns to scale. Since the empirical distribution of the DEA estimator of technical inefficiency asymptotically recovers the true distribution, we can construct different statistics to test hypotheses such as constant return to scale, based on assumptions about the true distribution of the technical inefficiency variable. In this section, I reproduce detailed materials from Banker and Chang (1993) that describe such tests for returns to scale based on the DEA estimator, and the results of a simulation study that compares their performance relative to conventional COLS-based tests of returns to scale.

Banker's (1993) four postulates are logically consistent with both increasing and decreasing returns to scale, they do not impose constant returns to scale. Next, impose the additional condition that the production set P exhibits constant returns to scale as embodied in postulate 6 below:

Postulate 6. *(Constant returns to scale)* If $(X, Y) \in P$ then $(rX, rY) \in P$ for all $r > 0$.

The DEA inefficiency estimator (denoted by $\hat{\theta}_j^C$) under the assumptions embodied in Postulates 1, 2, 3, 4, and 6, is obtained by solving the same linear program in (1) to (5) as before except that the constraint $\Sigma_{k=1}^{N} \lambda_k = 1$ in (4) is deleted. This corresponds to the so-called CCR model (Charnes, Cooper, and Rhodes 1978). Note that $\hat{\theta}_j^C \geq \hat{\theta}_j^B$ for all observations j because $\hat{\theta}_j^C$ is the solution to a less constrained linear program. It follows as in Banker (1993) that the estimator $\hat{\theta}^C$ is consistent under the assumptions of Postulates 1, 2, 3, 4, and 6, and the empirical distribution of $\hat{\theta}^C$ asymptotically recovers the true distribution of θ under these assumptions. Denote the distribution of $\hat{\theta}^C$ by $\hat{F}^C$.

To construct a test of the constant returns to scale hypothesis, maintain the assumptions in Postulates 1, 2, 3, and 4 under which the asymptotic empirical distribution of $\hat{\theta}^B$ is identical to the true distribution of θ. Further, under the null hypothesis of constant returns to scale as embodied in Postulate 6, in addition to the maintained assumptions in the other four postulates, the asymptotic empirical distribution of $\hat{\theta}^C$ also is identical to the true distribution of θ. This asymptotic correspondence between the empirical distributions of $\hat{\theta}^B$ and $\hat{\theta}^C$ under the null hypothesis of constant returns to scale motivates the following three test statistics analogous to those in Banker (1993):

(iv) Maintaining the additional assumption that θ is exponentially distributed, evaluate the test statistic $T_{EX} \equiv \Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$ relative to the critical value of the F-distribution with $(2N, 2N)$ degrees of freedom.

(v) Maintaining instead the assumption that θ is half-normally distributed, evaluate the test statistic $T_{HN} \equiv \Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$ relative to the critical value of the F-distribution with (N, N) degrees of freedom.

(vi) Maintaining no such assumptions about the inefficiency distribution, evaluate $T_{SM} = \max\{\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^B(\hat{\theta}_j^B) | j = 1, \ldots N\}$ as a Smirnov test statistic.

The above methods provide "two-sided" tests in the sense that the alternative hypothesis (to the null hypothesis of constant returns to scale) admits increasing or decreasing returns to scale, or both, in different regions of the production set. A one-sided test, with the alternative hypothesis of "decreasing returns to scale," can be constructed by considering the nondecreasing returns to scale assumption embodied in the following postulate:

Postulate 7. *(Nondecreasing returns to scale)* If $(X, Y) \in P$ then $(rX, rY) \in P$ for all $r \geq 1$.

The DEA inefficiency estimator denoted by $\hat{\theta}^E$ in this case, is obtained by solving the same linear programming formulation in (1) as before except that the constraint $\Sigma_{k=1}^{N} \lambda_k = 1$ is replaced by $\Sigma_{k=1}^{N} \lambda_k \geq 1$ as in Färe, Grosskopf, and Lovell (1985). Evidently this new linear program is less constrained than the first program to estimate θ_j^B, and more constrained than the second to estimate θ_j^C. Therefore, $\hat{\theta}_j^C \geq \hat{\theta}_j^E \geq \hat{\theta}_j^B$ for all observations $j = 1, \ldots N$. The inefficiency estimator $\hat{\theta}^E$ is weakly consistent under Postulates 1, 2, 3, 4, and 7, and its asymptotic empirical distribution recovers the true distribution of θ under these assumptions. If we maintain the assumptions embodied in Postulates 1, 2, 3, and 4, then we can test the null hypothesis of nondecreasing returns to scale in Postulate 7 against the alternative hypothesis of decreasing returns to scale exactly as before except that $\hat{\theta}^E$ and $\hat{F}^E(\hat{\theta}^E)$ are substituted for $\hat{\theta}^C$ and $\hat{F}^C(\hat{\theta}^C)$, respectively, in the earlier test statistics. We can develop tests of the null hypothesis of nonincreasing returns to scale against the alternative hypothesis now of increasing returns to scale in a similar manner. Note that if we keep $\hat{\theta}^C$ in the numerators of the earlier test statistics, and instead replace $\hat{\theta}^B$ in the denominators with $\hat{\theta}^E$ (i.e, treating nondecreasing returns as a maintained assumption) then we have tests of the null hypothesis of nonincreasing returns to scale against the alternative hypothesis of increasing returns to scale.

