



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

A Comparison Of Various Frontier Estimation Methods Under Differing Data Generation Assumptions

By

Daniel M. Settlage, Bruce L. Dixon, and Michael R. Thomsen

SP 06 2000

2000

University of Arkansas

Department of Agricultural Economics and Agribusiness
Agricultural Building Room 221
Fayetteville, AR 72701

A Comparison Of Various Frontier Estimation Methods Under Differing Data Generation Assumptions

Daniel M. Settlege, Bruce L. Dixon, and Michael R. Thomsen

Abstract:

Monte Carlo methods examine the accuracy of several production frontier approximating forms, estimators and methods to rank firms by level of predicted technical efficiency. Results show stochastic frontier methods superior to data envelopment analysis. The Cobb-Douglas approximating form is superior to the translog or generalized Leontief.

Presented as a selected paper at the Southern Agricultural Economics Association annual meetings, January 30-Feb. 2, 2000, Lexington KY.

Contact author Bruce Dixon bdixon@comp.uark.edu

Daniel M. Settlege is a former graduate research assistant in the Department of Agricultural Economics and Agribusiness at the University of Arkansas and is now a Ph.D. graduate research fellow in the Department of Agricultural Economics at Purdue University. Bruce L. Dixon and Michael R. Thomsen are professor and assistant professor of agricultural economics in the Department of Agricultural Economics and Agribusiness at the University of Arkansas.

A Comparison Of Various Frontier Estimation Methods Under Differing Data Generation Assumptions

Introduction

A major motivation for employing frontier functions is the measure of firm-specific efficiency that can be obtained from them (Førsund et al., 1980). Frontier functions are often used to estimate the extent of inefficiency in an industry or firm (e.g., Hjalmarsson et al., 1996), or to rank a sample of firms in order of efficiency (e.g., Gong and Sickles, 1989, 1992). Most frontier estimation methods fall into two categories: mathematical programming, commonly referred to as data envelopment analysis (DEA), and statistically based techniques such as corrected ordinary least squares (COLS) or maximum likelihood for estimating stochastic frontier functions.

This study uses a Monte Carlo approach to compare the accuracy of three modeling techniques, three approximating forms and two stochastic estimators applied to differing data generation processes (DGP's). The literature on frontier methods, with a few notable exceptions, has been sparse with regard to Monte Carlo comparisons. A notable exception are a pair of articles by Gong and Sickles (1989, 1992). Both studies utilize panel data and employ a variety of estimators and approximating forms, and compare models based on their ability to rank firms by level of technical efficiency.

Gong and Sickles (1989) applied several stochastic cost and production frontier estimators and approximating forms to data sets generated under various input elasticity conditions. Furthermore, sample size and simulated technical inefficiency distribution were varied. The present study is an extension on Gong and Sickles (1992) which compared the performance of cost stochastic frontiers and DEA on various panel data sets. The data were constructed using various assumptions about production technology and producer behavior. Our study extends this

work by relaxing the data requirements (cross sectional as opposed to panel sample) and by estimating production (rather than cost) functions using different functional forms. We examine the implications of simultaneity bias in estimating frontier production functions by creating data that conform with producer optimization given producer knowledge of their particular level of technical efficiency. An analysis of variance approach is employed in testing hypotheses about the impact of various factors on measuring the accuracy of ranking firms by their technical efficiency level. This study provides researchers additional information on how to utilize frontier methods more confidently and thus be able to rank firms by level of technical efficiency more accurately.

Data and Methods

The methodology of the experiment can be broken down into three main parts. First the producer input, output, and efficiency data are drawn under multiple data generation conditions. Second, several types of production frontier methods are applied to each data set and then estimated firm level efficiency scores are generated. Rank correlation coefficients (RCC) as defined in Mood, Graybill and Boes (1974) are calculated between the true and estimated levels of firm specific inefficiency. Finally, the RCC observations are used to test hypotheses about factors influencing ranking accuracy using a seemingly unrelated regression (SUR) analysis framework.

