Journal of Productivity Analysis, 17, 111–132, 2002

Evaluating the Adequacy of Parametric Functional Forms in Estimating Monotone and Concave Production Functions

RAJIV D. BANKER
School of Management, University of Texas at Dallas, Richardson, TX 75083-0688, USA

SURYA JANAKIRAMAN
School of Management, University of Texas at Dallas, Richardson, TX 75083-0688, USA

RAM NATARAJAN*
School of Management, University of Texas at Dallas, Richardson, TX 75083-0688, USA

Abstract

We consider situations where the a priori guidance provided by theoretical considerations indicates only that the function linking the endogenous and exogenous variables is monotone and concave (or convex). We present methods to evaluate the adequacy of a parametric functional form to represent the relationship given the minimal maintained assumption of monotonicity and concavity (or convexity). We evaluate the adequacy of an assumed parametric form by comparing the deviations of the fitted parametric form from the observed data with the corresponding deviations estimated under DEA. We illustrate the application of our proposed methods using data collected from school districts in Texas. Specifically, we examine whether the Cobb–Douglas and translog specifications commonly employed in studies of education production are appropriate characterizations. Our tests reject the hypotheses that either the Cobb–Douglas or the translog specification is an adequate approximation to the general monotone and concave production function for the Texas school districts.

Keywords: data envelopment analysis, production function estimation, evaluating parametric functional forms, Texas school districts

1. Introduction

A primary objective of much empirical research is the investigation of the relationship between an endogenous (or dependent) variable and a set of exogenous (or independent) variables. Researchers typically examine the relationship by making two major assumptions.

* Helpful comments and suggestions from participants at the Sixth European Workshop on Efficiency and Productivity Analysis in Copenhagen, Denmark in October 1999 are gratefully acknowledged.

The first assumption specifies the functional form that expresses the endogenous variable as a function of the exogenous variables. The second assumption specifies a probability distribution for the residual, the random variable that captures the difference between the actual and the predicted values of the endogenous variable. Based on these two assumptions, procedures such as maximum likelihood estimation are used to statistically estimate the relationship. These major assumptions about the functional form and the probability distribution are usually treated as maintained assumptions.

Frequently, the a priori guidance provided to a researcher by theoretical considerations is limited. For instance, in productivity analysis, theoretical models may specify that the function linking the endogenous variable to exogenous variables is monotone increasing and concave (or convex) but may not indicate a particular parametric form for the function. Similarly, while theory may suggest that the distribution of the residual is bounded, it may not provide a specific parametric form. It is common for researchers to carry out the estimation of production functions assuming a specific parametric form for the function.

While Data Envelopment Analysis (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. This is similar in spirit to the observation by Farrell (1957, pp. 275–276) that:

> (I)t is a laborious business both to specify and to analyze the efficient production function as it stands. This does not constitute a criticism of the method—if the world is complex, its analysis must needs be laborious. However, it would be a great convenience to approximate the efficient production function by a simple mathematical function. It would then be specified by a simple formula, together with the values of a few parameters, and its properties could be analysed algebraically.
>
> The best known and perhaps the most plausible of such approximations is the Cobb–Douglas function (Cobb and Douglas, 1928).... If it were to turn out to be a good approximation to the observed function, it would have the great advantage that it could be specified in terms of four or five parameters. Its disadvantage is that it makes quite strong assumptions about the shape of the efficient function, and for this reason it would not have been desirable to fit such a function at the beginning of the analysis.

In contrast, we present in this paper a DEA-based method to evaluate at the beginning of the analysis whether a specific parametric functional form, such as the Cobb–Douglas, is a good approximation to the general monotone and concave (or convex) function as evidenced by sample data.

Our objective is not the evaluation of whether DEA or parametric estimation method is superior. Rather, theory can usually justify no more than a minimal structure on the production correspondence such as that imposed in DEA, whereas parametric methods provide a more parsimonious representation while imposing additional structure. Our objective in this paper is to leverage the strengths of the two approaches by emphasizing parsimony in both structure and representation. For this purpose we develop procedures to evaluate whether a parsimonious parametric representation is adequate.[1]

Traditional DEA models estimate an "extremal" function with random deviation on only one side of the function (Banker, 1993). Our analysis in this paper accommodates the estimation of an extremal function with one sided errors as well as the estimation of an "average" function with random deviations on either side of the function or production frontiers with composed errors comprising a one-sided efficiency term and a two-sided noise term (Meeusen and van den Broeck, 1977; Aigner, Lovell and Schmidt, 1977). For this purpose we employ a transformation based on Gstach (1998) that links the average and the extremal function.

The remainder of this paper is structured as follows. In Section 2, we present the theoretical framework and details of our proposed method. In Section 3, we illustrate its application using data collected for school districts in the state of Texas. We conclude the paper in Section 4 with our closing remarks.

2. Theoretical Framework and Proposed Method

2.1. Theoretical Framework

Originally designed to evaluate the relative efficiency of decision making units (DMUs) in the not-for-profit and government sectors, such as schools and hospitals, that use multiple inputs to produce multiple outputs, DEA has since been extended to a variety of different applications with new and innovative mathematical formulations and used for many other types of entities in all sectors of the economy (Banker, 1996; Cooper, Thompson and Thrall, 1996). To benchmark targets against which a DMU's relative efficiency may be measured, DEA models have typically sought to estimate production frontiers. Thus, while regression models reflect the average relationship between the dependent and independent variables, traditional DEA models reflect the extremal relationship between the dependent and independent variables. While this feature makes the DEA particularly suitable for constructing efficient production frontiers, it used to be argued that the DEA technique was not suitable for capturing the average relationship between the dependent and the independent variables. In this paper, by employing a suitable transformation, we step beyond the traditional model of an extremal relationship and use the DEA technique to model also the average relationship between the independent and the dependent variables.

2.1.1. Data Generating Process We consider sample data on an endogenous variable and I exogenous variables for N observations. For the jth observation, we 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:

$$y_j = g(X_j)e^{\varepsilon_j} \tag{1}$$

where $g(.)$ is a monotone increasing and concave function. We assume 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.[2] Observe that the case of a full frontier function, as in the traditional DEA models, is accommodated by $S_U = 0$. In general, for the case of an average function, as in a regression model, or when the distribution of ε is a convolution of the distributions of an efficiency component and a noise component as in Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977), we have $S_U > 0$.[3]

We also assume that there is positive probability of having some observations with ε realizations that are arbitrarily close to S_U, i.e.,

$$\int_{S_U-\delta}^{S_U} f(\varepsilon)\,d\varepsilon > 0 \quad \forall\delta > 0 \tag{2}$$

2.1.2. DEA Estimation We begin as in Banker and Natarajan (2000) by defining

$$\tilde{g}(X) = g(X)e^{S_U} \tag{3}$$

and

$$\tilde{\varepsilon}_j = S_U - \varepsilon_j \geq 0 \tag{4}$$

Using equations (3) and (4), equation (1) can be rewritten as

$$y_j = \tilde{g}(X_j)e^{-\tilde{\varepsilon}_j} \tag{5}$$

Two observations are of interest. First, $\tilde{g}(.)$ is monotone increasing and concave since the multiplication of $g(.)$ by a constant preserves these properties. Second, it is evident that $\tilde{\varepsilon}$ takes values in the range $[0, S_L + S_U]$ and its probability density is given by the function $f(S_U - \tilde{\varepsilon})$. Further, equation (2) can be transformed to

$$\int_0^{\delta} f(S_U - \tilde{\varepsilon})\,d\tilde{\varepsilon} > 0 \quad \forall\delta > 0 \tag{6}$$

Thus, while $y_j \leq \tilde{g}(X_j)$, there is strictly positive probability that observed y_j is arbitrarily close to the true functional value $\tilde{g}(X_j)$.

An estimator of $\tilde{g}(X_j)$ is obtained by solving the following BCC (Banker, Charnes and Cooper, 1984) version of the DEA model individually for each observation in the sample:

$$\hat{\tilde{g}}^{DEA}(X_j) = \text{Max}\, y = \sum_{k=1}^{N} \lambda_{jk} y_k \tag{7.0}$$

subject to

$$\sum_{k=1}^{N} \lambda_{jk} x_{ik} \leq x_{ij} \quad \forall i = 1, \ldots, I \tag{7.1}$$

$$\sum_{k=1}^{N} \lambda_{jk} = 1 \tag{7.2}$$

$$\lambda_{jk} \geq 0 \quad \forall k = 1, \ldots, N \tag{7.3}$$

An estimator for $\tilde{\varepsilon}_j$ for each observation in the sample is obtained as $\hat{\tilde{\varepsilon}}_j^{DEA} = \ln(\hat{\tilde{g}}^{DEA}(X_j)) - \ln(y_j)$. Since S_U is unknown, the functional relation $g(.)$ and the p.d.f. $f(\varepsilon)$ for the residual ε are retrieved only up to a multiplicative constant. The following three propositions follow immediately from Banker (1993):

PROPOSITION 1 *$\hat{\tilde{g}}^{DEA}(X_j)$ and $\hat{\tilde{\varepsilon}}_j^{DEA}$ are consistent estimators of $\tilde{g}(X_j)$ and $\tilde{\varepsilon}_j$, respectively.*

PROPOSITION 2 *The asymptotic distribution of* $\hat{\tilde{\varepsilon}}^{DEA}$ *retrieves the true distribution of* $\tilde{\varepsilon}$.