In their simulation experiment, Banker and Chang (1993) considered three different production frontiers, four different inefficiency distributions, and four different sample sizes for a total of 48 (=3 * 4 * 4) cells, performing 30 replications for each cell to yield 1440 (= 30 * 48) data samples. They considered three different Cobb-Douglas type single output two inputs production technologies specified as:

$$q = 10x_1^{0.6}x_2^{0.4} \tag{11}$$

$$q = 10x_1^{0.4}x_2^{0.3} \tag{12}$$

$$q = 10x_1^{0.8}x_2^{0.5} \tag{13}$$

Table 3a. Summary of returns to scale test results for a decreasing returns to scale production function. (Source: Banker and Chang (1993))

True inefficiency is exponentially distributed with mean inefficiency = 1.6667.

Test Procedure and Test Statistics		Hypothesis		Sample Sizes			
		Null	Alt.	$N = 50$	$N = 100$	$N = 150$	$N = 200$
DEA (exponential)	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	CRS	VRS	24 (13)	27 (17)	30 (30)	30 (30)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	NDRS	DRS	11 (3)	15 (8)	29 (26)	30 (25)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)$	NDRS	DRS	7 (2)	15 (7)	29 (22)	29 (25)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	NIRS	IRS	2 (1)	0 (0)	0 (0)	0 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)$	NIRS	IRS	1 (0)	0 (0)	0 (0)	0 (0)
DEA (half-normal)	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	CRS	VRS	20 (11)	23 (16)	30 (30)	30 (29)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	NDRS	DRS	13 (5)	16 (11)	30 (26)	30 (26)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)^2$	NDRS	DRS	8 (3)	15 (9)	29 (25)	29 (24)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	NIRS	IRS	2 (1)	0 (0)	0 (0)	0 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)^2$	NIRS	IRS	1 (1)	0 (0)	0 (0)	0 (0)
DEA (Smirnov)	$\text{Max}[\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^B(\hat{\theta}_j^B)\mid j = 1..,N]$	CRS	VRS	22 (8)	24 (17)	30 (29)	30 (30)
	$\text{Max}[\hat{F}^E(\hat{\theta}_j^E) - \hat{F}^B(\hat{\theta}_j^B)\mid j = 1..,N]$	NDRS	DRS	5 (3)	8 (3)	24 (15)	26 (22)
	$\text{Max}[\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^D(\hat{\theta}_j^D)\mid j = 1..,N]$	NDRS	DRS	5 (3)	9 (3)	24 (15)	26 (21)
	$\text{Max}[\hat{F}^D(\hat{\theta}_j^D) - \hat{F}^B(\hat{\theta}_j^B)\mid j = 1..,N]$	NIRS	IRS	1 (0)	1 (0)	0 (0)	1 (0)
	$\text{Max}[\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^E(\hat{\theta}_j^E)\mid j = 1..,N]$	NIRS	IRS	0 (0)	0 (0)	0 (0)	0 (0)
COLS	$\hat{\alpha}_1 + \hat{\alpha}_2 = 1$	CRS	VRS	10 (6)	16 (12)	23 (19)	23 (21)
	$\hat{\alpha}_1 + \hat{\alpha}_2 \geq 1$	NDRS	DRS	15 (15)	21 (16)	27 (23)	27 (23)
	$\hat{\alpha}_1 + \hat{\alpha}_2 \leq 1$	NIRS	IRS	1 (0)	0 (0)	0 (0)	0 (0)

Note: The number reported in each cell is the number of iterations (out of a total of 30 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number is parentheses in each cell indicates the number of CRS = constant returns to scale; VRS = variable returns to scale; DRS (IRS) = decreasing (increasing) returns to scale; NDRS (NIRS) = nondecreasing (nonincreasing) returns to scale.

where the inputs x_1 and x_2 were generated randomly for independent uniform probability distributions over the interval [5, 15].

