Chapter 11

STATISTICAL TESTS BASED ON DEA EFFICIENCY SCORES

Rajiv D. Banker[1] and Ram Natarajan[2]
School of Management, The University of Texas at Dallas, Richardson, TX 75083-0688 USA
email: [1] rbanker@utdallas.edu [2]nataraj@utdallas.edu

Abstract: This chapter is written for analysts and researchers who may use Data Envelopment Analysis (DEA) to statistically evaluate hypotheses about characteristics of production correspondences and factors affecting productivity. Contrary to some characterizations, it is shown that DEA is a full-fledged statistical methodology, based on the characterization of DMU efficiency as a stochastic variable. The DEA estimator of the production frontier has desirable statistical properties, and provides a basis for the construction of a wide range of formal statistical tests (Banker 1993). Specific tests described here address issues such as comparisons of efficiency of groups of DMUs, existence of scale economies, existence of allocative inefficiency, separability and substitutability of inputs in production systems, analysis of technical change and productivity change, impact of contextual variables on productivity and the adequacy of parametric functional forms in estimating monotone and concave production functions.

Key words: Data envelopment analysis (DEA); Statistical tests

1. INTRODUCTION

Data Envelopment Analysis (DEA) continues to be used extensively in many settings to analyze factors influencing the efficiency of organizations. The DEA approach specifies the production set only in terms of desirable properties such as convexity and monotonicity, without imposing any parametric structure on it (Banker, Charnes and Cooper, 1984). Despite its widespread use, many persons continue to classify DEA as a non-statistical approach. However, many recent advances have established the statistical properties of the DEA efficiency estimators. Based on statistical representations of DEA, rigorous statistical tests of various hypotheses have also been developed.

To start with, Banker (1993) provided a formal statistical foundation for DEA by identifying conditions under which DEA estimators are statistically consistent and maximize likelihood. He also developed hypothesis tests for efficiency comparison when a group of DMUs (decision making units) is compared with another. Since the publication of Banker (1993), a number of significant advances have been made in developing DEA-based hypothesis tests to address a wide spectrum of issues of relevance to users of DEA. These include issues such as efficiency comparison of groups, existence of scale inefficiency, impact of contextual variables on productivity, adequacy of parametric functional forms in estimating monotone and concave production functions, examination of input separability and substitutability in production systems, existence of allocative inefficiency and evaluation of technical change and productivity change. In the rest of this chapter, we describe different DEA-based tests of hypotheses in the form of a reference list for researchers and analysts interested in applying DEA to efficiency measurement and production frontier estimation[1].

The paper proceeds as follows. In the next section, section 2, we present statistical tests that are relevant for applications of DEA in environments where the deviation from the frontier is caused by a single one-sided stochastic random variable representing DMU inefficiency. We describe salient aspects of the statistical foundation provided in Banker (1993), discuss hypothesis tests for efficiency comparison of groups of DMUs, for the existence of scale inefficiency or allocative inefficiency and for input substitutability. In section 3, we address situations where the production frontier shifts over time. We describe DEA-based techniques and statistical tests to evaluate and test for existence of productivity and technical change over time. In section 4, we discuss the application of DEA in environments where the deviation from the production frontier arises as a result of two stochastic variables, one representing inefficiency and the other random noise. We explain how efficiency comparisons across groups of DMUs can be carried out, and describe DEA-based methods to determine the impact of contextual or environmental variables on inefficiency. We also suggest methods to evaluate the adequacy of an assumed parametric form for situations where prior guidance specifies only monotonicity and concavity for a functional relationship. Finally, we summarize and conclude in section 5.

[1] Bootstrapping, discussed in chapter 10, offers an alternative approach for statistical inference within nonparametric efficiency estimation.

2. HYPOTHESIS TESTS WHEN INEFFICIENCY IS THE ONLY STOCHASTIC VARIABLE

The tests described in this section build on the work in Banker (1993) that provides a formal statistical basis for DEA estimation techniques. We briefly describe the salient aspects of Banker (1993) before discussing the hypothesis tests.

2.1 Statistical Foundation for DEA

Consider observations on j = 1,….N decision making units (DMUs), each observation comprising a vector of outputs $\mathbf{y}_j \equiv (y_{1j}, \ldots y_{Rj}) \geq 0$ and a vector of inputs $\mathbf{x}_j \equiv (x_{1j}, \ldots x_{Ij}) \geq 0$, where $y \in Y$ and $\mathbf{x} \in X$, and Y and X are convex subsets of $\Re^R$ and $\Re^I$, respectively. Input quantities and output mix proportion variables are random variables[2]. The production correspondence between the frontier output vector $\mathbf{y}^0$ and the input vector $\mathbf{x}^0$ is represented by the production frontier $g(\mathbf{y}^0, \mathbf{x}^0) = 0$. The support set of the frontier is a monotonically increasing and convex set $T \equiv \{(\mathbf{y}, \mathbf{x}) \mid \mathbf{y}$ can be produced from $\mathbf{x}\}$. The inefficiency of a specific DMU j is, $\theta_j \equiv \max \{\theta \mid (\theta \mathbf{y}_j, \mathbf{x}_j) \in T\}$. It is modeled as a scalar random variable that takes values in the range $[1, \infty)$ and is distributed with the probability density function $f(\theta)$ in this range. Banker (1993) imposes additional structure on the distribution of the inefficiency variable requiring that there is a non-zero likelihood of nearly efficient performance i.e., $\int_1^{1+\delta} f(\theta)\, d\theta > 0$ for all $\delta > 0$.

In empirical applications of DEA, the inefficiency variable θ is not observed and needs to be estimated from output and input data. The following Banker, Charnes and Cooper (BCC 1984) linear program is used to estimate the inefficiency:

$$\hat{\theta}_j = \text{argmax} \left\{ \theta \left| \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1,..R; \sum_{k=1}^{N} \lambda_k x_{ik} \leq x_{ij}, \forall i = 1,...I; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,..N \end{array} \right. \right\} \quad (11.1)$$

where $\hat{\theta}_j$ is the DEA estimator of θ_j.

Modeling the inefficiency deviation as a stochastic variable that is distributed independently of the inputs, enabled Banker (1993) to derive several results that provide a statistical foundation for hypothesis tests using

[2] Alternative specifications that specify output quantities and input mix proportion as random variables or endogenous input mix decisions based on input prices modeled as random variables can also be used.

DEA. He demonstrated that the DEA estimators of the true inefficiency values maximize likelihood provided the density function of the inefficiency random variable f(θ) is monotone decreasing. He also pointed out that a broad class of probability distributions, including the exponential and half-normal distributions, possesses monotone decreasing density functions. Banker (1993) also shows that the DEA estimator of the inefficiency underestimates the true inefficiency in finite samples. More importantly, he shows that asymptotically this bias reduces to zero; that is the DEA estimators are consistent if the probability of observing nearly efficient DMUs is strictly positive[3].

While consistency is a desirable property of an estimator, it does not by itself guide the construction of hypothesis tests. However, Banker (1993) exploits the consistency result to prove that for "large" samples the DEA estimators of inefficiency for any given subset of DMUs follow the same probability distribution as the true inefficiency random variable. This is, perhaps, the most important result in the Banker (1993) paper since it implies that, for large samples, distributional assumptions imposed for the true inefficiency variable can be carried over to the empirical distribution of the DEA estimator of inefficiency and test-statistics based on the DEA estimators of inefficiency can be evaluated against the assumed distribution of the true inefficiency.

2.2 Efficiency Comparison of Two Groups of DMUs

The asymptotic properties of the DEA inefficiency estimator are used by Banker (1993) to construct statistical tests enabling a comparison of two groups of DMUs to assess whether one group is more efficient than the other. Banker (1993) proposes parametric as well as nonparametric tests to evaluate the null hypothesis of no difference in the inefficiency distributions of two sub-samples, G_1 and G_2, that are part of the sample of N DMUs when the sample size, N, is large. For N_1 and N_2 DMUs in subgroups G_1 and G_2, respectively, the null hypothesis of no difference in inefficiency between the two subgroups can be tested using the following procedures:

[3] The consistency result does not require that the probability density function f(θ) be monotone decreasing. It only requires that there is a positive probability of observing nearly efficient DMUs, which is a much weaker condition than the monotonicity condition required for the DEA estimators to be maximum likelihood.