PROPOSITION 3 *If, and only if,* $f(\tilde{\varepsilon})$ *is increasing for all* $\tilde{\varepsilon}$, *then both* $\hat{\tilde{g}}^{DEA}(X_j)$ *and* $\hat{\tilde{\varepsilon}}_j^{DEA}$ *are also maximum likelihood estimators.*

2.1.3. Monotone and Convex Production Function The case when a priori considerations indicate that $g(X)$ is a monotone increasing and *convex* function is addressed in a similar manner by defining

$$\tilde{\tilde{g}}(X) = g(X)e^{-S_L} \tag{8}$$

and

$$\tilde{\tilde{\varepsilon}}_j = S_L + \varepsilon_j \tag{9}$$

Using equations (8) and (9), equation (1) can be rewritten as

$$y_j = \tilde{\tilde{g}}(X_j)e^{\tilde{\tilde{\varepsilon}}_j} \tag{10}$$

It follows as before that $\tilde{\tilde{g}}(.)$ is monotone increasing and convex, and the p.d.f. of $\tilde{\tilde{\varepsilon}}$ over the range $[0, S_L + S_U]$ is given by the function $f(\tilde{\tilde{\varepsilon}} - S_L)$. We also assume that there is positive probability of having some observations with ε realizations arbitrarily close to S_L, i.e.,

$$\int_{S_L}^{S_L+\delta} f(\varepsilon)\,d\varepsilon > 0 \quad \forall\delta > 0 \tag{11}$$

An estimator for $\tilde{\tilde{g}}(X_j)$ is obtained by solving the following DEA linear programming formulation individually for each observation in the sample:

$$\hat{\tilde{\tilde{g}}}^{DEA}(X_j) = \text{Max } y = \sum_{k=1}^{N} \lambda_{jk} y_k \tag{12.0}$$

subject to

$$\sum_{k=1}^{N} \lambda_{jk} x_{ik} \geq x_{ij} \quad \forall i = 1, \ldots, I \tag{12.1}$$

$$\sum_{k=1}^{N} \lambda_{jk} = 1 \tag{12.2}$$

$$\lambda_{jk} \geq 0 \quad \forall k = 1, \ldots, N \tag{12.3}$$

An estimator for $\tilde{\tilde{\varepsilon}}_j$ for each observation in the sample is obtained as $\hat{\tilde{\tilde{\varepsilon}}}_j^{DEA} = \ln(y_j) - \ln(\hat{\tilde{\tilde{g}}}^{DEA}(X_j))$. Analogs of the earlier propositions also follow for this case of monotone and convex functions.

2.1.4. Parametric Estimation We begin by specifying a parametric form $g(X;\beta)$ for the production function. Additional constraints are imposed when estimating the parameters β to ensure that $g(X;\beta)$ is monotone increasing and concave (or convex, as the case may be).

Any estimation procedure that yields consistent estimators of β under the maintained assumptions in Section 2.1.1 is satisfactory for the test procedures we describe below. Several alternatives are available. Maximum likelihood estimators (MLE) such as those described for stochastic frontier analysis by Meeusen and Van den Broeck (1977) and Aigner, Lovell and Schmidt (1977) may be obtained using a nonlinear constrained optimization procedure. However, such approaches also require the specification of a parametric form $f(\varepsilon; \alpha)$ for the probability density function over the range $[S_L, S_U] \subseteq \Re$, with additional constraints to ensure that $f(\varepsilon; \hat{\alpha})$ satisfies the condition in equation (2). An alternative approach that does not require such additional parametric assumptions about the density function is to estimate the transformed frontier as in equation (5). Corrected ordinary least squares (COLS) estimators may be obtained by first performing OLS and then shifting the intercept term by the magnitude of the largest negative residual as in Richmond (1974). This approach provides consistent estimators without imposing a parametric form for the density function (Greene, 1980, p. 34), and in some simulation studies its performance has been found to be superior to that of the MLE procedure (Olson, Schmidt and Waldman, 1980).

For expositional purposes, we assume that 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 $\hat{S}_U = \max\{\ln(y) - \ln(\hat{g}(X_j - \hat{\beta}))\}$, 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.

When $g(X)$ is monotone increasing and convex, the estimator of S_L is calculated from the residuals as $\hat{S}_L = -\min\{\ln(y_j) - \ln(\hat{g}(X_j; \hat{\beta}))\}$ and the deviation from the frontier is estimated as $\hat{\tilde{\varepsilon}}_j^{PARAM} = \ln(y_j) - \ln(\hat{g}(X_j, \hat{\beta})) - \hat{S}_L$.

We represent the parametric estimators of the transformed production frontier in equation (3) by $\hat{\tilde{g}}^{PARAM}(X_j; \hat{\beta})$ and of the residual from this estimated frontier by $\hat{\tilde{\varepsilon}}_j^{PARAM} = \ln(\hat{\tilde{g}}^{PARAM}(X_j)) - \ln(y_j)$. The following proposition can be proved along the lines of the proof of proposition 6 in Banker (1993):

PROPOSITION 4 *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).*

Proof. Since $\hat{\tilde{\varepsilon}}_j^{PARAM}$ is a consistent estimator of $\tilde{\varepsilon}_j$ under the null hypothesis that $g(X; \beta) = g(X)$ for all X, we have

$$\lim_{N\to\infty} \Pr\left\{\left|\hat{\tilde{\varepsilon}}_j^{PARAM} - \tilde{\varepsilon}_j\right| > \Delta\right\} = 0 \quad \text{for any} \quad \Delta > 0.$$

For any given constants $E_j \in \Re$,

$$\begin{aligned}\Pr\left\{\hat{\tilde{\varepsilon}}_j^{PARAM} \leq E_j \text{ for all } j \in J\right\} = {} & \Pr\left\{\text{Both } \hat{\tilde{\varepsilon}}_j^{PARAM}, \tilde{\varepsilon}_j \leq E_j \text{ for all } j \in J\right\} \\ & + \Pr\left\{\hat{\tilde{\varepsilon}}_j^{PARAM} \leq E_j \text{ for all } j \in J \text{ and} \right. \\ & \left. \tilde{\varepsilon}_j > E_j \text{ for some } j \in J\right\}.\end{aligned}$$

$$\begin{aligned}\Pr\{\tilde{\varepsilon}_j \le E_j \text{ for all } j \in J\} = \Pr\big\{&\text{Both } \hat{\hat{\varepsilon}}_j^{PARAM}, \tilde{\varepsilon}_j \le E_j \text{ for all } j \in J\big\}\\ &+\Pr\big\{\tilde{\varepsilon}_j \le E_j \text{ for all } j \in J \text{and}\\ &\quad \hat{\hat{\varepsilon}}_j^{PARAM} > E_j \text{ for some } j \in J\big\}.\end{aligned}$$

Since

$$\lim_{N\to\infty} \Pr\big\{\big|\hat{\hat{\varepsilon}}_j^{PARAM} - \tilde{\varepsilon}_j\big| > \triangle\big\} = 0 \quad \text{for any} \quad \Delta > 0,$$

$$\lim_{N\to\infty} \Pr\big\{\hat{\hat{\varepsilon}}_j^{PARAM} \le E_j \text{ for all } j \in J \text{ and } \tilde{\varepsilon}_j > E_j \text{ for some } j \in J\big\} = 0$$

and

$$\lim_{N\to\infty} \Pr\big\{\tilde{\varepsilon}_j \le E_j \text{ for all } j \in J \text{ and } \hat{\hat{\varepsilon}}_j^{PARAM} > E_j \text{ for some } j \in J\big\} = 0.$$

Therefore,

$$\lim_{N\to\infty} \Pr\big\{\hat{\hat{\varepsilon}}_j^{PARAM} \le E_j \text{ for all } j \in J\big\} = \Pr\{\tilde{\varepsilon}_j \le E_j \text{ for all } j \in J\}$$

and the asymptotic distribution of $\tilde{\varepsilon}_j^{PARAM}$ retrieves the true distribution of $\tilde{\varepsilon}$. ■

If both DEA and the parametric method are employed to estimate monotone increasing and concave production frontiers $\hat{\hat{g}}^{DEA}(.)$ and $\hat{\hat{g}}^{PARAM}(.)$ that envelop all observed data then we have an additional result from Banker (1993, proposition 1):

PROPOSITION 5 $\hat{\hat{\varepsilon}}_j^{DEA} \le \hat{\hat{\varepsilon}}_j^{PARAM}$ *for all* $j = 1, \ldots, N$.