Four different inefficiency distributions were considered. For the first inefficiency distribution, the random variable $\psi_j = \theta_j - 1 \geq 0$ was generated from a half-normal distribution $|N(0, 0.8356)|$ for all observations. For the second case, an exponential distribution with the same mean of 0.6667 was employed. For the third inefficiency distribution a half-normal distribution $|N(0.\ 0.2000)|$, and in the fourth case, an exponential distribution with a mean of 0.1595 were employed. Four different sample sizes (50, 100, 150, and 200) were considered. The same parametric functional form was used for COLS estimation as that used to generate the simulated production data in order to provide a challenging benchmark for the evaluation of the new DEA-based tests. Representative results appear in Tables 3(a) and 3(b). The proportion of Type I errors with the DEA-based tests was comparable

Table 3b. Summary of returns to scale test results for a decreasing returns to scale production function. (Source: Banker and Chang (1993))

True inefficiency is exponentially distributed with mean inefficiency = 1.6667.

Test Procedure and Test Statistics		Hypothesis Null	Alt.	Sample Sizes $N = 50$	$N = 100$	$N = 150$	$N = 200$
DEA (exponential)	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	CRS	VRS	17 (9)	26 (16)	27 (20)	30 (25)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	NDRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)$	NDRS	DRS	0 (0)	0 (0)	0 (0)	1 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)$	NIRS	IRS	16 (8)	26 (13)	26 (17)	27 (22)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)/\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)$	NIRS	IRS	15 (8)	26 (13)	26 (16)	29 (23)
DEA (half-normal)	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	CRS	VRS	11 (7)	21 (15)	27 (17)	30 (26)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	NDRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)^2$	NDRS	DRS	0 (0)	0 (0)	0 (0)	1 (0)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^D - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^B - 1)^2$	NIRS	IRS	10 (5)	20 (13)	25 (13)	27 (22)
	$\Sigma_{j=1}^{N}(\hat{\theta}_j^C - 1)^2/\Sigma_{j=1}^{N}(\hat{\theta}_j^E - 1)^2$	NIRS	IRS	10 (5)	20 (12)	24 (12)	29 (22)
DEA (Smirnov)	$\mathrm{Max}[\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^B(\hat{\theta}_j^B) \mid j = 1..,N]$	CRS	VRS	19 (16)	18 (8)	27 (19)	29 (26)
	$\mathrm{Max}[\hat{F}^E(\hat{\theta}_j^E) - \hat{F}^B(\hat{\theta}_j^B) \mid j = 1..,N]$	NDRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
	$\mathrm{Max}[\hat{F}^C(\hat{\theta}_j^C) - \hat{F}^D(\hat{\theta}_j^D) \mid j = 1..,N]$	NDRS	DRS	0 (0)	0 (0)	1 (0)	1 (1)
	$\mathrm{Max}[\hat{F}^D(\hat{\theta}_j^D) - \hat{F}^B(\hat{\theta}_j^B) \mid j = 1..,N]$	NIRS	IRS	15 (4)	12 (5)	25 (16)	27 (20)
COLS	$\hat{\alpha}_1 + \hat{\alpha}_2 = 1$	CRS	VRS	13 (9)	16 (13)	21 (20)	27 (25)
	$\hat{\alpha}_1 + \hat{\alpha}_2 \geq 1$	NDRS	DRS	0 (0)	0 (0)	0 (0)	0 (0)
	$\hat{\alpha}_1 + \hat{\alpha}_2 \leq 1$	NIRS	IRS	16 (13)	22 (16)	25 (21)	29 (27)

Note: The number reported in each cell is the number of iterations (out of a total of 30 iterations) for which the corresponding test statistic is significant at the 10% level. The corresponding number is parentheses in each cell indicates the number of CRS = constant returns to scale; VRS = variable returns to scale; DRS (IRS) = decreasing (increasing) returns to scale; NDRS (NIRS) = nondecreasing (nonincreasing) returns to scale.

to that for the COLS-based tests when the inefficiency distribution was exponential, but was worse when the distribution was half-normal. However, the frequency of Type II errors was lower for the DEA-based tests for both exponential and half-normal distributions, especially when the variance of the inefficiency distribution was high. Surprisingly, the DEA-based tests outperformed COLS-based tests even when the underlying production frontier violated the DEA assumption of concavity.