(i) If the logarithm[4] of the true inefficiency θ_j is distributed as exponential over $[0, \infty)$ for the two subgroups, then under the null hypothesis that there is no difference between the two groups, the test statistic is calculated as $\left[\sum_{j\in G_1} \ln(\hat{\theta}_j)/N_1\right] \Big/ \left[\sum_{j\in G_2} \ln(\hat{\theta}_j)/N_2\right]$ and evaluated relative to the critical value of the F distribution with $(2N_1, 2N_2)$ degrees of freedom.

(ii) If logarithm of the true inefficiency θ_j is distributed as half-normal over the range $[0, \infty)$ for the two subgroups, then under the null hypothesis that there is no difference between the two groups, the test statistic is calculated as $\left[\sum_{j\in G_1} \left\{\ln(\hat{\theta}_j)\right\}^2 /N_1\right] \Big/ \left[\sum_{j\in G_2} \left\{\ln(\hat{\theta}_j)\right\}^2 /N_2\right]$ and evaluated relative to the critical value of the F distribution with (N_1, N_2) degrees of freedom.

(iii) If no such assumptions are maintained about the probability distribution of inefficiency, a non-parametric Kolmogorov-Smirnov's test statistic given by the maximum vertical distance between $F^{G_1}(\ln(\hat{\theta}_j))$ and $F^{G_2}(\ln(\hat{\theta}_j))$, the empirical distributions of $\ln(\hat{\theta}_j)$ for the groups G_1 and G_2, respectively, is used. This statistic, by construction, takes values between 0 and 1 and a high value for this statistic is indicative of significant differences in inefficiency between the two groups.

2.3 Tests of Returns to Scale

Examining the existence of increasing or decreasing returns to scale is an issue of interest in many DEA studies. We provide a number of DEA-based tests to evaluate returns to scale using DEA inefficiency scores[5]. Consider the inefficiency $\hat{\theta}_j^C$ estimated using the CCR (Charnes, Cooper and Rhodes 1978) model obtained from the BCC linear program in (11.1) by deleting the constraint $\sum_{k=1}^{N} \lambda_k = 1$ i.e.,

$$\hat{\theta}_j^C = \operatorname{argmax}\left\{\theta \,\middle|\, \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_{rk} \geq \theta y_{rj}, \forall r = 1,\ldots R; \sum_{k=1}^{N} \lambda_k x_{ik} \leq x_{ij}, \forall i = 1,\ldots I; \\ \lambda_k \ \geq \ 0, \forall k = 1,\ldots N \end{array}\right\} \tag{11.2}$$

[4] Alternatively, the assumption may be maintained that $t(\theta_j)$ is distributed as exponential over $[0, \infty)$ where t(.) is some specified transformation function. Then the test statistic is given by $\left[\sum_{j\in G_1} t(\hat{\theta}_j)/N_1\right] \Big/ \left[\sum_{j\in G_2} t(\hat{\theta}_j)/N_2\right]$.

[5] Chapter 2 of this handbook provides a detailed discussion of qualitative and quantitative aspects of returns to scale in DEA.

By construction, $\hat{\theta}_j^C \geq \hat{\theta}_j$. Scale inefficiency is then estimated as $\hat{\theta}_j^S = \hat{\theta}_j^C / \hat{\theta}_j$. Values of scale inefficiency significantly greater than 1 indicate the presence of scale inefficiency to the extent operations deviate from the *most productive scale size (MPSS)* (Banker 1984, Banker, Charnes and Cooper 1984). All observations in the sample are scale efficient if and only if the sample data can be rationalized by a production set exhibiting constant returns to scale.

Under the null hypothesis of no scale inefficiency (or equivalently, under the null hypothesis of constant returns to scale), $\hat{\theta}_j^C$ is also a consistent estimator of θ_j (Banker 1993, 1996). The null hypothesis of no scale inefficiency in the sample can be evaluated by constructing the following test statistics:

(i) If the logarithm of the true inefficiency θ_j is distributed as exponential over $[0, \infty)$, then under the null hypothesis of constant returns to scale, the test statistic is calculated as $\sum \ln(\hat{\theta}_j^C) \big/ \sum \ln(\hat{\theta}_j)$. This test statistic is evaluated relative to the half-F distribution $|F_{2N,2N}|$ with 2N,2N degrees of freedom over the range $[1,\infty)$, since by construction the test statistic is never less than 1. The half-F distribution is the F distribution truncated below at 1, the median of the F distribution when the two degrees of freedom are equal.

(ii) If the logarithm of the true inefficiency θ_j is distributed as half-normal over the range $[0, \infty)$, then under the null hypothesis of constant returns to scale, the test statistic is calculated as $\sum \left\{\ln(\hat{\theta}_j^C)\right\}^2 \Big/ \sum \left\{\ln(\hat{\theta}_j)\right\}^2$ and evaluated relative to the half-F distribution $|F_{N,N}|$ with N,N degrees of freedom over the range $[1,\infty)$.

(iii) If no such assumptions are maintained about the probability distribution of inefficiency, a non-parametric Kolmogorov-Smirnov's test statistic given by the maximum vertical distance between $F^C(\ln(\hat{\theta}_j^C))$ and $F(\ln(\hat{\theta}_j))$, the empirical distributions of $\ln(\hat{\theta}_j^C)$ and $\ln(\hat{\theta}_j)$, respectively, is used. This statistic, by construction, takes values between 0 and 1 and a high value for this statistic is indicative of the existence of significant scale inefficiency in the sample.

The above tests evaluate the null hypothesis of constant returns to scale against the alternative of variable returns to scale. In addition, it is also possible to test the null hypothesis of non-decreasing returns to scale against the alternative of decreasing returns to scale and the null hypothesis of non-increasing returns to scale against the alternative of increasing returns to scale. Two additional inefficiency estimators $\hat{\theta}_j^D$ and $\hat{\theta}_j^E$ required for these tests are calculated by solving the program in (1) after changing the constraint $\sum_{k=1}^{N}\lambda_k = 1$ to $\sum_{k}\lambda_k \le 1$ for $\hat{\theta}_j^D$ and to $\sum_{k}\lambda_k \ge 1$ for $\hat{\theta}_j^E$. By construction, $\hat{\theta}_j^C \ge \hat{\theta}_j^D \ge \hat{\theta}_j$ and $\hat{\theta}_j^C \ge \hat{\theta}_j^E \ge \hat{\theta}_j$.

The following is a test of the null hypothesis of non-decreasing returns to scale against the alternative of decreasing returns to scale:

(i) If the logarithm of the true inefficiency θ_j is distributed as exponential over $[0, \infty)$, the test statistic is calculated as $\sum \ln(\hat{\theta}_j^E) \big/ \sum \ln(\hat{\theta}_j)$ or $\sum \ln(\hat{\theta}_j^C) \big/ \sum \ln(\hat{\theta}_j^D)$. Each of these statistics is evaluated relative to the half-F distribution $|F_{2N,2N}|$ with 2N,2N degrees of freedom over the range $[1,\infty)$.

(ii) If the logarithm of the true inefficiency θ_j is distributed as half-normal over the range $[0, \infty)$, the test statistic is calculated as either $\sum\left\{\ln(\hat{\theta}_j^E)\right\}^2 \Big/ \sum\left\{\ln(\hat{\theta}_j)\right\}^2$ or $\sum\left\{\ln(\hat{\theta}_j^C)\right\}^2 \Big/ \sum\left\{\ln(\hat{\theta}_j^D)\right\}^2$ and evaluated relative to the half-F distribution $|F_{N,N}|$ with N,N degrees of freedom over the range $[1,\infty)$.