2.2. Test Procedures

2.2.1. Primary Test Procedure To evaluate whether the parametric specification adequately fits the true production function, we would like to compare the empirical distribution of the deviations obtained from the parametric estimation with the true distribution of the deviations $\tilde{\varepsilon}$, and test whether they are identical. For instance, if $F(\tilde{\varepsilon})$ were known, we could employ the Kolmogorov–Smirnov test based on the maximum vertical distance between $\hat{F}(\hat{\hat{\varepsilon}}^{PARAM})$ and $F(\tilde{\varepsilon})$ to evaluate the fit. However, in most empirical applications, the true distribution $F(\tilde{\varepsilon})$ is not known. In the next proposition we address this problem by proving that:

PROPOSITION 6 *If the parametric specification is, in fact, the true specification of the production function then the maximum vertical distance between the empirical cumulative distributions of* $\hat{\hat{\varepsilon}}^{PARAM}$ *and* $\hat{\hat{\varepsilon}}^{DEA}$ *converges to the maximum vertical distance between the cumulative distributions of* $\tilde{\varepsilon}$ *(the true deviation) and* $\hat{\hat{\varepsilon}}^{PARAM}$ *when the sample size increases to infinity.*

Proof. It follows from Banker (1993) that

$$\tilde{\varepsilon}_j \geq \hat{\hat{\varepsilon}}_j^{DEA} \quad \text{and} \quad \hat{\hat{\varepsilon}}_j^{PARAM} \geq \hat{\hat{\varepsilon}}_j^{DEA}$$

This implies for a sample of size N that

$$\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{DEA}\right) \geq F(\tilde{\varepsilon}) \quad \text{and} \quad \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{DEA}\right) \geq \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{PARAM}\right) \quad \text{for all} \quad j.$$

We denote the maximum vertical distance (over all $j = 1, \ldots, N$) between $F(\tilde{\varepsilon}_j)$ and $\hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$ as $D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}}$ and that between $\hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$ and $\hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{DEA})$ as $D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}}$. Specifically, we write

$$D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}} = \max_j \left\{ \left| F(\tilde{\varepsilon}_j) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{PARAM}\right) \right| \right\}$$

and

$$D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}} = \max_j \left\{ \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{DEA}\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_j^{PARAM}\right) \right\}$$

While it is the case that the cumulative distribution of $\hat{\hat{\varepsilon}}^{DEA}$ always lies above those of $\tilde{\varepsilon}$ and $\hat{\hat{\varepsilon}}^{PARAM}$, the cumulative distribution of $\tilde{\varepsilon}$ and $\hat{\hat{\varepsilon}}^{PARAM}$ may cross.

Let the maximum (absolute) vertical distance between $F(\tilde{\varepsilon}_j)$ and $\lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$ occur at a point $j = A$ and that between $\lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{DEA})$ and $\lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$ occur at a point $j = B$. There are two possible cases. In the first case, the maximum (absolute) vertical distance between $F(\tilde{\varepsilon}_j)$ and $\lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$ occurs at a point where $F(\tilde{\varepsilon}_j) \geq \lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM})$. In the other case, the maximum occurs at a point where $\lim_{N\to\infty} \hat{F}^{(N)}(\hat{\hat{\varepsilon}}_j^{PARAM}) \geq F(\tilde{\varepsilon}_j)$.

Case 1

$$\begin{aligned}
\lim_{N\to\infty} D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}} &= \lim_{N\to\infty} \left\{ F(\tilde{\varepsilon}_A) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_A^{PARAM}\right) \right\} \\
&\geq \lim_{N\to\infty} \left\{ F(\tilde{\varepsilon}_B) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_B^{PARAM}\right) \right\} \\
&= \lim_{N\to\infty} \left\{ \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_B^{DEA}\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_B^{PARAM}\right) \right\} \\
&= \lim_{N\to\infty} D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}}
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty} D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}} &= \lim_{N\to\infty} \left\{ \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_B^{DEA}\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_B^{PARAM}\right) \right\} \\
&\geq \lim_{N\to\infty} \left\{ \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_A^{DEA}\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_A^{PARAM}\right) \right\} \\
&= \lim_{N\to\infty} \left\{ F(\tilde{\varepsilon}_A) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}_A^{PARAM}\right) \right\} \\
&= \lim_{N\to\infty} D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}}
\end{aligned}$$

Case 2

$$\begin{aligned}
\lim_{N\to\infty} D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}} &= \lim_{N\to\infty} \left\{\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_A\right) - F(\tilde{\varepsilon}_A)\right\} \\
&\geq \lim_{N\to\infty} \left\{F(\tilde{\varepsilon}_B) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_B\right)\right\} \\
&= \lim_{N\to\infty} \left\{\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{DEA}_B\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_B\right)\right\} \\
&= \lim_{N\to\infty} D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}}
\end{aligned}$$

$$\begin{aligned}
\lim_{N\to\infty} D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}} &= \lim_{N\to\infty} \left\{\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{DEA}_B\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_B\right)\right\} \\
&\geq \lim_{N\to\infty} \left\{\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_A\right) - \hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{DEA}_A\right)\right\} \\
&= \lim_{N\to\infty} \left\{\hat{F}^{(N)}\left(\hat{\hat{\varepsilon}}^{PARAM}_A\right) - F(\tilde{\varepsilon}_A)\right\} \\
&= \lim_{N\to\infty} D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}}
\end{aligned}$$

Therefore, in both cases,

$$\lim_{N\to\infty} D^N_{\hat{\hat{\varepsilon}}^{DEA},\hat{\hat{\varepsilon}}^{PARAM}} = \lim_{N\to\infty} D^N_{\tilde{\varepsilon},\hat{\hat{\varepsilon}}^{PARAM}} \qquad \blacksquare$$

Thus, the Kolmogorov–Smirnov statistic to evaluate whether the distribution of the deviations $\hat{\hat{\varepsilon}}^{PARAM}$ using the parametric method is identical to the true (but unknown) distribution of $\tilde{\varepsilon}$ is approximated for large samples by a corresponding Kolmogorov–Smirnov test statistic to evaluate whether the empirical distributions of $\hat{\hat{\varepsilon}}^{PARAM}$ and $\hat{\hat{\varepsilon}}^{DEA}$ are identical. Therefore, we construct our primary test procedure based on the Kolmogorov–Smirnov test statistic given by the maximum vertical distance between $\hat{F}(\hat{\hat{\varepsilon}}^{DEA}_j)$ and $\hat{F}(\hat{\hat{\varepsilon}}^{PARAM}_j)$, where $\hat{F}(\hat{\hat{\varepsilon}}^{DEA}_j)$ and $\hat{F}(\hat{\hat{\varepsilon}}^{PARAM}_j)$ denote the empirical distributions of $\hat{\hat{\varepsilon}}^{DEA}_j$ and $\hat{\hat{\varepsilon}}^{PARAM}_j$, respectively. By construction, this statistic takes values between 0 and 1 and a low value is indicative of support for the null hypothesis that $g(X;\beta)$ adequately represents $g(X)$.

2.2.2. Additional Test Procedures The results in propositions 4 and 5 above immediately imply the following:

PROPOSITION 7 *If the parametric specification is, in fact, the true specification of the production function then as $N \to \infty$, (a) the asymptotic distribution of $\hat{\hat{\varepsilon}}^{PARAM}$ converges to that of $\hat{\hat{\varepsilon}}^{DEA}$ and (b) both $\hat{\hat{\varepsilon}}^{PARAM}_j$ and $\hat{\hat{\varepsilon}}^{DEA}_j$ converge asymptotically to $\tilde{\varepsilon}_j$ for all $j \in J$, where J is a given set of observations.*

Proof. (a) $|\hat{F}(\hat{\hat{\varepsilon}}^{PARAM}) - \hat{F}(\hat{\hat{\varepsilon}}^{DEA})| \leq |\hat{F}(\hat{\hat{\varepsilon}}^{PARAM}) - F(\tilde{\varepsilon})| + |\hat{F}(\hat{\hat{\varepsilon}}^{DEA}) - F(\tilde{\varepsilon})|$

The asymptotic distribution of $\hat{\hat{\varepsilon}}^{DEA}$ retrieves the true distribution of $\tilde{\varepsilon}$ (Banker, 1993, Proposition 6) and under the premise of this proposition, the asymptotic distribution of $\hat{\hat{\varepsilon}}^{PARAM}$ retrieves the true distribution of $\tilde{\varepsilon}$ (Proposition 4 above). Therefore, each of the two terms on the right-hand side above converges to zero as the sample size increases to infinity, which in turn implies that the left-hand side converges to zero as well.