5. Tests of Input Substitutability

The materials in this section are drawn from Banker, Chang, and Sinha (1994). Implicit in the design of most operations control and evaluation systems are assumptions about the nature of the production system. A common assumption is the one embedded in the separate estimation or optimization of the requirements for different inputs (such as labor and materials) independent of the estimation or optimization for other inputs. The opposite assumption that one input is substitutable for another (such as capital for labor) is also common in the evaluation of production system alternatives. Separability of inputs is a common assumption in costing products for purposes of making pricing and mix decisions, and for designing products for manufacturability by costing alternative design specifications.[4]

Conventional econometric methods calculate point estimates of elasticities of substitution. For example, afer estimating a translog cost frontier, Allen elasticity of substitution, σ_{ij}, are estimated at the sample mean as:

$$\sigma_{ij} = \frac{A_{ij} + s_i s_j}{s_i s_j};\ i, j = 1, \ldots, I; j \neq i \tag{14}$$

where, A_{ij} is the estimated second-order input price coefficient in the translog cost frontier; and s_i and s_j are the cost shares of inputs i and j, respectively. The test of whether the inputs are separable is operationalized as a test of the null hypothesis $\sigma_{ij} = 0$ at the sample mean (Berndt and Wood 1975). In contrast, the new DEA-based tests developed by Banker, Chang, and Sinha (1994) evaluate the null hypothesis over the entire sample data. The four DEA postulates (1, 2, 3, and 4) are consistent with both input substitutability and separability, but do not impose either substitutability or separability. Refer to a production set that does not preclude input substitutability as P^{SUB}, and the corresponding inefficiency measure and its distribution functions as θ^{SUB} and F^{SUB}. In contrast, if the production set P exhibits input separability then it can be expressed as follows (Banker 1992):

$$P^{\text{SEP}} \equiv \{X, Y) | (x_i, Y) \in P_i,\ i = 1, \ldots, I\} \tag{15}$$

where

$$P_i \equiv \{(x_i, Y) | \text{input } x_i \text{ is adequate for the production of outputs } Y\}. \tag{16}$$

This generalizes the conventional Leontief technology that uses linear functions $g_i(Y) - x_i \leq 0$ to represent the relation $(x_i, Y) \in P_i$.

If each P_i and the corresponding probability density function $f_i(\theta_i)$ for input $i = 1, \ldots, I$, satisfy the four DEA postulates with X modified to read as x_i, then it can be shown that P^{SEP} and $f^{\text{SEP}}(\theta^{\text{SEP}})$ also satisfy the four postulates, where θ^{SEP} is the reciprocal of Shephard's (1970) distance measure given by:

$$\begin{aligned} \theta^{\text{SEP}}(X_o, Y_o) &\equiv \sup\{\theta | X_o/\theta, Y_o) \in P^{\text{SEP}}\} \\ &= \min\{\sup\{\theta_i | (x_{io}/\theta_i, Y_o) \in P_i\} | i = 1, \ldots, I\} \\ &= \min\{\theta_i(x_{io}, Y_o) | i = 1, \ldots, I\} \end{aligned} \tag{17}$$

and f^{SEP} and F^{SEP} are density and distribution functions respectively. Therefore, under the null hypothesis that the production set P is separable as in (15), the asymptotic empirical distributions of DEA estimates of θ^{SUB} and of θ^{SEP} are identical, with each retrieving the true distribution of θ.

The DEA estimates of $\hat{\theta}^{\text{SUB}}$ are obtained by solving the linear program for the DEA (Banker, Charnes, and Cooper 1984) model which admits variable returns to scale. The DEA estimates $\hat{\theta}_i(x_{io}, Y_o)$ for each separate set P_i are obtained from a similar linear program except that each X_k and X_j in (2) is replaced by x_{ik} and x_{ij} (and, of course, the right-hand sides of the constraints in (2) and (3) for the input inefficiency model are x_{ij}/θ and Y_j respectively.) Then, $\hat{\theta}^{\text{SEP}}(X_j, Y_j) = \min\{\hat{\theta}_i(x_{ij}, Y_j) | i = 1, \ldots, I\}$. It should be noted that $\hat{\theta}_i^{\text{SEP}}$ is determined from a less constrained mathematical program, and therefore, $\hat{\theta}^{\text{SEP}} \geq \hat{\theta}^{\text{SUB}}$ for all $j = 1, \ldots, N$.