(iii) If no such assumptions are maintained about the probability distribution of inefficiency, a non-parametric Kolmogorov-Smirnov's test statistic given by either the maximum vertical distance between $F^E(\ln(\hat{\theta}_j^E))$ and $F(\ln(\hat{\theta}_j))$, or that between $F^C(\ln(\hat{\theta}_j^C))$ and $F^D(\ln(\hat{\theta}_j^D))$ is used.

The test statistics for testing the null of non-increasing returns to scale against the alternative of increasing returns to scale can be developed in a similar fashion by interchanging $\hat{\theta}_j^E$ and $\hat{\theta}_j^D$ in the statistics above.

2.4 Tests of Allocative Efficiency

In this section, we describe DEA-based tests that can be used to examine the existence of allocative inefficiencies associated with input utilization. In many DEA studies that have examined inefficiency associated with input utilization, inefficiency is often estimated using aggregate cost expenditure information. Banker et al. (2003) address the situation when information about input prices is not available, except for the knowledge that the firms procure the inputs in the same competitive market place. They employ the result that the DEA technical inefficiency measure using a single aggregate cost variable, constructed from multiple inputs weighted by their unit prices, reflects the aggregate technical and allocative inefficiency. This result is then used to develop statistical tests of the null hypothesis of no allocative inefficiency analogous to those of the null hypothesis of no scale inefficiency described earlier.

For the purposes of this section, consider observations on j = 1,….N DMUs, each observation comprising an output vector $\mathbf{y}_j \equiv (y_{1j}, \ldots y_{Rj}) \geq 0$ and a vector of input costs $\mathbf{c}_j \equiv (c_{1j}, \ldots c_{Ij}) \geq 0$ for i=1,…I inputs. Each input i, i=1,…I, is bought by all firms in the same competitive market at a price p_i. Let $\mathbf{p} = (p_1, \ldots, p_I)$ be the vector of input prices. The cost of input i for DMU j is then $c_{ij} = p_i x_{ij}$. The total cost of inputs for DMU j is $c_j = \sum_i p_i x_{ij} = \sum_i c_{ij}$. The input quantities x_{ij} and the price vector $\mathbf{p}$ are not observable by the researcher[6]. Only the output and cost information are observed.

The aggregate technical and allocative inefficiency estimator, $\hat{\theta}_j^Z \geq 1$, is estimated using the following linear program that utilizes output and aggregate cost data:

$$\hat{\theta}_j^Z = \text{argmax}\left\{\theta \,\middle|\, \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_{rk} \geq y_{rj}, \forall r = 1,\ldots R; \sum_{k=1}^{N} \lambda_k c_k \leq c_j / \theta; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,\ldots N \end{array} \right\} \quad (11.3)$$

The technical inefficiency estimator $\hat{\theta}_j^B \geq 1$ is estimated as

$$\hat{\theta}_j^B = \text{argmax}\left\{\theta \,\middle|\, \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_{rk} \geq y_{rj}, \forall r = 1,\ldots R; \sum_{k=1}^{N} \lambda_k c_{ik} \leq c_{ij} / \theta \;\forall i = 1,\ldots I; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,\ldots N \end{array} \right\} \quad (11.4)$$

$\hat{\theta}_j^B$ is a consistent estimator of the true technical inefficiency θ_j^B (Banker 1993). Further the estimator for the allocative inefficiency, $\hat{\theta}_j^V$, can be calculated as $\hat{\theta}_j^Z / \hat{\theta}_j^B$. Under the null hypothesis that the sample data does not exhibit any allocative inefficiency, $\hat{\theta}_j^Z$ is also a consistent estimator of the true technical inefficiency θ_j^B. This leads to the following tests of the null hypothesis of no allocative inefficiency in the utilization of the inputs as opposed to the alternative of existence of allocative inefficiency:

(i) If $\ln(\theta_j^B)$ is distributed as exponential over $[0, \infty)$, then under the null hypothesis of no allocative inefficiency, the test statistic is calculated as $\sum_{j=1}^{N} \ln(\hat{\theta}_j^Z) \Big/ \sum_{j=1}^{N} \ln(\hat{\theta}_j^B)$ and evaluated relative to the critical value of the half-F distribution with (2N, 2N) degrees of freedom.

(ii) If $\ln(\theta_j^B)$ is distributed as half-normal over the range $[0, \infty)$, then under the null hypothesis of no allocative inefficiency, the test statistic is

[6] Sections 5 and 6 of chapter 1 discuss efficiency estimation for situations in which unit prices and unit costs are available.

calculated as $\sum_{j=1}^{N}(\ln(\hat{\theta}_j^Z))^2 \Big/ \sum_{j=1}^{N}(\ln(\hat{\theta}_j^B))^2$ and evaluated relative to the critical value of the half-F distribution with (N, N) degrees of freedom.

(iii) If no such assumptions are maintained about the probability distribution of inefficiency, a non-parametric Kolmogorov-Smirnov's test statistic given by the maximum vertical distance between $F(\ln(\hat{\theta}_j^Z))$ and $F(\ln(\hat{\theta}_j^B))$, the empirical distributions of $\ln(\hat{\theta}_j^Z)$ and $\ln(\hat{\theta}_j^B)$, respectively, is used. This statistic, by construction, takes values between 0 and 1 and a high value is indicative of the existence of allocative inefficiency.

There could also be situations involving multiple outputs and multiple inputs where output quantity information may not be available but monetary value of the individual outputs along with input quantity information may be available. In such situations, output based allocative inefficiency can be estimated and tested using procedures similar to those outlined above for input-based allocative efficiency (Banker et al. 1999).

2.5 Tests of Input Separability

In this section we describe DEA-based tests that can be used to evaluate the null hypothesis of input separability, i.e., the influence of each of the inputs on the output is independent of other inputs, against the alternative hypothesis that the inputs are substitutable. The DEA-based tests proposed in this section evaluate the null hypothesis of input separability over the entire sample data in contrast to the parametric (Berndt and Wood 1975) tests which are operationalized only at the sample mean.

Once again, consider observations on j = 1,….N DMUs, each observation comprising an output vector $\mathbf{y}_j \equiv (y_{1j}, \ldots y_{Rj}) \geq 0$, a vector of inputs $\mathbf{x}_j \equiv (x_{1j}, \ldots x_{Ij}) \geq 0$ and a production technology characterized by a monotone increasing and convex production possibility set $T \equiv \{(\mathbf{y}, \mathbf{x}) \mid \mathbf{y}$ can be produced from $\mathbf{x}\}$. The input-oriented inefficiency measure for this technology is estimated using the following BCC -- Banker, Charnes and Cooper (1984) -- linear program:

$$\hat{\theta}_j^{SUB} = \text{argmax}\left\{\theta \left| \begin{array}{l} \sum_{k=1}^{N}\lambda_k y_{rk} \geq y_{rj}, \forall r = 1,\ldots R; \sum_{k=1}^{N}\lambda_k x_{ik} \leq x_{ij}/\theta \;\forall i = 1,\ldots I; \\ \sum_{k=1}^{N}\lambda_k = 1, \lambda_k \geq 0, \forall k = 1,\ldots N \end{array}\right.\right\} \quad (11.5)$$

When the inputs are separable, input inefficiency is first estimated considering only one input at a time, resulting in I different inefficiency measures corresponding to the I inputs. The overall DMU inefficiency is then estimated as the minimum of these I inefficiency measures. Specifically, the inefficiency corresponding to input i is measured as:

$$\hat{\theta}_j^i = \text{argmax} \left\{ \theta_i \left| \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_{rk} \geq y_{rj}, \forall r = 1,...R; \sum_{k=1}^{N} \lambda_k x_{ik} \leq x_{ij} / \theta_i ; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,...N \end{array} \right. \right\} \quad (11.6)$$

The inefficiency measure under the input separability assumption is then estimated as $\hat{\theta}_j^{SEP} = \text{Min} \left\{ \hat{\theta}_j^i \mid \text{i} = 1,.....,\text{I} \right\}$. Since $\hat{\theta}_j^{SEP}$ is estimated from a less constrained program, $\hat{\theta}_j^{SEP} \geq \hat{\theta}_j^{SUB}$. Under the null hypothesis of input separability, the asymptotic empirical distributions of $\hat{\theta}_j^{SEP}$ and $\hat{\theta}_j^{SUB}$ are identical, with each retrieving the distribution of the true input inefficiency θ.