(b) This result follows immediately from the consistency property of the two estimators. ■

Thus, under the null hypothesis that the parametric form is an adequate representation of the true production function, the empirical distributions of the DEA and the parametric estimates are asymptotically identical. For comparison with our primary Kolmogorov–Smirnov test, we present three additional test procedures to evaluate the hypothesis that the functional form specified as $g(X;\beta)$ is adequate in representing the monotone increasing and concave function $g(X)$. These test procedures are based on a comparison of the empirical distributions of the DEA estimator $\hat{\tilde{\varepsilon}}_j^{DEA}$ and the parametric methods estimator $\hat{\tilde{\varepsilon}}_j^{PARAM}$. Our test procedures directly evaluate whether the empirical distributions of the two sets of estimates themselves are identical. Our test statistics are not based on some unobserved true parameters characterizing the distribution of $\tilde{\varepsilon}$ that may have been approximated by their estimated values.

The DEA based procedures we present to test the adequacy of an assumed parametric form $g(X;\beta)$ to represent the functional relationship $y = g(X)$ exploit the asymptotic relationship between the empirical distributions of the DEA and the parameter estimates if the null hypothesis is true. That is, these procedures rely on the property that the statistically consistent DEA and parametric estimators $\hat{\tilde{\varepsilon}}^{DEA}$ and $\hat{\tilde{\varepsilon}}^{PARAM}$ retrieve the true distribution of the deviation $\tilde{\varepsilon}$ from the production frontier $\tilde{g}(x)$ when the null hypothesis of $g(X;\beta) = g(X)$ for all X holds true. We describe our nonparametric tests of the degree of fit for the case when $g(X)$ is monotone increasing and concave.

In the first test procedure, the rank of $\hat{\tilde{\varepsilon}}_j^{DEA}$ is regressed on the rank of $\hat{\tilde{\varepsilon}}_j^{PARAM}$ (Iman and Conover, 1979). Under the null hypothesis that the parametric form is adequate, the expected value of the coefficient on $\hat{\tilde{\varepsilon}}_j^{PARAM}$ 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. The second test procedure employs the Wilcoxon rank-sum test to evaluate whether the empirical distributions $\hat{F}(\hat{\tilde{\varepsilon}}_j^{DEA})$ and $\hat{F}(\hat{\tilde{\varepsilon}}_j^{PARAM})$ are different. The third test procedure that we use is based on Theil's (1950) distribution-free test to evaluate whether the regression coefficient obtained by regressing $\hat{\tilde{\varepsilon}}_j^{DEA}$ on $\hat{\tilde{\varepsilon}}_j^{PARAM}$ is equal to 1.

2.3. An Alternative Approach

Our alternative approach relies on Wooldridge's (1992) Davidson–Mackinnon type test to evaluate a linear null model against a nonparametric alternative. Wooldridge provides relevant asymptotic theory to show that the test is consistent when the alternative nonparametric model is based on a sieve estimator. Wooldridge's test cannot be applied directly when the nonparametric alternative is based on the traditional DEA estimator. However, the sieve DEA estimator described by Banker and Maindiratta (1992) (BM hereafter) falls in the category of alternative nonparametric estimation procedures to which Wooldridge's test can be applied directly. We first describe the BM sieve estimation procedure as it applies to the specification in equation (1) and then we describe the estimation of the Wooldridge (1992) test statistic to evaluate the adequacy of the parametric specifications.

While BM's (1992) formulation does not impose any parametric functional form on $g(X)$ it does require the specification of a parametric form $f(\varepsilon; \alpha)$ for the probability density function of ε over the range $[S_L, S_U] \subseteq \Re$. Some additional notation is needed before presenting the BM formulation. We represent the sieve by the set of vectors $X_k, k = 1, \ldots, K$. We also represent the fitted value for the jth observation, $j = 1, \ldots, N$, obtained from the BM estimation procedure by $y_j^f \equiv \hat{g}^{BM}(X_j)$ and the vector of fitted values by $Y^f \equiv (y_1^f, \ldots, y_n^f)$. The BM maximum likelihood estimation procedure for the joint determination of the vector of fitted values Y^f and the parameter vector $\boldsymbol{\alpha}$ is:

$$\underset{Y^f,\alpha}{\text{Maximize}} \ln L \equiv \sum_{j=1}^{N} \ln f\left(\ln y_j - \ln y_j^f; \alpha\right) \tag{13.0}$$

subject to

$$(i)\ y_j^f \geq \phi_j(Y^f) \text{ for each } j = 1, \ldots, N. \tag{13.1}$$

$$(ii)\ \phi_j(Y^f) \equiv \max\left[\sum_{k=1}^{K} \lambda_{jk} y_k^f \,\middle|\, \sum_{k=1}^{K} \lambda_{jk} X_k \leq X_j; \sum_{k=1}^{K} \lambda_{jk} = 1; \lambda_{jk} \geq 0 \,\forall k\right] \tag{13.2}$$

$$(iii)\ Y^f \geq 0. \tag{13.3}$$

Sarath and Maindiratta (1997) prove that the BM procedure yields consistent estimators of the fitted outputs and the error density function parameters if this imposed structure is valid.

The following three-step process describes the application of the Wooldridge (1992) test:

(1) For the assumed parametric form $g^{PARAM}(X; \beta)$, estimate the parametric fitted value $\hat{g}^{PARAM}(X_j)$ and $\varepsilon_j^{PARAM} = (\ln y_j - \ln \hat{g}^{PARAM}(X_j; \beta))$ by maximizing $\ln L \equiv \sum_{j=1}^{n} \ln f(\ln y_j - \ln \hat{g}^{PARAM}(X_j; \beta))$ subject to constraints to ensure that $\hat{g}^{PARAM}(.)$ is monotone increasing and concave.

(2) Estimate $y_j^f \equiv \hat{g}^{BM}(X_j)$ by using the BM procedure outlined in equation (13).

(3) Let $h(X_j, \ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}^{PARAM}(X_j)); \beta, \gamma)$ be the linear function that has all the terms in $\ln(g^{PARAM}(X_j))$ as well as an extra term based on the difference between the BM fitted value and the parametric fitted value, and let (β, γ) be the parameter vector characterizing this function h(.). Estimate the parameters (β, γ) by maximizing $\ln L \equiv \sum_{j=1}^{N} \ln f(\varepsilon_j^{PARAM} - h(.; \beta, \gamma))$. If the t-statistic of the coefficient γ on $\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}^{PARAM}(X_j))$ is significant at conventional levels, then the null hypothesis of the adequacy of the parametric form is rejected against the nonparametric BM alternative.

We next describe the probability density function $f(\varepsilon; \alpha)$ used in the empirical analysis for the ML estimation. As before the maintained assumption is that $f(\varepsilon)$ is defined over the range $[-S_L, S_U] \subseteq \Re$, where $S_L \geq 0$, and $S_U \geq 0$. We further assume that ε consists of two components: a two-sided error term v defined in the range $(-\infty, S_U]$ and a one-sided error term u defined in the range $[0, \infty)$. We also assume that u is distributed as Gamma $(2, \lambda)$

and v as $N(0, \sigma_v^2)$ truncated above at S_U. Let $\varepsilon_1 = (\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\lambda})$, $\varepsilon_2 = (\frac{S_U}{\sigma_v} + \frac{\sigma_v}{\lambda})$, and $f^*(.)$ and $F^*(.)$ be the standard normal density and distribution functions, respectively. Then, the probability density function of $\varepsilon = v - u$ defined in the range $(-\infty, S_U]$ is given by:

$$f(\varepsilon) = \frac{\sigma_v e^{\sigma_v^2/2\lambda^2 + \varepsilon/\lambda}}{\lambda^2 F^*\left(\frac{S_U}{\sigma_v}\right)}[\{f^*(\varepsilon_1) - f^*(\varepsilon_2)\} + \varepsilon_1\{F^*(\varepsilon_1) - F^*(\varepsilon_2)\}] \tag{14}$$

It can be verified that $f(\varepsilon = S_U) = 0$. This additional assumption about a parametric form for the probability distribution enables the ML estimation of the parametrically specified function $g(X; \beta)$ in the first step, of the nonparametric sieve DEA function of Banker and Maindiratta (1992) in the second step, and of the augmented parametric function $h(.)$ in the third step to construct Wooldridge's test.