Based on the asymptotic correspondence between the empirical distributions of $\hat{\theta}^{\text{SUB}}$ and $\hat{\theta}^{\text{SEP}}$ under the null hypothesis of input separability we can construct the following three test statistics to test the input separability hypothesis analogous to those described earlier:

$$\text{(vii)}\quad T_{EX} \equiv \sum_{j=1}^{N} (\hat{\theta}_j^{\text{SEP}} - 1) / \sum_{j=1}^{N} (\hat{\theta}_j^{\text{SUB}} - 1)$$

$$\text{(viii)}\quad T_{HN} \equiv \sum (\hat{\theta}_j^{\text{SEP}} - 1)^2 / \sum_{j=1}^{N} (\hat{\theta}_j^{\text{SUB}} - 1)^2_{j=1}^{N}$$

$$\text{(ix)}\quad T_{SM} \equiv \max\{\hat{F}^{\text{SUB}}(\hat{\theta}_j^{\text{SUB}}) - \hat{F}^{\text{SEP}}(\hat{\theta}_j^{\text{SEP}}) | j = 1, \ldots, N\}$$

Rejection of the null hypothesis indicates that the sample data do not support the input separability hypothesis under the maintained assumptions embodied in the four DEA postulates specified by Banker (1993).

Banker, Chang, and Sinha (1994) generated 480 data sets considering two different production technologies (separable and substitutable inputs), two parametric inefficiency distributions (exponential and half-normal), two levels of mean inefficiencies (high and low), and three sample sizes (50, 100, and 150), and replicated each of the 24(= 2 * 2 * 2 * 3) settings 20 times to yield 480 data sets. They used the following two production frontiers:

$$q_{\text{SUB}} = 10x_1^{0.6}\, x_2^{0.4} \tag{18}$$

$$q_{\text{SEP}} = \text{Min}\{x_1, x_2\} \tag{19}$$

where x_1 and x_2 were drawn randomly and independently from uniform probability distributions over the interval [5, 15].

Representative results of the simulation experiments appear in Tables 4a and 4b. The DEA-based tests outperform the COLS-based test when outputs (q_{SEP}) are generated using the Leontief production technology in (19) that assumes separability of inputs. In fact, the performance of the COLS-based test worsens for larger samples with the number of cases for which the null hypothesis of input separability is falsely rejeced increasing with the increase in sample size. When outputs (q_{SUB}) are generated using the Cobb-Douglas production technology in (18) that assumes substitutability of inputs, the performances of the

Table 4a. Summary of test results for a Leontief-type (separable) production function. (Source: Banker, Chang and Sinha (1994))

True inefficiency is half-normally distributed with mean inefficiency = 1.6667.

Test Procedure	Test Statistics	Hypothesis		Sample Sizes		
		Null	Alt.	$N = 5$	$N = 100$	$N = 150$
1 DEA (exponential)	$T_{EXP} \equiv \Sigma_{j=1}^{N} (\hat{\theta}_j^{SEP} - 1)/\Sigma_{j=1}^{N} (\hat{\theta}_j^{SUB} - 1)$	SEP	SUB	0 (0)	0 (0)	0 (0)
2 DEA (half-normal)	$T_{HN} \equiv \Sigma_{j=1}^{N} (\hat{\theta}_j^{SEP} - 1)^2/\Sigma_{j=1}^{N} (\hat{\theta}_j^{SUB} - 1)^2$	SEP	SUB	0 (0)	0 (0)	0 (0)
3 DEA (Smirnov)	$T_{SM} \equiv \text{Max}[F^{SEP}(\hat{\theta}_j^{SEP}) - F^{SUB}(\hat{\theta}_j^{SUB})\| j = 1,..N]$	SEP	SUB	0 (0)	0 (0)	0 (0)
4 COLS	$T_{COLS} \equiv \hat{\sigma}_{12}/\hat{S}_{\sigma}$	SEP	SUB	12 (10)	16 (12)	15 (15)

Notes: The number reported in each cell is the number of iterations (out of a total 20 iterations) for which the corresponding test statistic is significant at the 10 (5)% level.

SEP = input separability; SUB = substitutability.

Table 4b. Summary of test results for a Cobb-Douglas (substitutable) production function. (Source: Banker, Chang and Sinha (1994))

True inefficiency is half-normally distributed with mean inefficiency = 1.6667.