The above discussion leads to the following tests of the null hypothesis of separability in the utilization of the inputs as opposed to the alternative of substitutability of inputs:

(i) If $\ln(\theta_j)$ is distributed as exponential over $[0, \infty)$, then under the null hypothesis of separability of inputs, the test statistic is calculated as $\sum_{j=1}^{N} \ln(\hat{\theta}_j^{SEP}) \Big/ \sum_{j=1}^{N} \ln(\hat{\theta}_j^{SUB})$ and evaluated relative to the critical value of the half-F distribution with (2N, 2N) degrees of freedom.

(ii) If $\ln(\theta_j)$ is distributed as half-normal over the range $[0, \infty)$, then under the null hypothesis of input separability, the test statistic is calculated as $\sum_{j=1}^{N} (\ln(\hat{\theta}_j^{SEP}))^2 \Big/ \sum_{j=1}^{N} (\ln(\hat{\theta}_j^{SUB}))^2$ and evaluated relative to the critical value of the half-F distribution with (N, N) degrees of freedom.

(iii) If no such assumptions are maintained about the probability distribution of inefficiency, a non-parametric Kolomogorov-Smirnov's test statistic given by the maximum vertical distance between $F(\ln(\hat{\theta}_j^{SEP}))$ and $F(\ln(\hat{\theta}_j^{SUB}))$, the empirical distributions of $\ln(\hat{\theta}_j^{SEP})$ and $\ln(\hat{\theta}_j^{SUB})$, respectively, is used. This statistic, by construction, takes values between 0 and 1 and a high value is indicative of the rejection of input separability in the production technology.

3. HYPOTHESIS TESTS FOR SITUATIONS CHARACTERIZED BY SHIFTS IN FRONTIER

In the previous section, we presented DEA tests that are useful in situations where cross-sectional data on DMUs is used for efficiency analysis. In this section, we describe estimation procedures and statistical tests when both longitudinal and cross-sectional data on DMUs are available and where the object of interest is change in productivity over time[7]. Productivity researchers have advocated both parametric and nonparametric approaches to estimate and analyze the impact of technical change and efficiency change on productivity change. The nonparametric literature (e.g. Färe et al 1997, Ray and Desli 1997, Førsund and Kittelsen 1998) has focused exclusively on the measurement of productivity change using Data Envelopment Analysis (DEA) without attempting to provide a statistical basis to justify those methods. Recently, Banker, Chang and Natarajan (2002) have developed DEA-based estimation methods and tests of productivity change and technical change. This section summarizes salient aspects of the Banker et al. (2002) study. For ease of exposition, we focus on a single output rather than a multiple output vector to illustrate the techniques and tests in this section.

Let $\mathbf{x_t} = (x_{1t},...x_{it},\ldots x_{It}) \in X$, $x_{it} > 0$, t=0,1 be the I-dimensional input vector in period t. Consider two possible values of the time subscript t corresponding to a base period (t = 0) and another period (t = 1). The production correspondence in time t between the frontier output y_t^* and the input $\mathbf{x_t}$, is represented as

$$y_t^* = \phi^t(\mathbf{x_t}),\ t = 0,1 \tag{11.7}$$

The function $\phi^t(.)$: $X \rightarrow \Re^+$ is constrained to be monotone increasing and concave in the input $\mathbf{x_t}$ but not required to follow any specific parametric functional form. The input vector $\mathbf{x_t}$, the production function $\phi^t(\mathbf{x_t})$, and a relative efficiency random variable, α_t, that takes values between $-\infty$ and 0, together determine the realized output, y_t, in period t[8]. Specifically,

$$y_t \equiv e^{\alpha_t}\,\phi^t(\mathbf{x_t}) \tag{11.8}$$

For the above setup, estimators of technical change, relative efficiency change and productivity change, respectively, $\hat{b}_j^{(N)}$, $\hat{r}_j^{(N)}$ and $\hat{g}_j^{(N)}$, for the j^{th} DMU can be estimated from observed output values and *estimators* of frontier outputs using the following expressions:

$$\hat{b}_j^{(N)} = \left\{\ln\left(\hat{\phi}^1(\mathbf{x}_{j1})\right) - \ln\left(\hat{\phi}^0(\mathbf{x}_{j1})\right)\right\}$$

$$\hat{r}_j^{(N)} = \left\{\ln\left(y_{j1}/\hat{\phi}^1(\mathbf{x}_{j1})\right) - \ln\left(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})\right)\right\}$$

$$\hat{g}_j^{(N)} = \left\{\ln\left(y_{j1}/\hat{\phi}^0(\mathbf{x}_{j1})\right) - \ln\left(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})\right)\right\} \tag{11.9}$$

[7] The treatment of productivity change in this section provides an alternative to treatments using the Malmquist index described in chapter 5 and additionally has the advantage of providing explicit statistical characterizations.

[8] The relative efficiency random variable α_t and the inefficiency measure θ_t are linked by the relationship $\alpha_t = -\ln(\theta_t)$.

These estimators satisfy the fundamental relationship that productivity change is the sum of technical change and relative efficiency change i.e., $\hat{g}_j^{(N)} = \hat{b}_j^{(N)} + \hat{r}_j^{(N)}$. The frontier outputs required for the estimation of the various change measures in (11.9) are estimated using linear programs. The estimation of $\hat{\phi}^0(\mathbf{x}_{j0})$ and $\hat{\phi}^1(\mathbf{x}_{j1})$, is done using only the input-output observations from the base period or period 1, as the case may be. The linear program for estimating $\hat{\phi}^0(\mathbf{x}_{j0})$ is the following Banker, Charnes and Cooper (BCC 1984) model:

$$\hat{\phi}^0(\mathbf{x}_{j0}) = \text{argmax}\left\{\hat{\phi}\,\middle|\, \begin{array}{l} \sum_{k=1}^{N}\lambda_{k0}^0 y_{k0} \geq \hat{\phi};\ \sum_{k=1}^{N}\lambda_{k0}^0 x_{ik0} \leq x_{ij0}, \forall i = 1,...I; \\ \sum_{k=1}^{N}\lambda_{k0}^0 = 1, \lambda_{k0}^0 \geq 0, \forall k = 1,...N \end{array}\right\} \quad (11.10)$$

A similar program is used to estimate $\hat{\phi}^1(\mathbf{x}_{j1})$ from input and output data from period 1. $\hat{\phi}^0(\mathbf{x}_{j0})$ and $\hat{\phi}^1(\mathbf{x}_{j1})$ are consistent estimators of $\phi^0(\mathbf{x}_{j0})$ and $\phi^1(\mathbf{x}_{j1})$, respectively, (Banker 1993).

The base period frontier value, $\phi^0(\mathbf{x}_{j1})$, corresponding to *period 1* input, $\mathbf{x}_{j1}$, is estimated based on the following linear program:

$$\hat{\phi}^0(\mathbf{x}_{j1}) = \text{argmax}\left\{\hat{\phi}\,\middle|\, \begin{array}{l} \sum_{k=1}^{N}\lambda_{k0}^0 y_{k0} \geq \hat{\phi};\ \sum_{k=1}^{N}\lambda_{k0}^0 x_{ik0} \leq x_{ij1}, \forall i = 1,...I; \\ \sum_{k=1}^{N}\lambda_{k0}^0 = 1, \lambda_{k0}^0 \geq 0, \forall k = 1,...N \end{array}\right\} \quad (11.11)$$

or $=$ y_{j1} when the above linear program is not feasible

Note that the difference between the above model and the traditional BCC model is that the observation under evaluation is not included in the reference set for the constraints in (11.11) as in the super-efficiency model described first in Banker, Das and Datar (1989) and Anderson and Petersen (1993). It is the case that $\hat{\phi}^0(\mathbf{x}_{j1})$ is a consistent estimator of $\phi^0(\mathbf{x}_{j1})$.