3. Application to Texas School Districts

Considerable prior research has sought to evaluate the efficiency of school districts. While several studies have employed DEA to estimate the underlying production function (Ahn et al., 1988; Bessent et al., 1982; Ray, 1991), many others have employed a parametric functional specification. For instance, Chizmar and Zak (1984) and Butler and Monk (1985) use a Cobb–Douglas specification while Jimenez (1986), Callan and Santerre (1990), Gyimah-Brempong and Gyapong (1992) and Smet and Nonneman (1998) use a translog specification to model the relationship between the inputs and the outputs in educational institutions. Sengupta and Sfeir (1986) use both DEA and translog specifications in their analysis. A question that has not been addressed in prior research is whether either the Cobb–Douglas or the translog functional form is an adequate representation of the education production function.

We collected data on school districts in Texas for three fiscal years: 1994–1995, 1995–1996 and 1996–1997. Consistent with many earlier studies on the education production function (Butler and Monk, 1985; Jimenez, 1986; Callan and Santerre, 1990) we measure output (*students*) as the total number of students enrolled. As in Grosskopf et al. (1999) we restrict our attention to school districts with at least 50 students in both the 6th and 10th grades to avoid sampling problems that might be introduced by a small number of students. We use full-time instructional staff (*teachers*), support staff (*support*) and administrative staff (*admin*) as the three inputs, consistent with Callan and Santerre (1990) and Ray (1991).

Our final sample consists of 570, 586, and 583 school districts, respectively, for the three years in our sample. Two very large districts, Dallas and Houston with 154,847 and 209,375 enrolled students in academic year 1996–1997 are excluded from the analysis because of their uniqueness.[4] Table 1 presents descriptive statistics on the output and input measures. The mean (median) number of *students* is 5,542 (2,101), 5,512 (2,062) and 5,663 (2,129) respectively for the academic years 1994–1995, 1995–1996 and 1996–1997. There is considerable variation in the size of the school districts as measured by student enrollment. The median school district has a student-teacher ratio of 13 to 1 and a student-staff ratio of 16 to 1. The median support staff to administrator ratio is 11 to 1.

Our objective is to use the data for these three years to evaluate the adequacy of fitting a particular parametric form to the data. Prior studies have used either the loglinear or the translog form to characterize the association between student enrollment and the labor

Table 1. Descriptive statistics of inputs and outputs of Texas school districts.

	Mean	Std. Deviation	25th Percentile	Median	75th Percentile
Panel A (Year 1994–1995)					
Students	5542	9676	1192	2101	4894
Teachers	350.5	590.9	85.0	137.7	309.5
Support	303.6	521.5	64.3	118.3	258.0
Admin	22.5	34.0	7.0	10.9	20.7
Panel B (Year 1995–1996)					
Students	5512	9734	1170	2062	4767
Teachers	351.8	593.8	85.8	139.8	305.2
Support	305.7	524.9	65.0	117.0	261.4
Admin	22.5	34.3	6.9	11.0	20.8
Panel C (Year 1996–1997)					
Students	5663	9999	1201	2129	4876
Teachers	364.7	612.7	89.4	144.5	322.0
Support	317.8	544.3	67.7	121.7	276.9
Admin	23.3	35.4	6.9	11.2	21.0

The variable *students* denote the total number of students enrolled in the school district for a given fiscal year. *Teachers, support* and *admin* are variables representing full time equivalent teachers, support staff and administrators employed by the school district during the year. There are 570, 583 and 586 school districts in our sample corresponding to the years 1994–1995, 1995–1996 and 1996–1997, respectively.

inputs. In our context, the loglinear form may be specified as:

$$\log(\mathit{students}) = \beta_0 + \beta_1 \log(\mathit{teachers}) + \beta_2 \log(\mathit{support}) + \beta_3 \log(\mathit{admin}) + \varsigma \quad (15)$$

The translog specification for our model is:

$$\begin{aligned}\log(\mathit{students}) = {} & \beta_0 + \beta_1 \log(\mathit{teachers}) + \beta_2 \log(\mathit{support}) + \beta_3 \log(\mathit{admin}) \\ & + \gamma_1 (\log(\mathit{teachers}))^2 + \gamma_2 (\log(\mathit{support}))^2 + \gamma_3 (\log(\mathit{admin}))^2 \\ & + \eta_{12} \log(\mathit{teachers}) \log(\mathit{support}) + \eta_{23} \log(\mathit{support}) \log(\mathit{admin}) \\ & + \eta_{31} \log(\mathit{admin}) \log(\mathit{teachers}) + \zeta \end{aligned} \quad (16)$$

The adequacy of the loglinear specification in capturing the functional relationship between the output and inputs when the true functional form is translog, is evaluated by testing whether the estimated coefficients on the second order terms of equation (16) are all zero.

To evaluate the adequacy of a parametric form in representing the monotone and concave production function, we compare the empirical distribution of the estimated DEA deviations with that of the estimated loglinear or translog deviations, as the case may be. For the inputs (*teachers, support, admin*) and output (*students*) observation for each school district, we first estimate the deviations from the DEA frontier for each of the three years estimated using all the available data for a particular year. The deviation score for each observation is computed as the difference between the logarithm of the frontier value and the logarithm of the observed output value.

The loglinear and the translog parametric forms of the production function are estimated for each year after imposing restrictions (as explained below) on the parameters to ensure that the fitted functions are monotone increasing and concave at each observation.

For the loglinear form of the production function, the estimation is carried out after constraining the coefficients of the inputs to the production process to lie in the range [0,1] and the sum of the three coefficients to be less than 1. Referring to equation (15), this requirement to ensure monotonicity and concavity translates into the constraints $0 \leq \beta_1, \beta_2, \beta_3 \leq 1$ and $\sum_{i=1}^{3} \beta_i \leq 1$.

In the case of translog estimation, the following two sets of condition are imposed on the estimated parameters. The first set of conditions ensures that the output is monotonically increasing in the inputs. Let A_1, A_2 and A_3 denote the partial derivatives of the output (*students*) with respect to each of the three inputs (*teachers, support, admin*) respectively. Then,

$$A_1 = \frac{students}{teachers}[\beta_1 + 2\gamma_1 \log(teachers) + \eta_{12} \log(support) + \eta_{31} \log(admin)] \quad (17)$$

$$A_2 = \frac{students}{support}[\beta_2 + 2\gamma_2 \log(support) + \eta_{12} \log(teachers) + \eta_{23} \log(admin)] \quad (18)$$

$$A_3 = \frac{students}{admin}[\beta_3 + 2\gamma_3 \log(admin) + \eta_{23} \log(support) + \eta_{31} \log(teachers)] \quad (19)$$

Monotonicity requires that A_1, A_2 and A_3 are all non-negative.

The second set of conditions ensures that the production function is concave. Let B_1, B_2 and B_3 represent the second order derivatives of the output (*students*) with respect to each of the three inputs (*teachers, support, admin*) respectively. Then

$$B_1 = \frac{(A_1)^2}{students} - \frac{A_1}{teachers} + \frac{2\gamma_1(students)}{(teachers)^2} \quad (20)$$

$$B_2 = \frac{(A_2)^2}{students} - \frac{A_2}{support} + \frac{2\gamma_2(students)}{(support)^2} \quad (21)$$

$$B_3 = \frac{(A_3)^2}{students} - \frac{A_3}{admin} + \frac{2\gamma_3(students)}{(admin)^2} \quad (22)$$

Further, let C_{ij}, $(i, j) = (1, 2), (2, 3)$ and $(3, 1)$ denote the cross partial derivatives defined as follows:

$$C_{12} = \frac{(A_1)(A_2)}{students} + \frac{\eta_{12}(students)}{(teachers)(support)} \quad (23)$$

$$C_{23} = \frac{(A_2)(A_3)}{students} + \frac{\eta_{23}(students)}{(support)(admin)} \quad (24)$$

$$C_{31} = \frac{(A_3)(A_1)}{students} + \frac{\eta_{31}(students)}{(teachers)(admin)} \quad (25)$$

Next, define H as the determinant of the Hessian matrix where

$$H = (B_1)(B_2)(B_3) + 2(C_{12})(C_{23})(C_{31}) - (B_1)(C_{23})^2 - (B_2)(C_{31})^2 - (B_3)(C_{12})^2 \quad (26)$$

The parameters are estimated subject also to the constraints that B_1, B_2 and B_3 are non-positive and H is non-negative to ensure that the estimated translog function is concave at each observation in the sample.

Table 2 shows the estimated coefficients of the loglinear and the translog specifications of the education production function for the Texas school district data for each of the three years. The estimated parameters for the Cobb–Douglas specification are comparable for all three years. While it appears that *admin* does not have any influence in the production function based on the loglinear parametric form results, the translog parametric form indicates that all three labor inputs are positively related to the number of students taught. The estimation results for both the loglinear and the translog functions are also qualitatively similar to the estimation results (not reported here) when no constraints are imposed on the parameters to ensure monotonicity and concavity.