Test Procedure	Test Statistics	Hypothesis		Sample Sizes		
		Null	Alt.	$N = 5$	$N = 100$	$N = 150$
1 DEA (exponential)	$T_{EXP} \equiv \Sigma_{j=1}^{N} (\hat{\theta}_j^{SEP} - 1)/\Sigma_{j=1}^{N} (\hat{\theta}_j^{SUB} - 1)$	SEP	SUB	13 (7)	20 (19)	20 (20)
2 DEA (half-normal)	$T_{HN} \equiv \Sigma_{j=1}^{N} (\hat{\theta}_j^{SEP} - 1)^2/\Sigma_{j=1}^{N} (\hat{\theta}_j^{SUB} - 1)^2$	SEP	SUB	14 (10)	20 (20)	20 (20)
3 DEA (Smirnov)	$T_{SM} \equiv \text{Max}[F^{SEP}(\hat{\theta}_j^{SEP}) - F^{SUB}(\hat{\theta}_j^{SUB})\| j = 1,..N]$	SEP	SUB	15 (9)	16 (14)	20 (19)
4 COLS	$T_{COLS} \equiv \hat{\sigma}_{12}/\hat{S}_{\sigma}$	SEP	SUB	12 (10)	16 (13)	15 (13)

Notes: The number reported in each cell is the number of iterations (out of a total 20 iterations) for which the corresponding test statistic is significant at the 10 (5)% level.

SEP = input separability; SUB = substitutability.

DEA-based tests and COLS-based test are comparable for the small sample size. However, with an increase in the size of the samples the DEA-based tests outperform the COLS-based test.

Overall, the results of the simulation experiments suggest that even when the parametric form specified for the COLS estimation is the same as the true production frontier used to simulate the data, the DEA-based tests perform very satisfactorily compared to the COLS-based test in distinguishing between separable and substitutable forms of input characterization. The DEA-based tests are robust and perform well regardless of the form of the true production technology and inefficiency distribution. The experiments also indicate that the performance of the tests improves as the sample size increases and as the level of mean inefficiency decreases. There are no substantial differences in performance across the three DEA-based tests.

6. Some Tests of Model Specification

The materials in this section are drawn from Banker, Devaraj, and Sinha (1995). A question that is of considerable interest in many applications is whether a set of variables is significant at the margin in characterizing the production correspondence between inputs and outputs. For specificity, let X and Y represent vectors of input and output variables respectively in the base model, and let Z represent the vector of additional input variables whose significance we wish to evaluate. (While Z is specified here as a vector of input variables, the analysis can be adapted easily to address the case when some of these variables are outputs.) We can estimate $\hat{\theta}^B(X_j, Y_j)$ and $\hat{\theta}^B(X_j, Z_j; Y_j)$ using the BCC-model in (1) to (5) first with only X and then with (X, Z) as input vectors, respectively. Evidently, $\hat{\theta}^B(X_j; Y_j) \geq \hat{\theta}^B(X_j, Z_j; Y_j)$. Therefore, under the maintained assumptions embodied in the four DEA postulates corresponding to the input vector (X, Z), and the null hypothesis that input variables Z do not influence the production correspondence, and proceeding as before, we can construct the following test statistics:

$$\text{(x)} \quad T_{EX} = \left[\sum_{j=1}^{N} \hat{\theta}^B(X_j; Y_j) - 1 \right] \Big/ \left[\sum_{j=1}^{N} \hat{\theta}^B(X_j, Z_j; Y_j) - 1 \right]$$

$$\text{(xi)} \quad T_{HN} = \left[\sum_{j=1}^{N} \hat{\theta}^B(X_j; Y_j) - 1)^2 \right] \Big/ \left[\sum_{j=1}^{N} \hat{\theta}^B(X_j, Z_j; Y_j) - 1)^2 \right]$$

$$\text{(xii)} \quad T_{SM} = \max\{\hat{F}_x^B(\hat{\theta}^B(X_j; Y_j)) - \hat{F}_{x,z}^B(\hat{\theta}^B(X_j, Z_j; Y_j)) | j = 1, \ldots N\}.$$

Simulation studies designed in a manner similar to those described in the earlier sections indicate that these tests perform as well or better than COLS-based tests even when the estimated parametric form for COLS estimation is identical to the one used to generate the simulated data. Representative results appear in Tables 5a and 5b.

The general approach described here can be applied to develop a wide variety of tests of model specification. For instance, let $z_j = z(Z_j)$ be an aggregation of the vector Z. Is this aggregate sufficient in capturing the impact of the variables Z on the production correspondence? Such a hypothesis can be evaluated by comparing the inefficiencies estimated with (X, z) as the input vector against the inefficiencies estimated with (X, z, Z) as the input vector, and proceeding using the test statistics described above.