Data Generation

The underlying production technology used to generate the producer level data is a three input, one output stochastic production frontier. The functional form used to generate observations in the Monte Carlo experiments is the generalized constant ratio elasticity of substitution, homothetic (CRESH) production function (Mukerji, 1963) written as:

$$y_i = [\ddot{a}_1 x_{i1}^{-\tilde{n}_1} + \ddot{a}_2 x_{i2}^{-\tilde{n}_2} + \ddot{a}_3 x_{i3}^{-\tilde{n}_3}]^{-1/\tilde{n}} \exp(v_i - u_i) \quad (1)$$

In (1), y_i is the output for the i^{th} firm, x_{i1} , x_{i2} , and x_{i3} are the three inputs, and \bar{a}_1 , \bar{a}_2 , \bar{a}_3 , \bar{n}_1 , \bar{n}_2 , \bar{n}_3 , and \bar{n} are the technology specific parameters. Following Hanoch (1971), $\bar{a}_1 + \bar{a}_2 + \bar{a}_3 = 1$ and $\bar{a}_1 = 0.3$, $\bar{a}_2 = 0.3$ and $\bar{a}_3 = 0.4$ to conform with Guilkey et al. (1983) and subsequent studies such as Dixon, Garcia, and Anderson (1987) and Gong and Sickles (1989,1992). The sample size is set at 100 observations for estimating a given frontier model. The three sets of technology parameters used are given in table 1 and are taken from Guilkey et al. (1983). Each of the three technologies exhibit differing input substitutabilities. The ease of substitutability for each technology is indicated by the overall mean of the Allen-Uzawa partial elasticities of substitution in table 2.

The statistical noise, represented by v_i in (1), is simulated as an independently and identically distributed (IID) $N(0,0.01)$ across all observations as well as independently of technical inefficiency (u_i). Technical inefficiency is simulated using three different non-negative, one-sided distributions: exponential (EX), truncated normal (TN), and half normal (HN). The EX error vector was drawn from a distribution with a characteristic parameter $\bar{\epsilon} = 3.165$, which implies a mean of 0.316. The TN error vector was calculated as one plus a random vector drawn from a truncated standard normal distribution (truncated from the left at -1). This sum was then multiplied by 0.35. The HN error vector was calculated as the absolute value of a normally distributed variable with a mean of 0 and a variance of 0.299.

Technical efficiency (TE) is computed as $\exp(-u_i)$. The sample mean and variance of the u_i and TE error terms are presented in table 3. Depending on the technology, average levels of technical efficiency are between 61 and 78 percent. Previous studies (Olsen et al. (1980), Coelli (1995), and Khoju and Dixon (1994)) have found that the ratio of variances of noise (v_i) to TE ($\exp(-u_i)$) is a significant factor in the performance of an estimator. Hence the variances of all three TE distributions are fixed at approximately the same level, forcing the variance ratio to be

fixed. As a consequence, the means of the TE terms vary slightly between different distributions.

A set of three input vectors (x_1 , x_2 , and x_3) of 100 observations are generated from a multivariate lognormal distribution as reported in Driscoll and Boisvert (1991). The observations on y_i are generated by using the x_{i1} , x_{i2} , and x_{i3} , v_i , one of the three sets of u_i (EX, TN or HN), and one of the three sets of technology parameters. One hundred observations on x_{i1} , x_{i2} , x_{i3} , v_i , and u_i compose one sample.

All possible combinations of the three technologies and three technical inefficiency distributions result in nine data sets, each with 100 observations. Then, each data set is replicated 100 times by randomly drawing observations on the statistical noise term (v_i). For one half of the samples, referred to as the uncorrelated data, the u_i are fixed in repeated samples for a given technical efficiency distribution, i.e. the same set of 100 observations for any given TE distribution is used repeatedly. These data sets exhibit minimal correlation between the x_{ik} 's ($k = 1,2,3$) and u_i . An additional nine sets of data, referred to as the correlated data, are constructed that have correlation between the x_{ik} 's and u_i . These correlated data simulate the correlation between the inputs and technical efficiency that would arise if producers were aware of their efficiency level and optimized accordingly. Under these conditions, econometric frontier estimation would become tainted by simultaneous equations bias.

These nine sets of correlated data are constructed in a similar manner to the uncorrelated data. The correlation is induced by solving for the x_k 's assuming producers maximize expected profit for a given input price vector. There are nine distinct sets of x_i for the correlated data since the x_k vary by the assumed distribution of the u_i and by the technology scheme. The mean level of correlation between all three input vectors and TE in the uncorrelated data sets¹ is -0.05239 with

¹ The overall mean of all nine correlation coefficients between the three inputs and the three possible TE vectors (EX, HN, and TN).

a variance of 0.009944, whereas the overall mean level of correlation between all three input vectors and TE in the correlated data sets is 0.686852 with a variance of 0.051493. The mean Allen-Uzawa partial elasticities of substitution for the correlated data were similar to the mean Allen-Uzawa partial elasticities of substitution for the uncorrelated data reported in table 2.