Likelihood ratio tests indicate for all three years that the loglinear functional form is rejected given a translog production function. The last row of Table 2 provides the value of the likelihood ratio and the p value at which the Cobb–Douglas specification is rejected given the translog function.

Next, we obtain an estimate of $\tilde{\varepsilon}$, the deviation of each observed output value from the predicted value of the output variable based on the parametrically estimated production frontier, and compare it with the estimate of $\tilde{\varepsilon}$ obtained using DEA. As observed in Section 2.1, $\hat{\tilde{\varepsilon}}^{DEA}$ is a consistent estimator of $\tilde{\varepsilon}$. Under the null hypothesis that the parametric function $g(X;\beta) = g(X)$ for all X, $\hat{\tilde{\varepsilon}}_j^{PARAM} = \ln(\hat{g}^{PARAM}(X_j,\hat{\beta})) + \hat{S}_U^{PARAM} - \ln(y_j)$ is also a consistent estimator of $\tilde{\varepsilon}$, where $PARAM = Cobb\text{–}Douglas$ or *translog* as the case may be.

The summary statistics of $\hat{\tilde{\varepsilon}}$, the estimated deviations of the actual value of output from the frontiers fitted using DEA, Cobb–Douglas and translog specifications are presented in Table 3. The deviations $\hat{\tilde{\varepsilon}}_j^{DEA}$ obtained from DEA are generally much lower than the deviations $\hat{\tilde{\varepsilon}}_j^{Cobb-Douglas}$ or $\hat{\tilde{\varepsilon}}_j^{trans\log}$ obtained from fitting either of the two parametric forms considered in this paper. This is as expected because production frontier estimated by DEA always fits more tightly than any other monotone concave production frontier (Banker, 1993, Proposition 1).

As described in Section 2.2.1, our primary test procedure makes use of the Kolmogorov–Smirnov statistic, which is based on the maximum vertical distance between the empirical cumulative distribution of $\hat{\tilde{\varepsilon}}_j^{PARAM}$ and $\hat{\tilde{\varepsilon}}_j^{DEA}$. Panel A of Table 4 presents the Kolmogorov–Smirnov statistic for the comparison of deviations from DEA and Cobb–Douglas parametric frontiers and indicates that the null hypothesis of identical distributions of deviations from DEA and Cobb–Douglas frontiers is rejected. Panel A of Table 5 presents similar statistics for the comparison of DEA deviations and translog deviations, and indicates that the null hypothesis of adequacy of the translog form is rejected.

Next, we present the results of the three additional tests that are motivated in Section 2.2.2. All of these three test procedures compare the deviations from the estimated DEA and the estimated parametric frontiers. The first of these three tests is based on the regression of the rank of the DEA deviations on the rank of deviations from each parametric frontier. Panel B of Table 4 provides the results of the comparison of DEA deviations to Cobb–Douglas deviations. The estimated regression coefficient on the rank of Cobb–Douglas deviation ranges from 0.7970 to 0.9419 for the three sample years. In all three years, the

Table 2. Estimation of the relationship between the number of students in a school district and the full time equivalent teachers, administrators and support staff.

Loglinear specification:

$$\log(students) = \beta_0 + \beta_1 \log(teachers) + \beta_2 \log(support) + \beta_3 \log(admin) + \varsigma$$

Translog specification:

$$\begin{aligned}\log(students) = {} & \beta_0 + \beta_1 \log(teachers) + \beta_2 \log(support) + \beta_3 \log(admin) + \gamma_1 (\log(teachers))^2 \\ & + \gamma_2 (\log(support))^2 + \gamma_3 (\log(admin))^2 + \eta_{12} \log(teachers) \log(support) \\ & + \eta_{23} \log(support) \log(admin) + \eta_{31} \log(admin) \log(teachers) + \zeta\end{aligned}$$

Estimated Parameter	Year 1994–1995		Year 1995–1996		Year 1996–1997	
	Loglinear	Translog	Loglinear	Translog	Loglinear	Translog
β_0	2.7172 (0.0050)	2.2301 (0.2439)	2.6965 (0.0060)	1.5920 (0.3093)	2.6870 (0.0059)	2.2432 (0.2847)
β_1	0.9092 (0.0151)	1.0301 (0.1403)	0.9188 (0.0231)	1.4269 (0.1976)	0.9045 (0.0224)	1.0078 (0.1749)
β_2	0.0908 (0.0151)	−0.0506 (0.0280)	0.0812 (0.0231)	−0.0174 (0.03160)	0.0955 (0.0224)	−0.0675 (0.0327)
β_3	0.0000 (0.0000)	0.3045 (0.1209)	0.0000 (0.0000)	−0.0976 (0.1968)	0.0000 (0.0000)	0.3373 (0.1498)
γ_1	—	−0.0060 (0.0178)	—	−0.0648 (0.0254)	—	−0.0229 (0.0148)
γ_2	—	−0.0177 (0.0058)	—	−0.0063 (0.0104)	—	−0.0386 (0.0305)
γ_3	—	0.0444 (0.0162)	—	−0.0095 (0.0351)	—	0.0410 (0.0120)
η_{12}	—	0.0515 (0.0101)	—	0.0272 (0.0272)	—	0.0935 (0.0436)
η_{23}	—	−0.0125 (0.0164)	—	−0.0249 (0.0105)	—	−0.0127 (0.0311)
η_{31}	—	−0.0799 (0.0400)	—	0.0605 (0.0647)	—	−0.0811 (0.0564)
Log-likelihood	496.55	605.47	464.39	614.91	463.33	588.89
Likelihood ratio		217.84 (0.001)		301.4 (0.001)		251.12 (0.001)

The standard error of the estimated parameter is given in parentheses except in the last row. For the likelihood ratio statistic, the p value at which the Cobb–Douglas specification is rejected given the translog specification is in parentheses.
The estimation is carried out after imposing appropriate constraints to ensure monotonicity and concavity of the estimated production function.
For fiscal years 1994–1995, 1995–1996 and 1996–1997, 570, 583 and 586 observations respectively were used for estimation.
Likelihood ratio is the test statistic for testing whether the coefficients of the second order terms (γ_1, γ_2, γ_3, η_{12}, η_{23} and η_{31}) are all equal to zero.

Table 3. Descriptive statistics of deviations from estimated DEA, Cobb–Douglas and translog frontiers.

	Mean	Std. Deviation	25th Percentile	Median	75th Percentile
Panel A (Year 1994–1995)					
$\hat{\varepsilon}^{DEA}$	0.1662	0.0949	0.1036	0.1638	0.2288
$\hat{\varepsilon}_j^{Cobb-Douglas}$	0.2398	0.1013	0.1682	0.2343	0.3083
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{Cobb-Douglas}$	−0.0735	0.0531	−0.0979	−0.0561	−0.0366
$\hat{\varepsilon}_j^{trans\log}$	0.2251	0.0837	0.1688	0.2251	0.2769
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{trans\log}$	−0.0589	0.0466	−0.0793	−0.0473	−0.0245
Panel B (Year 1995–1996)					
$\hat{\varepsilon}^{DEA}$	0.1752	0.0985	0.1075	0.1717	0.2395
$\hat{\varepsilon}_j^{Cobb-Douglas}$	0.2703	0.1059	0.1976	0.2608	0.3381
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{Cobb-Douglas}$	−0.0951	0.0634	−0.1301	−0.0771	−0.0472
$\hat{\varepsilon}_j^{trans\log}$	0.2581	0.0848	0.2013	0.2571	0.3128
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{trans\log}$	−0.0829	0.0507	−0.0982	−0.0685	−0.0499
Panel C (Year 1996–1997)					
$\hat{\varepsilon}^{DEA}$	0.2330	0.1175	0.1486	0.2259	0.3150
$\hat{\varepsilon}_j^{Cobb-Douglas}$	0.2748	0.1094	0.1981	0.2639	0.3500
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{Cobb-Douglas}$	−0.0418	0.0417	−0.0485	−0.0312	−0.0187
$\hat{\varepsilon}_j^{trans\log}$	0.4136	0.0882	0.3565	0.4149	0.4699
$\hat{\varepsilon}^{DEA} - \hat{\varepsilon}_j^{trans\log}$	−0.1806	0.0793	−0.2278	−0.1687	−0.1209

$\hat{\varepsilon}_j^{DEA}$ is computed as equal to $\ln(\hat{g}^{DEA}(X_j)) - \ln(y_j)$, where y_j is the actual output value and $\hat{g}^{DEA}(X_j)$ is the estimated frontier value corresponding to the jth observation.
$\hat{\varepsilon}_j^{Cobb-Douglas}$ is calculated as equal to $\ln(\hat{g}^{Cobb-Douglas}(X_j;\hat{\beta})) + \hat{S}_U - \ln(y_j)$ where y_j is the actual output value and $\hat{g}^{Cobb-Douglas}(X_j;\hat{\beta})$ is the estimated value of the output using the Cobb–Douglas specification and $\hat{S}_U^{Cobb-Douglas}$ is the estimated value of the upper support of the error distribution in (1).
$\hat{\varepsilon}_j^{trans\log}$ is calculated as equal to $\ln(\hat{g}^{trans\log}(X_j;\hat{\beta})) + \hat{S}_U - \ln(y_j)$ where y_j is the actual output value and $\hat{g}^{trans\log}(X_j;\hat{\beta})$ is the estimated value of the output using the translog specification and $\hat{S}_U^{trans\log}$ is the estimated value of the upper support of the error distribution in (1).

null hypothesis that the regression coefficient equals one is rejected at the 5% significance level. Panel B of Table 5 provides the results of the comparison of DEA deviations to translog deviations. The estimated regression coefficient on the rank of translog deviation varies from 0.7174 to 0.8593. In every year, the null hypothesis of unit slope is rejected at the 5% significance level.