Given the N values of technical, relative efficiency and productivity changes estimated using (11.9), the estimators of the medians of these performance measures are

$$\hat{b}^{MD(N)} = \text{argmin}\ \frac{1}{N}\sum_{j=1}^{N}|\,\hat{b}_j^{(N)} - b\,|$$

$$\hat{r}^{MD(N)} = \text{argmin}\ \frac{1}{N}\sum_{j=1}^{N}|\,\hat{r}_j^{(N)} - r\,|$$

$$\hat{g}^{MD(N)} = \text{argmin}\ \frac{1}{N}\sum_{j=1}^{N}|\,\hat{g}_j^{(N)} - g\,| \quad (11.12)$$

Banker et al. (2002) show that $\hat{b}^{MD(N)}$, $\hat{r}_j^{MD(N)}$ and $\hat{g}_j^{MD(N)}$ are consistent estimators of the population median technical change β^{MD}, population median relative efficiency change ρ^{MD} and population median productivity change γ^{MD}, respectively.

Consider the number of observations, $\hat{p}_\beta^{(N)}$, $\hat{p}_\rho^{(N)}$ and $\hat{p}_\gamma^{(N)}$ (out of the sample of N observations) for which $\hat{b}_j^{(N)}$, $\hat{r}_j^{(N)}$ and $\hat{g}_j^{(N)}$, respectively, is strictly positive. Tests for the median of the various performance measures being equal to zero are conducted as follows:

(a) Under the null hypothesis of zero median technical change between the base period and period t i.e., $\beta^{MD} = 0$, the statistic $\hat{p}_\beta^{(N)}$ is asymptotically distributed as a binomial variate with parameters N and 0.5 i.e., $\hat{p}_\beta^{(N)} \sim b(N,0.5)$.

(b) Under the null hypothesis of zero median relative efficiency change between the base period and period t i.e., $\rho^{MD} = 0$, the statistic $\hat{p}_\rho^{(N)}$ is asymptotically distributed as a binomial variate with parameters N and 0.5 i.e., $\hat{p}_\rho^{(N)} \sim b(N,0.5)$.

(c) Under the null hypothesis of zero median technical change between the base period and period t i.e., $\gamma^{MD} = 0$, the statistic $\hat{p}_\gamma^{(N)}$ is asymptotically distributed as a binomial variate with parameters N and 0.5 i.e., $\hat{p}_\gamma^{(N)} \sim b(N,0.5)$.

Banker et al. (2002) also provide methods for estimating and testing the location of the population mean of the various performance measures. The mean technical change $\hat{\bar{b}}^{(N)}$, mean relative efficiency change $\hat{\bar{r}}^{(N)}$ and mean productivity change $\hat{\bar{g}}^{(N)}$ are estimated as

$$\hat{\bar{b}}^{(N)} = \frac{1}{N}\sum_{j=1}^{N}(\ln(\hat{\phi}^1(\mathbf{x}_{j1})) - \ln(\hat{\phi}^0(\mathbf{x}_{j1})))$$

$$\hat{\bar{r}}^{(N)} = \frac{1}{N}\left\{\ln\left(y_{j1}/\hat{\phi}^1(\mathbf{x}_{j1})\right) - \ln\left(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})\right)\right\}$$

$$\hat{\bar{g}}^{(N)} = \frac{1}{N}\left\{\ln\left(y_{j1}/\hat{\phi}^0(\mathbf{x}_{j1})\right) - \ln\left(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})\right)\right\} \quad (11.13)$$

$\hat{\bar{b}}^{(N)}$, $\hat{\bar{r}}^{(N)}$ and $\hat{\bar{g}}^{(N)}$ are consistent estimators of the population mean technical change, $\bar{\beta}$, population mean relative efficiency change, $\bar{\rho}$, and population productivity change, $\bar{\gamma}$, respectively.

Next consider the statistics $\hat{t}^{(N)}(\beta) = \sqrt{N}\,\hat{b}^{(N)} \big/ \hat{s}(\beta)$, $\hat{t}^{(N)}(\rho) = \sqrt{N}\,\hat{r}^{(N)} \big/ \hat{s}(\rho)$ and $\hat{t}^{(N)}(\gamma) = \sqrt{N}\,\hat{g}^{(N)} \big/ \hat{s}(\gamma)$ where

$$\hat{s}^2(\beta) = \left(\sum_{j=1}^{N} \left(\ln(\hat{\phi}^1(\mathbf{x}_{j1})) - \ln(\hat{\phi}^0(\mathbf{x}_{j1})) \right)^2 - N\left(\hat{b}^{(N)}\right)^2 \right) \Big/ (N-1)$$

$$\hat{s}^2(\rho) = \left(\sum_{j=1}^{N} \left(\ln(y_{j1}/\hat{\phi}^1(\mathbf{x}_{j1})) - \ln(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})) \right)^2 - N\left(\hat{r}^{(N)}\right)^2 \right) \Big/ (N-1)$$

$$\hat{s}^2(\gamma) = \left(\sum_{j=1}^{N} \left(\ln\left(y_{j1}/\hat{\phi}^0(\mathbf{x}_{j1})\right) - \ln\left(y_{j0}/\hat{\phi}^0(\mathbf{x}_{j0})\right) \right)^2 - N\left(\hat{g}^{(N)}\right)^2 \right) \Big/ (N-1) \quad (11.14)$$

For large samples, the distribution of each of $\hat{t}^{(N)}(\beta)$, $\hat{t}^{(N)}(\rho)$ and $\hat{t}^{(N)}(\gamma)$ approaches that of a Student's T variate with N-1 degrees of freedom. Therefore, a simple t-test for the mean of the DMU-specific estimators for the various performance measures estimated using (11.9) is appropriate when the sample size is large.

4. HYPOTHESIS TESTS FOR COMPOSED ERROR SITUATIONS

The tests described in the previous sections are conditioned on a data generating process that characterizes the deviation of the actual output from the production frontier as arising only from a stochastic inefficiency term. In this section, we describe the application of DEA-based tests for composed error situations where the data generating process involves not only the one-sided inefficiency term but also a noise term that is independent of the inefficiency. Recently, Gsatch (1998) and Banker and Natarajan (2001) have developed DEA-based estimation procedures for environments characterized by both inefficiency and noise. The tests developed by Banker (1993) can be adapted to these environments through an appropriate transformation of the inefficiency term. We describe these tests below:

4.1 Tests for Efficiency Comparison

Consider observations on j = 1,….N DMUs, each observation comprising a single output $y_j \geq 0$ and a vector of inputs $\mathbf{x}_j \equiv (x_{1j}, \ldots x_{Ij}) \geq 0$. The production correspondence between the frontier output y^0 and the I inputs is represented as $y^0 = g(\mathbf{x})$ subject to the assumption that g(.) is monotonically increasing and concave in **x**. The deviation from the frontier for the jth DMU could be positive or negative and is represented as $\varepsilon_j = u_j - v_j$

$= g(\mathbf{x}_j) - y_j$. Thus, the deviation is modeled as the sum of two components, a one-sided inefficiency term, u_j, and a two-sided random noise term v_j bounded above at V^M, analogous to composed error formulations in parametric stochastic frontier models (Aigner et al. 1977, Meussen and van den Broeck 1977, Banker and Natarajan 2001). In this stochastic framework, Banker and Natarajan (2003) propose two statistical tests to compare the efficiency of two groups of DMUs.