Frontier Estimation

There are three frontier estimation methods are used in this study. DEA is a nonparametric approach to the estimation of frontier functions. DEA's major advantage is that no explicit functional form is imposed on the data other than that the frontier is quasi-convex (Bauer, 1990). This allows the shape of the DEA frontier to be extremely flexible. DEA does not require any distributional assumptions about the way in which producer inefficiency is distributed across the sample, but is critiqued by econometricians because the estimated frontier lacks statistical properties.

COLS is a stochastic method used to impart some statistical properties to the frontier (Greene, 1997). The model's parameters (excluding the constant term) can be consistently estimated using OLS (Greene, 1997). To estimate the constant, the OLS intercept is shifted upward until only one or more observations lie on the frontier and all others lie below it. This method ignores the possibility of exogenous shocks, measurement error, and statistical noise since all deviations from the frontier are attributed to the one-sided error component that captures producer inefficiency.

The stochastic frontier specification includes the effects of nontechnically related random factors. This model utilizes a two part, composed error term with both symmetric and one-sided error components. The symmetric component captures the effects of statistical noise, measurement error, and random shocks not associated with firm specific characteristics. The one-

sided component captures the inefficiency effect and is usually incorporated into the econometric framework by assuming a one-sided distribution (e.g., HN, EX, and TN). Given the stochastic specification in terms of specific distributions, maximum likelihood (ML) estimation may be used.

Each data set is modeled using three approximating forms: the translog (TL), generalized Leontief (GL), and Cobb-Douglas (CD). Each approximating form is estimated by two estimators, ML and COLS. The ML estimator utilizes a two part, composed error specification as in Greene (1995). The one-sided error component in the ML estimator is modeled as an exponential. Residuals from the ML estimated models are used to estimate technical efficiency. These residuals are predictions of the composed error term ($\exp(v_i - u_i)$) and can be decomposed into two parts: technical inefficiency and statistical noise. The Jondrow et al. (1982) method of recovering firm specific measures of technical inefficiency from the residuals is used in this study. For COLS, firm level inefficiency is ranked by the size of the residual.

In addition to the six stochastic methods above, a DEA model is also used on each sample and replication. The DEA approach in this study is an output oriented model that satisfies variable returns to scale (see Charnes et al. (1994)). The technical efficiency measure for the j^{th} production plan is given by $1/\theta_j$ where j , like i , goes from 1 to 100. The value θ_j is the solution to

the following linear programming model.

$$\begin{aligned} \theta_j &= \max \theta + \alpha s^y + \alpha \sum_{k=1}^3 s_k \\ \text{s.t.} \\ \sum_{i=1}^n z_i x_{ik} + s_k &= x_{jk} \quad \forall k = 1, 2, 3 \\ \sum_{i=1}^n z_i y_i - s^y &= \theta y_j \\ \sum_{i=1}^n z_i &= 1 \\ z_i \geq 0 \quad \forall i, s_k &\geq 0 \quad \forall k, s^y \geq 0 \end{aligned}$$

In this model, i, j index the production plans, α is a non-Archimedian constant value, x_{ik} is the level of the k^{th} input used in the i^{th} production plan, and the z_i are intensity variables that enable the creation of feasible production plans through the radial contraction, expansion, or convex combination of observed production plans. The s^y and s_k are slack variables.

For each estimation method applied to a given sample, a RCC is computed between the true and predicted level of technical efficiency. The RCC is a distribution free measure of the performance of the estimation methods which relates the strength of the monotonic relationship between true and predicted TE. The RCCs are the dependent variables in the SUR analysis.

SUR Analysis Framework

The hypotheses of the study are tested using a SUR framework. The ability of SUR to model inter-equation error term correlations is important in the context of this study because such correlations are generated by the structure of the experiment. By design, each individual data set (combination of technology, correlation structure, technical inefficiency distribution, and noise term) is estimated seven ways. This repeated use of each sample enables making direct comparisons between estimation methods, but it induces correlation among the parameter estimates used to compare accuracy. In this setting, the SUR method offers more efficient hypotheses tests than does a simple ANOVA or OLS system that ignores the correlations when cross-equation parameter restrictions are maintained.