The next test is based on the Wilcoxon rank sum statistic. This involves ordering of all the deviations from DEA as well as a parametric frontier from the smallest to the largest, and summing the ranks of the DEA deviations to obtain the Wilcoxon statistic. Panel C in

Table 4. Tests of adequacy of Cobb–Douglas functional form.

		Year 1994–1995	Year 1995–1996	Year 1996–1997
Panel A: Kolmogorov–Smirnov tests				
Kolmogorov–Smirnov statistic		0.1474	0.1775	0.0729
p value at which the null hypothesis of identical distributions is rejected		0.0001	0.0001	0.0001
Panel B: Rank regression tests				
$Rank(\hat{\varepsilon}^{DEA}) = \alpha_0 + \alpha_1 Rank(\hat{\varepsilon}^{Cobb-Douglas}) + \varphi$				
α_0		44.1922	59.59	16.6966
		(5.98)	(7.04)	(3.61)
α_1		0.8452	0.7970	0.9419
		(37.70)	(31.89)	(67.59)
Adjusted R^2		0.7139	0.6346	0.8870
Test of $\alpha_1 = 1$	t-statistic	6.91	8.12	4.17
	p value	0.00	0.00	0.00
Panel C: Wilcoxon tests				
Wilcoxon statistic		−11.66	−14.3396	−5.9355
p value at which the null hypothesis of identical distributions is rejected		0.0001	0.0001	0.0001
Panel D: Theil tests				
$\hat{\varepsilon}^{DEA} = \mu_0 + \mu_1 \hat{\varepsilon}^{Cobb-Douglas} + \nu$				
Theil's statistic		−9.8106	−9.7072	1.7511
p value at which the null hypothesis $\mu_1 = 1$ is rejected		0.0001	0.0001	0.0791
Panel E: Wooldridge tests				
$\hat{\varepsilon}^{Cobb-Douglas} = \beta_0 + \beta_1 \ln(teachers) + \beta_2 \ln(Support) + \beta_3 \ln(Admin) + \gamma\{\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}(X_j;\beta))\} + \zeta$				
Independent variable				
$\ln(\hat{g}^{BM}(X_j) - \ln(\hat{g}(X_j;\beta))$		0.5469	0.4459	0.3469
		(7.53)	(7.72)	(3.61)

$\hat{\varepsilon}^{Cobb-Douglas}$ is computed as equal to $\ln(y_j) - \ln(\hat{g}^{Cobb-Douglas}(X_j;\hat{\beta}))$; t-statistics are given in parentheses.

each of Tables 4 and 5 displays the results of such comparisons between DEA deviations and Cobb–Douglas deviations, and DEA deviations and translog deviations, respectively. In all three years, for both parametric forms, the null hypothesis that the distribution of the deviations from DEA is the same as the distribution of the deviations from either of the two parametric specifications is rejected at the 5% significance level.

Table 5. Tests of adequacy of translog functional form.

	Year 1994–1995	Year 1995–1996	Year 1996–1997
Panel A: Kolmogorov–Smirnov tests			
Kolmogorov–Smirnov statistic	0.1465	0.1834	0.3165
p value at which the null hypothesis of identical distributions is rejected	0.0001	0.0001	0.0001
Panel B: Rank regression tests			
$Rank(\hat{\hat{\varepsilon}}^{DEA}) = \alpha_0 + \alpha_1 Rank(\hat{\hat{\varepsilon}}^{trans\log}) + \eta$			
α_0	40.1829 (5.68)	46.6307 (6.15)	82.5143 (8.47)
α_1	0.8593 (40.04)	0.8411 (37.59)	0.7174 (24.82)
Adjusted R^2	0.7379	0.7070	0.5139
Test of $\alpha_1 = 1$ t-statistic	6.56	7.10	9.78
p value	0.00	0.00	0.00
Panel C: Wilcoxon tests			
Wilcoxon statistic	10.7292	14.2374	23.12
p value at which the null hypothesis of identical distributions is rejected	0.0001	0.0001	0.0001
Panel D: Theil tests			
$\hat{\hat{\varepsilon}}^{DEA} = \mu_0 + \mu_1 \hat{\hat{\varepsilon}}^{trans\log} + \tau$			
Theil's statistic	−0.1473	0.5629	0.0832
p value at which the null hypothesis $\mu_1 = 1$ is rejected	0.8820	0.5734	0.9336
Panel E: Wooldridge tests			
$\hat{\varepsilon}^{Trans\log} = \beta_0 + \beta_1 \ln(teachers) + \beta_2 \ln(Support) + \beta_3 \ln(Admin) + \beta_4 \ln(Teachers^2) + \beta_5 \ln(Support^2) + \beta_6 \ln(Admin^2) + \beta_7 \ln(Teachers^{*}Support) + \beta_8 \ln(Support^{*}Admin) + \beta_9 \ln(Admin^{*}Teachers) + \gamma\{\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}(X_j;\beta))\} + \zeta$			
Independent variable			
$\ln(\hat{g}^{BM}(X_j) - \ln(\hat{g}(X_j;\beta))$	0.4180 (2.79)	0.6040 (3.97)	0.4451 (2.28)

$\hat{\varepsilon}^{trans\log}$ is computed as equal to $\ln(y_j) - \ln(\hat{g}^{trans\log}(X_j;\hat{\beta}))$; t-statistics are given in parentheses.

The last of the three tests motivated in Section 2.2.2 is based on Theil's (1950) test to evaluate 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, we first measure the difference $D_j = \hat{\hat{\varepsilon}}_j^{DEA} - \hat{\hat{\varepsilon}}_j^{PARAM}$ and then sort the data by $\hat{\hat{\varepsilon}}_j^{PARAM}$. Next, we assign a score $c_{ji} = 1, 0$

or -1 for each $i < j$ depending on whether $D_j - D_i > 0, = 0$ or < 0, respectively. Then 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 is indicative of a significant difference between the parametric frontier and the DEA frontier. The results of Theil's test are presented in panel D of Table 4. They indicate that the adequacy of Cobb–Douglas specification is rejected at the 5% level for the years 1994–1995, 1995–1996 but not for the year 1996–1997 ($p = 0.0791$). The results reported in panel D of Table 5 indicate that Theil's test fails to reject the null hypothesis of adequacy of the translog specification.

Finally, we present the results from the Wooldridge (1992) type test described in Section 2.3. We begin this test procedure by first estimating β by maximizing the loglikelihood $\ln L \equiv \sum_{j=1}^{n} \ln f(\ln y_j - \ln \hat{g}^{PARAM}(X_j; \beta))$ based on the probability density function specified in equation (14). Here $PARAM = Cobb\text{–}Douglas$ or *translog* as the case may be. The difference between the actual value ε and the predicted value $\hat{\varepsilon}^{PARAM} = \ln(y_j) - \ln(\hat{g}^{PARAM}(X_j; \hat{\beta}))$ is then calculated for each observation. Next, because of the computational complexity of solving the nonlinear embedded mathematical program in equation (13) to obtain the BM sieve estimator, we employ the DEA estimator as a substitute for $\ln(\hat{g}^{BM}(X_j))$ for illustration here. Finally, we repeat the parametric maximum likelihood estimation with $\hat{\varepsilon}_j^{PARAM}$ as the dependent variable and $\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}^{PARAM}(X_j; \hat{\beta}))$ included in addition to the independent variables used in the earlier estimation of the parametric function. The Wooldridge test is based on the t statistic of the coefficient on $\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}^{PARAM}(X_j))$. The test statistic is asymptotically normally distributed.