As before, consider two sub-samples, G_1 and G_2, that are part of the sample of N DMUs when the sample size, N, is large. Let the true inefficiency, u_j, be distributed with means $\bar{u}_1$ and $\bar{u}_2$ in the two groups. Further assume that the variance of the inefficiency is the same in both groups. Define $\tilde{u}_j = V^M - v_j + u_j$. We can estimate $\hat{\tilde{u}}_j$, a consistent estimator of $\tilde{u}_j$, by applying DEA on input and output data from the full sample of N DMUs. For N_1 and N_2 DMUs in subgroups G_1 and G_2, respectively, the null hypothesis of no difference in mean inefficiency between the two subgroups can be tested using the following procedures:

(i) Consider the OLS regression $\hat{\tilde{u}}_j = a_0 + a_1 z_j + e_j$ estimated using a total of N_1+N_2 DEA inefficiency scores. z_j is a dummy variable that takes a value of 0 if a particular DMU belongs to group G_1 and 1 if it belongs to G_2 and e_j is an i.i.d error term. The regression coefficient $\widehat{a_1}$ is a consistent estimator of $\bar{u}_2 - \bar{u}_1$, the difference in mean inefficiency between groups G_2 and G_1. The t-statistic associated with this regression coefficient can be used to evaluate whether two groups are significantly different in terms of mean inefficiency.

(ii) Assume that the probability distributions of the inefficiency random variable u_j and the noise random variable v_j are such that $\tilde{u}_j = V^M - v_j + u_j$ is distributed as a log-normal variable in the two groups. Under the null hypothesis that the mean inefficiencies are equal i.e., $\bar{u}_1 = \bar{u}_2$ and assuming that the variance of u_j is the same in the two groups, the Student-t statistic $\hat{t}$ $= \left(\overline{\widehat{\widetilde{lu}}}_1 - \overline{\widehat{\widetilde{lu}}}_2\right) \Big/ \hat{S}\sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$ distributed with (N_1+N_2-2) degrees of freedom can be used to evaluate the null hypothesis of no difference in mean inefficiency across the two groups. $\overline{\widehat{\widetilde{lu}}}_1$ is $\frac{1}{N_1}\sum_{j=1}^{N_1} \ln(\hat{\tilde{u}}_{j1})$, $\overline{\widehat{\widetilde{lu}}}_2 = \frac{1}{N_2}\sum_{j=1}^{N_2} \ln(\hat{\tilde{u}}_{j2})$ and $\hat{S} = \left(\frac{1}{N_1 + N_2 - 2}\left\{\sum_{j=1}^{N_1}\left\{\ln(\hat{\tilde{u}}_{j1}) - \overline{\widehat{\widetilde{lu}}}_1\right\}^2 + \sum_{j=1}^{N_2}\left\{\ln(\hat{\tilde{u}}_{j2}) - \overline{\widehat{\widetilde{lu}}}_2\right\}^2\right\}\right)^{0.5}$.

4.2 Tests for Evaluating the Impact of Contextual Variables on Efficiency

Analysis of factors contributing to efficiency differences has been an important area of research in DEA. Ray (1991), for instance, regresses DEA scores on a variety of socio-economic factors to identify key performance drivers in school districts. The two-stage approach of first calculating productivity scores and then seeking to correlate these scores with various explanatory variables has been in use for over twenty years but explanations of productivity differences using DEA are still dominated by ad hoc speculations (Førsund 1999). Banker and Natarajan (2001) provide a general framework for the evaluation of contextual variables affecting productivity by considering a variety of Data Generating Processes (DGPs) and present appropriate estimation methods and statistical tests under each DGP. In this section, we describe the DEA-based tests developed in Banker and Natarajan (2001) that can be used to determine the impact of contextual or environmental variables on efficiency.

Consider observations on j = 1,….N decision making units (DMUs), each observation comprising a single output $y_j \geq 0$, a vector of inputs $\mathbf{x}_j \equiv (x_{1j}, \ldots x_{Ij}) \geq 0$, and a vector of contextual variables $\mathbf{z}_j \equiv (z_{1j}, \ldots z_{Sj})$ that may influence the overall efficiency in transforming the inputs into the outputs. The production function g(.) is monotone increasing and concave in **x**, and relates the inputs and contextual variables to the output as specified by the equation

$$y_j = g(\mathbf{x}_j) + v_j - h(\mathbf{z}_j) - u_j \qquad (11.15)$$

where v_j is a two-sided random noise term bounded above at V^M, $h(\mathbf{z}_j)$ is a non-negative monotone increasing function, convex in **z** and u_j is a one-sided inefficiency term. The inputs, contextual variables, noise and inefficiency are all independently distributed of each other. Defining $\tilde{g}(x_j) = g(\mathbf{x}_j) + V^M$ and $\tilde{\delta}_j = (V^M - v_j) + h(\mathbf{z}_j) + u_j \geq 0$, (11.15) can be expressed as

$$y_j = \tilde{g}(x_j) - \tilde{\delta}_j \qquad (11.16)$$

Since $\tilde{g}(.)$ is derived from g(.) by multiplication with a positive constant, $\tilde{g}(.)$ is also monotone increasing and concave. Therefore, the DEA inefficiency estimator, $\hat{\tilde{\delta}}_j$, obtained by performing DEA on the inputs-output observations (y_j, $\mathbf{x}_j$), j = 1, . . .N, is a consistent estimator of $\tilde{\delta}_j$ (Banker 1993). This consistency result is used by Banker and Natarajan (2001) to develop the following DEA-based tests corresponding to different specifications of h(.), the function linking contextual variables to inefficiency.

Consider the case where h(**z**) = h(**z**;β) and h(**z**;β) is a non-negative function, monotone increasing in **z**, linear in β. In this case, the impact of contextual variables can be consistently estimated by regressing the first stage DEA estimate $\hat{\tilde{\delta}}_j$ on the various contextual variables associated with the various components of the β vector. This procedure yields consistent estimators of the parameter vector β. In the special case where h(**z**;β) = **z'**β = $\sum_{j=1}^{N} z_j \beta_j$, the independent variables in the regression are the same as the S contextual variables.

Now, suppose no additional structure can be placed on h($\mathbf{z}_j$) except that it is a non-negative monotone increasing function, convex in **z**. Let $\tilde{\varepsilon}_j = (V^M - v_j) + u_j$. Then $\tilde{\delta}_j$ = h($\mathbf{z}_j$) + $\tilde{\varepsilon}_j$ and a second stage DEA estimation on the pseudo "input-outputs" observations $(\hat{\tilde{\delta}}_j, \mathbf{z}_j)$ yields a consistent estimator $\hat{\tilde{\varepsilon}}_j$ for $\tilde{\varepsilon}_j$ (Banker 1993). This estimator is obtained by solving the following linear programming formulation analogous to the BCC model (Banker, Charnes and Cooper 1984) in DEA individually for each observation in the sample:

$$\hat{\psi}_j = \operatorname{argmin}\left\{\psi \left| \begin{array}{l} \sum_{k=1}^{N} \lambda_k \hat{\tilde{\delta}}_k = \psi; \sum_{k=1}^{N} \lambda_k z_{ik} \geq z_{ij}, \forall i = 1,\ldots S; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,\ldots N \end{array} \right.\right\} \quad (11.17)$$

A consistent estimator for $\tilde{\varepsilon}_j$ for each observation in the sample is obtained as $\hat{\tilde{\varepsilon}}_j = \hat{\tilde{\delta}}_j - \hat{\psi}_j$.

To evaluate the statistical significance of individual z_s, a third stage DEA estimation is first performed on the pseudo observations ($\hat{\tilde{\delta}}_j, \mathbf{z}_j^{-S}$) where $\mathbf{z}^{-s}$ is the original **z** vector without the z_s variable. The following modified version of the program in (11.17) is used for this purpose.

$$\hat{\psi}_j^{-s} = \operatorname{argmin}\left\{\psi \left| \begin{array}{l} \sum_{k=1}^{N} \lambda_k \hat{\tilde{\delta}}_k = \psi; \sum_{k=1}^{N} \lambda_k z_{ik} \geq z_{ij}, \forall i = 1,.,s-1,s+1,..S; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \geq 0, \forall k = 1,\ldots N \end{array} \right.\right\} \quad (11.18)$$

Let the resulting estimator of $\tilde{\varepsilon}_j$ be $\hat{\tilde{\varepsilon}}_j^{-s} = \hat{\tilde{\delta}}_j - \hat{\psi}_j^{-s}$. Since (11.18) is a less constrained program than (11.17), $\hat{\tilde{\varepsilon}}_j^{-s} \geq \hat{\tilde{\varepsilon}}_j$ for all observations j = 1,…N.