The dependent variables in the SUR model are the RCC from the estimated frontier and DEA models. The observations on the RCCs can be partitioned into seven separate vectors that constitute the dependent variables of the seven equations in the SUR model. Let the seven SUR equations be indexed by the letter m . For $m = 1$, the RCC observations are for those frontier models estimated by ML as a TL approximating form. Similarly, for $m = 2$, TL models are

estimated by COLS. For $m = 3$ and 4, GL models are estimated by ML and COLS, respectively. For $m = 5$ and 6, CD models are estimated by ML and COLS, respectively. Finally, for $m = 7$, DEA models are estimated. The independent variables in each equation are identical. They are composed of eight binary variables corresponding to the data generation conditions and a constant term representing the equation mean.

Let \hat{a}_{mp} denote the p^{th} coefficient in the m^{th} equation of the SUR system. The independent variables are binary and they are structured such that: \hat{a}_{m2} represents the deviation from the equation mean when the input variables and TE terms are not correlated and \hat{a}_{m3} is the deviation when input and TE terms are correlated. The \hat{a}_{m4} , \hat{a}_{m5} and \hat{a}_{m6} represent the derivations from the equation mean for technologies 1, 2, and 3, respectively. Likewise, \hat{a}_{m7} , \hat{a}_{m8} and \hat{a}_{m9} are the deviations when technical inefficiencies are distributed as EX, TN and HN, respectively. Thus, the equation restrictions are: $\hat{a}_{m2} + \hat{a}_{m3} = 0$; $\hat{a}_{m4} + \hat{a}_{m5} + \hat{a}_{m6} = 0$ and $\hat{a}_{m7} + \hat{a}_{m8} + \hat{a}_{m9} = 0$.

Using within equation restrictions and an identical set of regressors for each equation, OLS gives identical parameter estimates to SUR. However, the SUR approach is used because information contained in the off-diagonal elements of the error covariance matrix is used in hypothesis testing. In doing parameter hypothesis tests from multiple equations, the feasible generalized least squares (FGLS) parameter matrix provides the correct covariance needed for the tests. Efficiency gains are realized when cross equation hypothesis restrictions are imposed to test for various forms of interaction effects.

Results and Analysis

The summary statistics for the RCC observations partitioned by dependent variable are found in table 4. As is evident from the table, ML slightly outperformed COLS for a given approximating form. In addition, DEA performed poorly compared with the other methods.

The parameter estimates and relevant statistics for the seven SUR equations are presented in table 5. Equation fit was generally good and most coefficients were significant at the 0.05 level or less. Cross equation parameter tests reflect the overall effect of data generation processes and modeling choices.

Single Variable Tests of Hypotheses

Tests of the mean of each of the eight binary categorical variables across all seven equations yield information about the overall effect of a particular aspect of the DGP on estimating model performance in general. Table 6 lists the test variables, values of test statistics, standard errors, and the associated z and p-values. As the table illustrates, all hypotheses tested were significant at the 0.05 level or better.

Results on hypothesis 1 indicate that the models performed substantially better with input vectors that were uncorrelated with the level of technical efficiency. This was true for each equation. On average, the RCC of the models estimated under uncorrelated data conditions exceeded the RCC of the models estimated with correlated data by 0.69574. This is a large difference in light of the RCC being bound between negative one and one. This result is consistent with other Monte Carlo studies (e.g. Gong and Sickles (1992)).

Mean RCC for technologies 1 and 2 were distinctly inferior to technology 3. The mean value of model performance degradation under T1 and T2 are -0.01385 and -0.07931 respectively. T3 enhanced mean RCC by an average of 0.09316. This is a substantial performance increase, which indicates that the models performed the best under conditions of difficult input substitution.

Of the three technical inefficiency error distributions, EX was the form that the models had the most difficulty approximating. Average performance degradation was -0.02527. Although the magnitude of this number is relatively small, the negative sign is contrary to expectations. Since the ML estimator assumed that technical inefficiency was distributed exponentially, intuition suggests that it should perform the best when technical inefficiency was actually generated as an exponential. Additional computations show the average performance degradation for ML models with technical inefficiency generated as an exponential is -0.02246. The HN and TN were approximated better than the EX, their overall coefficient estimates being 0.01330 and 0.01197 respectively.