Panel E of Table 4 presents the results of the test of the Cobb–Douglas functional form. The t statistic of the coefficient on $\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}(X_j))$ is statistically significant at the 1% level for all three years of estimation (t statistic is 7.53, 7.72 and 3.61 for the years 1994–1995, 1995–1996 and 1996–1997, respectively) resulting in the rejection of the null hypothesis of the adequacy of the Cobb–Douglas form for representing the functional relationship between the number of students and the number of teachers, support and administrative staff. Panel E of Table 5 presents the results of the test of the translog functional form. The t statistic of the coefficient on $\ln(\hat{g}^{BM}(X_j)) - \ln(\hat{g}(X_j))$ is significantly different from zero at the 1% level for all the three years of data that we analyzed. Again, this leads to the rejection of the adequacy of the translog functional form.

4. Conclusion

In this paper we have extended the DEA methodology to present test procedures for evaluating how well a parametric functional form approximates a monotone and concave (or convex) function. In parametric estimation, researchers often maintain assumptions about the parametric form of the relationship between the exogenous and endogenous variables. However, the question of whether the maintained assumption about the parametric form is justifiable is often left unanswered. Theory may provide only limited guidance; for instance, only a monotone and concave production function relating output to input variables may be indicated for productivity analysis. Since a parametric representation of the production function may provide parsimony and tractability in analyzing its properties, test procedures to evaluate the adequacy of alternative parametric forms to represent such relations is of considerable interest.

We illustrated our proposed methods to evaluate the adequacy of an assumed parametric form with an application to the estimation of an education production function based on data collected from the Texas school districts for three years. Researchers have used both the Cobb–Douglas and the translog functional forms to estimate the education production function. Our empirical results indicate that both the Cobb–Douglas and the translog parametric forms are inadequate in approximating a monotone increasing and concave educational production function for the Texas school districts.

Notes

1. Theoretical considerations justify only minimal structure such as monotonicity and concavity of the production frontier, and Banker (1993) has proved that DEA provides a consistent estimator when only this minimal structure is imposed. We are not aware of any theoretical work that results in the translog specification being the *minimal* structure for the production correspondence. Consequently, we do not see any justifiable way to maintain the hypothesis that the production correspondence is translog rather than monotone increasing and concave.
2. The analysis depends fundamentally on the assumption that S_U is finite. This assumption is not satisfied by normal-half normal and normal-exponential convolutions that are commonly specified in the composed error literature. In reality, however, physical limits on the maximum output that can be produced with existing facilities imply an upper bound S_U such that $f(\epsilon) = 0$ for all $\epsilon > S_U$. For instance, let each firm comprise of a finite number of plants $p = 1, \ldots, P$, with finite physical capacities K_p. Then $f(\epsilon) = 0$ for all $\varepsilon > \sum_{p=1}^{P} K_p$. Observe that it is not sufficient that $f(\epsilon) \to 0$ as $\epsilon \to \infty$ in case $f(\epsilon)$ remains strictly positive.
3. If we write $\varepsilon = v - u$ as a composed error such that v is a two-sided error component and u is a one-sided error component, then there are convolutions of v and $-u$ for which $f(\varepsilon) = 0$ at $\varepsilon = S_U$. For example, when v is distributed as truncated normal in the range $(-\infty, S_U]$ and u is distributed as gamma over the range of $[0, \infty)$, the composed error $\varepsilon = v - u$ has a distribution that has $f(\varepsilon = S_U) = 0$. More generally, $f(\varepsilon) = \int_{\varepsilon}^{S_U} f_v(v) f_u(v - \varepsilon)\, dv$ and $f(\varepsilon) = 0$ when evaluated at $\varepsilon = S_U$.
4. Our empirical results are robust to the inclusion of these two outlier observations.

References

Ahn, T., A. Charnes and W. W. Cooper. (1988). "Some Statistical and DEA Evaluations of Relative Efficiencies of Public and Private Institutions of Higher Learning." *Socio-Economic Planning Science,* 259–269.

Aigner, D., C. A. K. Lovell and P. Schmidt. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6, 211–237.

Banker, R. D. (1993). "Maximum Likelihood, Consistency and Data Envelopment Analysis." *Management Science* 39, 1265–1273.

Banker, R. D. (1996). "Hypothesis Tests Using Data Envelopment Analysis." *Journal of Productivity Analysis* 7, 139–159.

Banker, R. D., A. Charnes and W. W. Cooper. (1984). "Models for the Estimation of Technical and Scale Efficiencies in Data Envelopment Analysis." *Management Science* 30, 1078–1092.

Banker, R. D. and A. Maindiratta. (1992). "Maximum Likelihood Estimation of Monotone and Concave Production Frontiers." *Journal of Productivity Analysis* 3, 401–415.

Banker, R. D. and R. Natarajan. (2000). "Evaluating Contextual Variables Affecting Productivity using Data Envelopment Analysis." Working paper, The University of Texas at Dallas.

Bessent, A., W. Bessent, J. Kennington and B. Reagan. (1982). "An Application of Mathematical Programming to Assets Productivity in the Houston Independent School District." *Management Science* 28, 1255–1367.

Butler, R. and D. Monk. (1985). "The Cost of Public Schooling in New York State: The Role of Scale and Efficiency in 1978–79." *Journal of Human Resources* 20, 361–381.

Callan, S. and R. Santerre. (1990). "The Production Characteristics of Local Public Education: A Multiple Product and Input Analysis." *Southern Economic Journal* 57(2), 68–480.

Chizmar, J. F. and T. A. Zak. (1983). "Modeling Multiple Outputs in Educational Production Functions." *American Economic Review* 73, 18–28.

Cooper, W. W., R. G. Thompson and R. M. Thrall. (1996). "Introduction: Extensions and New Developments in DEA." *Annals of Operations Research* 66, 3–45.

Davidson, R. and J. G. MacKinnon. (1981). "Several Model Specification Tests in the Presence of Alternative Hypotheses." *Econometrica* 49, 781–793.

Farrell, M. J. (1957). "The Measurement of Production Efficiency." *Journal of the Royal Statistical Society Series A* 120, 254–290.

Greene, W. H. (1980). "Maximum Likelihood Estimation of Econometric Frontier Production Functions." *Journal of Econometrics* 13, 27–56.

Grosskopf, S., K. Hayes, L. Taylor and W. Weber. (1999). "On the Determinants of School District Efficiency: Competition and Monitoring." Working paper, Oregon State University.

Gstach, D. (1998). "Another Approach to Data Envelopment Analysis in Noisy Environments: DEA+." *Journal of Productivity Analysis* 9, 161–176.

Gyimah-Brempong, K. and A. Gyapong. (1992). "Elasticities of Factor Substitution in the Production of Education." *Economics of Education Review,* 205–217.

Iman, R. L. and W. J. Conover. (1979). "The Use of Rank Transform in Regression." *Technometrics* 499–509.

Jimenez, E. (1986). "The Structure of Education Costs: Multiple Cost Functions for Primary and Secondary Schools in Latin America." *Economics of Education Review* 15, 25–39.

Meeusen, W. and J. van den Broeck. (1977). "Efficiency Estimation for Cobb–Douglas Production Function with Composed Error." *International Economic Review* 18, 435–455.

Ohlson, J. A., P. Schmidt and D. A. Waldman. (1980). "A Monte Carlo Study of Estimators of Stochastic Frontier Production Functions." *Journal of Econometrics* 13, 67–82.

Ray, S. (1991). "Resource-use Efficiency in Public Schools: A study of Connecticut Data." *Management Science* 37, 1620–1628.

Richmond, J. (1974). "Estimating the Efficiency of Production." *International Economic Review* 15, 515–521.

Sarath, B. and A. Maindiratta. (1997). "On the Consistency of Maximum Likelihood Estimation of Monotone and Concave Production Frontiers." *Journal of Productivity Analysis* 8, 239–246.

Schmidt, P. (1976). "On the Statistical Estimation of Parametric Frontier Production Functions." *Review of Economics and Statistics* 58(1), 238–239.

Sengupa, J. and R. Sfeir. (1986). "Production Frontier Estimates of Scale in Public Schools in California." *Economics of Education Review,* 297–307.

Smet, M. and W. Nonneman. (1988). "Economies of Scale and Scope in Flemish Secondary Schools." *Applied Economics* 30, 1251–1258.

Theil, H. (1950). "A Rank-invariant Method of Linear and Polynomial Regression Analysis." *I. Proc. Kon. Ned. Akad. V. Wetensch.* A 53, 386–392.

Wooldridge, J. M. (1992). "A Test for Functional Form Against Nonparametric Alternatives." *Econometric Theory* 8, 452–475.