Table 5a. Summary of test results. Test of the marginal effect (number of Type II errors).

True inefficiency is exponentially distributed with mean inefficiency = 1.18.

Test Procedure		Sample Size 50	100	150
DEA-based tests	exponential	1 (1)	0 (0)	0 (0)
	half-normal	1 (2)	0 (0)	0 (0)
COLS test		5 (6)	4 (9)	4 (8)

Notes: True production frontier: $y = x_1^{0.65}$
H_o: x_2 has no marginal effect on production.

The number reported in each cell is the number of iterations (out of 100) for which the test statistic is significant at the 5% level (numbers in parentheses are at 10%).

Table 5b. Summary of test results. Test of the marginal effect (100 less number of Type I errors).

True inefficiency is exponentially distributed with mean inefficiency = 1.18.

Test Procedure		Sample Size 50	100	150
DEA-based tests	exponential	89 (96)	95 (100)	96 (100)
	half-normal	69 (83)	73 (96)	82 (98)
COLS test		67 (78)	93 (98)	96 (98)

Notes: True production frontier: $y = x_1^{0.65}\, x_2^{0.15}$
H_o: x_2 has no marginal effect on production.

The number reported in each cell is the number of iterations (out of 100) for which the test statistic is significant at the 5% level (numbers in parentheses are at 10%).

7. Concluding Remarks

The principal limitation of the different DEA-based tests described here is that for any finite sample, the random variates in the numerator or the denominator of the test statistics T_{EX} and T_{HN} need not be distributed as chi-squared. Also, they may not be distributed independently of each other and, therefore, their ratio need not follow the F-distribution. A similar caveat applies also to the Smirnov-type test described here.

An interesting possibility is to use bootstrapping methods to augment the tests described in this survey. For instance, since the DEA estimator of inefficiency is biased downward in finite samples (Banker 1993, Proposition 4), it may be possible to reduce the bias using bootstrapping methods, and construct these tests based on the bootstrapping estimators. A caveat is in order, however. Standard proofs for the consistency of the bootstrap estimator do not apply for the DEA-type nonparametric estimation of general monotone and concave production frontiers because of the incidental parameters problem described by Banker (1993, p. 1266). Thus, much remains to be accomplished in providing a theoretical basis for applying the bootstrapping methods to nonparametric estimation of general monotone and concave production frontiers. Furthermore, similar to the DEA-based tests described in this paper, the distribution of the bootstrap estimator is not known for finite samples. Unlike these DEA-based tests, however, the performance of the bootstrap estimator in the DEA context has not been evaluated, to date, using systematic simulation studies.

It is important, therefore, to evaluate the comparative performance of these tests for finite sample sizes using simulation studies. The initial Monte Carlo studies mentioned here seem to be promising for the tests described in this paper. However, any such evidence must be interpreted cautiously as no simulation study exhausts all possibilities. We must continue similar experimentation to identify conditions under which the DEA-based tests perform well and conditions under which they do not (Banker et al. 1987).[5]

Many interesting directions remain open for extending this line of research. If, for instance, the expected value of DMU inefficiency is modeled as a linear function of a set of explanatory variables, then consistent estimators of the parameters of this function are obtained by a two-stage procedure, first obtaining DEA inefficiency estimates and then regressing them (or appropriate transforms) on the explanatory variables (Theil, 1971; Banker and Johnston, 1995; Gstach, 1995). While the performance of this procedure remains to be evaluated in simulation studies, it suggests how the consistency property of the DEA inefficiency estimator can be employed to estimate a wide variety of models.

Acknowledgments

This paper has benefited from the comments of Leopold Simar and other participants in the Efficiency Measurement Research Workshop at Odense University in May 1995.

Notes

1. Alternative specifications based on output quantities and input mix proportion variables, or on endogenous input mix decisions based on random input price variables can be also employed to obtain results similar to those described here.
2. The postulates are specified here for the multiple-input and multiple-output case with a multiplicative inefficiency term modeled as a random variable. The extension from Banker (1993), where the single-output case is described with an additive inefficiency term, is straightforward.
3. The proof of Theorem 6 in Banker (1993) also applies equally well to any parametric estimator of an extremal production function obeying the consistency property. That is, under Postulates 3 and 4, the empirical distribution of inefficiency scores evaluated relative to such a parametrically estimated production frontier also recovers the true distribution of θ.
4. Separability of inputs here refers to the determination of the demand for each input (given output quantities and input prices) independent of the levels of all other inputs. While this is consistent with common usage in accounting and operations management, it does not correspond to the concept of separability of production frontiers in economics.
5. It is important, therefore, to allocate time and effort in conducting simulation studies to evaluate performance under a large number of different conditions rather than conducting a large number of replications for relatively few conditions.