Under the null hypothesis that the marginal impact of z_s (i.e. $\partial h(\mathbf{z})/\partial z_s$ if h(.) is differentiable) is zero, the asymptotic distributions of $\tilde{\varepsilon}$ and $\tilde{\varepsilon}^{-s}$ are identical (Banker 1993). If the asymptotic distribution of $\tilde{\varepsilon}_j$ is assumed to be exponential or half-normal, the null hypothesis of no impact of z_s is tested by comparing the ratios $\sum_{j=1}^{N} \hat{\tilde{\varepsilon}}_j^{-s} \Big/ \sum_{j=1}^{N} \hat{\tilde{\varepsilon}}_j$ or $\sum_{j=1}^{N} \left[\hat{\tilde{\varepsilon}}_j^{-s} \right]^2 \Big/ \sum_{j=1}^{N} \left[\hat{\tilde{\varepsilon}}_j \right]^2$ against critical values obtained from half-F distributions with (2N,2N) or (N,N) degrees of freedom, respectively. If $\tilde{\varepsilon}_j$ is exponentially distributed, the test statistic is evaluated relative to the half-F distribution $|F_{2N,2N}|$ with 2N,2N degrees of freedom over the range $[1,\infty)$, since by construction the test statistic is never less than 1. If $\tilde{\varepsilon}_j$ is a half-Normal variate then the test statistic is evaluated relative to the half-F distribution $|F_{N,N}|$ with N,N degrees of freedom over the range $[1,\infty)$. Recall that $\tilde{\varepsilon} = u + (V^M - v)$. Therefore, statistics based on Banker (1993) may not be valid unless the variance of the noise term v is considerably smaller than the variance of the inefficiency term, u, and u is distributed with a mode at its lower support.

The Kolmogorov-Smirnov statistic, which is based on the maximum vertical distance between the empirical cumulative distribution of $\hat{\tilde{\varepsilon}}_j^{-s}$ and $\hat{\tilde{\varepsilon}}_j$, can also be used to check whether the empirical distributions of $\hat{\tilde{\varepsilon}}_j^{-s}$ and $\hat{\tilde{\varepsilon}}_j$ are significantly different. If they are significantly different, then it can be established that z_s has a significant impact on productivity.

4.3 Tests for Evaluating the Adequacy of Parametric Functional Forms

While DEA provides a theoretically correct way to estimate monotone and concave (or convex) functional relationships, it is often useful to represent the relationship in a more parsimonious functional form that is afforded by a parametric specification. Specific parametric functional forms, such as the Cobb-Douglas, are useful if they provide a good approximation to the general monotone and concave (or convex) function as evidenced by sample data. In this section, we present methods developed in Banker, Janakiraman and Natarajan (2002) to evaluate the adequacy of a parametric functional form to represent the functional relationship between an endogenous variable and a set of exogenous variables given the minimal maintained assumption of monotonicity and concavity.

Consider sample data on an endogenous variable and I exogenous variables for N observations. For the j^{th} observation, denote the endogenous variable as y_j and the vector of exogenous variables as $X_j \equiv (x_{1j}, x_{2j}, \ldots x_{Ij})$. The relationship between y_j and X_j is specified as:

$$y_j = g(X_j) \;\; e^{\varepsilon_j} \tag{11.19}$$

where $g(.)$ is a monotone increasing and concave function. It is assumed further that ε_j is independent of X_j and i.i.d. with a probability density function $f(\varepsilon)$ over the range $[-S_L,\; S_U] \subseteq \Re$, where $S_L \geq 0$, and $S_U \geq 0$ are unknown parameters that describe the lower and upper supports of the distribution. Define $\tilde{g}(X) = g(X)e^{S_U}$ and $\tilde{\varepsilon}_j = S_U - \varepsilon_j \geq 0$ such that (11.19) can be rewritten as $y_j = \tilde{g}(X_j) \;\; e^{-\tilde{\varepsilon}_j}$. The DEA estimator of $\tilde{g}(X_j)$ can be estimated using the following linear program:

$$\hat{\tilde{g}}^{DEA}(X_j) = \operatorname{argmax}\left\{ y \,\middle|\, \begin{array}{l} \sum_{k=1}^{N} \lambda_k y_k = y; \sum_{k=1}^{N} \lambda_k x_{ik} \leq x_{ij}, \forall i = 1, \ldots I; \\ \sum_{k=1}^{N} \lambda_k = 1, \lambda_k \;\geq\; 0, \forall k = 1, \ldots N \end{array} \right\} \tag{11.20}$$

An estimator for $\tilde{\varepsilon}_j$ for each observation in the sample can then be obtained as $\hat{\tilde{\varepsilon}}_j^{DEA} = \ln(\hat{\tilde{g}}^{DEA}(X_j)) - \ln(y_j)$. Further, $\hat{\tilde{g}}^{DEA}(X_j)$ and $\hat{\tilde{\varepsilon}}_j^{DEA}$ are consistent estimators of $\tilde{g}(X_j)$ and $\tilde{\varepsilon}_j$ respectively (Banker 1993).

The parametric estimation is carried out by specifying a parametric form $g(X;\beta)$ and regressing ln(y) on the exogenous variables in ln($g(X;\beta)$). The residuals from the regression are used to obtain $\hat{S}_U = \max\{\ln(y) - \ln(\hat{g}(X_j;\hat{\beta}))\}$ which is a consistent estimator of S_U (Greene 1980). The estimated deviation from the parametric frontier is then calculated as $\hat{\tilde{\varepsilon}}_j^{PARAM} = \ln(\hat{g}(X_j;\hat{\beta})) + \hat{S}_U - \ln(y_j)$. In addition to $\hat{\tilde{\varepsilon}}^{DEA}$, $\hat{\tilde{\varepsilon}}_j^{PARAM}$ also is a consistent estimator of $\tilde{\varepsilon}_j$ under the null hypothesis that $g(X;\beta) = g(X)$ for all *X*. Banker et al. (2002) prove that the asymptotic distribution of $\hat{\tilde{\varepsilon}}^{PARAM}$ retrieves the true distribution of $\tilde{\varepsilon}$ if the parametric specification is, in fact, the true specification of the production function (i.e., $g(X;\beta) = g(X)$ for all X). Further, they also show that if the parametric specification is, in fact, the true specification of the production function then as $N \to \infty$, (a) the asymptotic distribution of $\hat{\tilde{\varepsilon}}^{PARAM}$ converges to that of $\hat{\tilde{\varepsilon}}^{DEA}$ and (b) both $\hat{\tilde{\varepsilon}}_j^{PARAM}$ and $\hat{\tilde{\varepsilon}}_j^{DEA}$ converge asymptotically to $\tilde{\varepsilon}_j$ for all $j \in J$, where J is a given set of observations. Based on these results, they suggest the following four tests for testing the adequacy of the parametric functional form:

(a) The first test uses the Kolmogorov-Smirnov test statistic given by the maximum vertical distance between $\hat{F}(\hat{\hat{\varepsilon}}_j^{DEA})$ and $\hat{F}(\hat{\hat{\varepsilon}}_j^{PARAM})$, where $\hat{F}(\hat{\hat{\varepsilon}}_j^{DEA})$ and $\hat{F}(\hat{\hat{\varepsilon}}_j^{PARAM})$ denote the empirical distributions of $\hat{\hat{\varepsilon}}_j^{DEA}$ and $\hat{\hat{\varepsilon}}_j^{PARAM}$ respectively. A low value is indicative of support for the null hypothesis that $g(X;\beta)$ adequately represents $g(X)$.

(b) The second procedure is based on the regression of rank of $\hat{\hat{\varepsilon}}_j^{DEA}$ on the rank of $\hat{\hat{\varepsilon}}_j^{PARAM}$ (Iman and Conover 1979). Under the null hypothesis that the parametric form is adequate, the expected value of the coefficient on $\hat{\hat{\varepsilon}}_j^{PARAM}$ in the rank regression is asymptotically equal to 1. The null hypothesis is evaluated against the alternative hypothesis that the regression coefficient has a value less than 1.

(c) The third test procedure employs the Wilcoxon rank-sum test to evaluate whether the empirical distributions $\hat{F}(\hat{\hat{\varepsilon}}_j^{DEA})$ and $\hat{F}(\hat{\hat{\varepsilon}}_j^{PARAM})$ are different. If the test shows these distributions to be different, the adequacy of the parametric form is rejected.

(d) The fourth procedure is based on Theil's (1950) distribution-free test. This test evaluates the null hypothesis that $\mu_1 = 1$ against the alternative $\mu_1 \neq 1$ in the relation $\hat{\hat{\varepsilon}}^{DEA} = \mu_0 + \mu_1 \hat{\hat{\varepsilon}}^{PARAM}$. To compute Theil's statistic, the difference $D_j = \hat{\hat{\varepsilon}}_j^{DEA} - \hat{\hat{\varepsilon}}_j^{PARAM}$ is calculated and then the data are sorted by $\hat{\hat{\varepsilon}}_j^{PARAM}$. Next, a score c_{ji} = 1, 0 or -1 is assigned for each i< j depending on whether $D_j - D_i > 0$, $= 0$ or < 0 respectively. Theil's test statistic C is defined as $C = \sum_{j=1}^{N} c_{ji}$. Theil's test statistic is distributed as a standard normal variate for large samples and a high absolute value of C rejects the adequacy of the parametric functional form.

Banker et al. (2002) also propose an alternative approach to evaluating the adequacy of a parametric functional form. The approach suggested by them relies on Wooldridge's (1992) Davidson-Mackinnon type test to evaluate a linear null model against a nonparametric alternative. Banker et al. (2002) propose the use of a sieve DEA estimator as the nonparametric alternative for the purposes of Wooldridge's test since the test cannot be applied directly when the nonparametric alternative is based on the traditional DEA estimator. Interested readers are referred to section 2.3 of Banker et al. (2002) for additional details on how Wooldridge's test can be applied to examine the adequacy of a specified parametric form for a monotone and concave function.

5. CONCLUDING REMARKS

We have described here several statistical tests that can be used to test hypotheses of interest and relevance to applied users of Data Envelopment Analysis. A common underlying theme of these tests is that the deviation from the DEA frontier can be viewed as a stochastic variable. While the DEA estimator is biased in finite samples, the expected value of the DEA estimator is almost certainly the true parameter value in large samples. The tests described in this paper rely on this asymptotic property of the DEA estimator.

An important caveat is that the tests described in this paper are designed for large samples. Results of simulation studies conducted on many of the tests proposed in this study suggest that these tests perform very well for sample sizes similar to those used in many typical applications of DEA[9]. These tests need to be used with caution in small samples. We believe additional simulation studies are warranted to provide evidence on small sample performance of the tests described here. Clearly, this is an important area for future research.

We believe that there are many more avenues and areas where DEA-based statistical tests can be applied. This is because the flexible structure of DEA facilitates application in a large number of situations where insufficient information or guidance may preclude the use of parametric methods. Statistical tests developed during the past 10 years have contributed significantly to the reliability of managerial and policy implications of DEA studies and we believe that they will continue to enrich future applications of DEA.

REFERENCES

1. Aigner, D.J., Lovell, C.A.K. and Schmidt, P., 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, 6, 21-37.
2. Anderson, P., and Petersen, N.C., 1993. A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science* 39, 1261-1264.
3. Banker, R.D., 1984. Estimating Most Productive Scale Size Using Data Envelopment Analysis. *European Journal of Operations Research,* July, 35-44.

[9] Banker (1996) provides details on some of these simulation studies. Based on the results of these studies, it appears that the tests described in this paper perform well in sample sizes of the order of 50 or more.

4. Banker, R.D., 1993. Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation. *Management Science,* October, 1265-1273.
5. Banker, R.D., 1996. Hypothesis Tests Using Data Envelopment Analysis. *Journal of Productivity Analysis,* 7, 139-159.
6. Banker, R.D., Charnes, A. and Cooper, W.W., 1984. Models for the Estimation of Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*. 30, 1078-1092.
7. Banker, R.D., Chang, H. and Natarajan, R., 1999. Statistical Tests of Allocative Efficiency Using DEA: An Application to the U.S. Public Accounting Industry. Working Paper, The University of Texas at Dallas.
8. Banker, R.D., Chang, H. and Natarajan, R., 2002. Nonparametric Analysis of Productivity Change and Technical Progress. Working Paper, The University of Texas at Dallas.
9. Banker, R.D., Das, S. and Datar S., 1989. Analysis of Cost Variances for Management Control in Hospitals. *Research in Governmental and Nonprofit Accounting,* 5, 269-291.
10. Banker, R.D., Janakiraman, S. and Natarajan, R., 2002. Evaluating the Adequacy of Parametric Functional Forms in Estimating Monotone and Concave Production Functions. *Journal of Productivity Analysis.* Volume 17, 111-132.
11. Banker, R.D., Janakiraman, S. and Natarajan, R., 2003. Analysis of Trends in Technical and Allocative Efficiency: An Application to Texas Public School Districts. Forthcoming *European Journal of Operations Research*.
12. Banker, R. D., and Natarajan, R., 2001. Evaluating Contextual Variables Affecting Productivity using Data Envelopment Analysis. Working paper, The University of Texas at Dallas.
13. Banker, R.D., and Natarajan, R., 2003. DEA-based Hypothesis Tests for Comparison of Two Groups of Decision Making Units. Working paper, The University of Texas at Dallas.
14. Berndt, E., and Wood, D., 1975. Technology, Prices and the Derived Demand for Energy. *Review of Economics and Statistics*. 259-268.
15. Charnes, A., Cooper, W.W., and Rhodes, E., 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research,* 429-444.
16. Färe, R., Griffel-tatje, E., Grosskopf, S., Lovell, C.A.K., 1997. Biased Technical Change and the Malmquist Productivity Index. *Scandinavian Journal of Economics*, Volume 99: 119-127.
17. Førsund, F.R., 1999. The Evolution of DEA – The Economics Perspective. *Working Paper*, University of Oslo, Norway.

18. Førsund, F., and Kittelsen, S., 1998. Productivity Development of Norwegian Electricity Distribution Utilities. *Resource and Energy Economics*. 20:207-224.
19. Greene, W.H., 1980. Maximum Likelihood Estimation of Econometric Frontier Production Functions. *Journal of Econometrics* 13, 27-56.
20. Gstach, D., 1998. Another Approach to Data Envelopment Analysis in Noisy Environments: DEA+. *Journal of Productivity Analysis,* 9, 161-176.
21. Iman, R. L. and Conover, W. J., 1979. The Use of Rank Transform in Regression. *Technometrics.* 499-509.
22. Meeusen, W. and van den Broeck, J., 1977. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review.* June, 435-444.
23. Ray, S., 1991. Resource-use Efficiency in Public Schools: A Study of Connecticut Data. *Management Science.* 1620-1628.
24. Ray, S., and Desli, E., 1997. Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries: Comment. *American Economic Review* 87: 1033-1039.
25. Theil, H., 1950. A Rank-invariant Method of Linear and Polynomial Regression Analysis, *I. Proc. Kon. Ned. Akad. V. Wetensch.* A 53, 386-392.
26. Wooldridge, J. M., 1992. A Test for Functional Form Against Nonparametric Alternatives. *Econometric Theory* 8, 452-475.