References

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

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

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

Banker, R.D. (1992). "Selection of Efficiency Evaluation Models." *Contemporary Accounting Research* Fall 343–355.

Banker, R.D. (1993). "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation." *Management Science*, Oct. 1265–1273.

Banker, R.D. and H. Chang. (1993). "Tests of Returns to Scale for Monotone Concave Production Functions."

Banker, R.D. and H. Chang. (1994). "A Monte Carlo Evaluation of Hypothesis Tests for Differences in Efficiencies." *International Journal of Production Economics*.

Banker, R.D. and H. Chang. (1995). "A Simulation Study of Hypothesis Tests for Differences in Efficiencies for the Case of Multiple Outputs and Measurement Error." 1995.

Banker, R.D., H. Chang, and K.K. Sinha. (1994). "Tests to Evaluate the Separability or Substitutability of Inputs to a Production System."

Banker, R.D., A. Charnes, and W.W. Cooper. (1984). "Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* Sept. 1078–1092.

Banker, R.D., A. Charnes, W.W. Cooper, and A. Maindiratta. (1987). "A Comparison of Data Envelopment Analysis and Translog Estimates of Production Frontiers Using Simulated Observations from a known Technology." In A. Dogramaci and R. Färe (eds.), *Applications of Modern Production Theory in Efficiency and Productivity*. Boston: Kluwer Academic Publishers.

Banker, R.D., R. Conrad, and R. Strauss. (1986). "A Comparative Application of Data Envelopment Analysis and Translog Methods: An Illustrative Study of Hospital Production," *Management Science* Jan. 31–44.

Banker, R.D., S. Devaraj, and K.K. Sinha. (1995). "Some Tests of Model Specification in Data Envelopment Analysis."

Banker, R.D., V. Gadh, and W. Gorr. (1993). "A Monte Carlo Comparison of Two Production Frontier Estimation Methods: Corrected Ordinary Least Squares and Data Envelopment Analysis." *European Journal of Operational Research* 332–343.

Banker, R.D. and H.H. Johnston. (1995). "Evaluating the Impacts of Operating Strategoes on Efficiency in the U.S. Airline Industry." In A. Charnes, W.W. Cooper, A. Lewin, and L. Seiford (eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*, Boston: Kluwer Academic Publishers.

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

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

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

Berndt, E. and D. Wood. (1975). "Technology, Prices and the Derived Demand for Energy." *Review of Economics and Statistics* 259–268.

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

Charnes, A., W.W. Cooper, and E. Rhodes. (1981). "Program Evaluation and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through." *Management Science* 668–697.

Färe, R., S. Grosskopf, and C.A.K. Lovell. (1985). *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff.

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

Gong, B. and R.C. Sickles. (1992). "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data." *Journal of Econometrics* 259–284.

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

Grosskopf, S. and V. Valdmanis. (1987). "Measuring Hospital Performance: A Nonparametric Approach." *Journal of Health Economics* 89–107.

Gstach, D. (1995). "Efficiency of Austrian Combined Transport: Application of DEA in a Statistical Setting."

Kittelsen, S.A.C. (1995). "Monte Carlo Simulations of DEA Efficiency Measures and Hypothesis Tests."

Korostelev, A.P., L. Simar, and A.B. Tsybakov. (1995). "On Estimation of Monotone and Convex Boundaries." *Public Institute of Statistics of University of Paris.* pp. 3–18.

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

Mensah, Y.M. and S. Li. (1993). "Measuring Production Efficiency in a Not-for-Profit Setting: An Extension." *The Accounting Review* Jan. 66–88.

Olson, J.A., P. Schmidt, and D.A. Waldman. (1980). "A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions." *Journal of Econometrics* 67–82.

Schmidt, P. (1976). "On the Statistical Estimation of Parametric Frontier Production Functions." *Review of Economics and Statistics*, May 238–339.

Schmidt, P. (1985). "Frontier Production Functions." *Econometric Review*, 289–328.

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

Siegel, S. and N.J. Castellan Jr. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Theil, Henri. (1971). *Principles of Econometrics*. New York: John Wiley & Sons, Inc.

Varian, H.R. (1978). "The Nonparametric Approach to Production Analysis." *Econometrica*, 579–597.