One aspect of the tests presented in table 6 is that they are not independent from each other. Hypotheses within a given DGP treatment effect are related. Hence, the overall significance level of the tests in assessing the significance of the treatment effects is diminished from the nominal p-value. To gain information on the overall significance of the tests, Wald tests are employed. As indicated in table 7, all three sets of restrictions are significant at the 0.05 level. This indicates that the differences found in the data generation categories are statistically significant.

Multiple Variable Hypothesis Tests

Table 6 presents tests of one independent variable over all seven equations. It does not explicitly address comparisons within a treatment effect nor does it examine the relative accuracy of differing approximating forms or estimation approaches. To test these hypotheses, multiple independent variable combinations are tested. This allows testing if any given estimator,

approximating form, or DGP give results that are superior to an alternative. Table 8 lists the hypotheses, test statistics, standard errors, asymptotic t-values, and the associated p-values.

The preferred method of stochastic estimation (ML vs. COLS) is tested by hypothesis 9. ML provides an average RCC performance increase over COLS of 0.03366. The magnitude of this significant statistic is fairly small, indicating ML is not vastly superior to COLS. By utilizing hypotheses 9, 10, and 11, it is clear that ML outperforms COLS, and both substantially outperform DEA. Gong and Sickles (1992) found a DEA model to be unaffected by input vector correlation with technical inefficiency in estimating cost functions.

Hypotheses 12, 13, and 14—which test approximating form—indicate the approximating forms can be ranked from best to worst performing as follows: CD, TL, and GL. The differences are fairly large, indicating that the CD outperformed the other two approximating forms by a wide margin. These results may be due to the lower number of parameters needed to estimate the CD as opposed to the TL or GL. Clearly, the TL and GL are characterized by higher levels of multicollinearity than the CD.

In a similar manner, model performance by technology can be ranked. The order of ranking from best to worst performance is T3, T1, and T2. All differences are significant at the 0.05 level or better. Hence, the models performed best under conditions of difficult input substitution (T3), followed by easy substitution (T1), and finally mixed input substitution (T2).

Model performance under different inefficiency distributions can also be ranked. The models did the worst under the EX distribution, and there was no significant difference at the 0.05 level between the HN and TN distributions.

Discussion

Model performance under differing technical inefficiency distributions was surprising. Since the ML estimator assumed an exponential distribution, intuition indicates it would approximate the exponential data sets well, but exactly the opposite was true. With regard to underlying production technology, overall model performance was enhanced by difficult input substitutability, while mixed substitution tended to degrade performance.

The performance of the approximating forms was also surprising. The Cobb-Douglas performed better than the two more flexible forms, even though the CD is simply a restricted version of the TL. Apparently the performance of the TL was degraded due to over-parameterization (multicollinearity). Across all DGP's, DEA performed worse on average than all combinations of approximating form/estimator except the GL/COLS combination, with which its performance is not statistically different.

Our results indicate the applied researcher may wish to choose the approximating form based partially upon the intended use of the model. When the analyst's main objective is to examine the level and ordering of technical efficiency, these results suggest that the simple CD functional form is best. However, if the object of the study is to estimate elasticities of substitution, then a CD is obviously inadequate.

Choice of stochastic estimator was not as important for accurately estimating RCC as expected. Nonetheless, both stochastic estimators were almost always better than DEA. The ML estimator outperformed COLS in terms of RCC, but difference in accuracy was small. This suggests that the gains from moving from the simple COLS estimator to the more complex ML estimator are minimal.

Overall, the largest factor effecting model performance is input vector/technical efficiency correlation. Correlated data were the single largest factor in the overall performance of any

model, with frontiers estimated under uncorrelated conditions outperforming frontiers estimated under correlated conditions by a wide margin. The mean RCC for models without input and TE correlation is 0.69574 higher than models with correlation. This clearly suggests that simultaneity bias is a large problem that can result in distinctly inferior results. The empirical researcher may wish to test to determine if the exogenous variables and error terms are independent of each other. Specification tests such as the Hausman test and the RESET (regression error specification test) can be employed to determine if correlation is present.

DEA has been proposed as a possible alternative to econometric methods under input and TE correlation. The performance of DEA was actually worse instead of better in most cases in this study. Other studies using cost functions instead of production functions found DEA to be unaffected by simultaneity problems. Our results indicate that using production functions on such data may lead to poor results. Such a finding argues in favor of estimating cost functions when there is doubt about input exogeneity.

Table 1. Technology Parameters

Technology	Parameters			
	$\tilde{\eta}_1$	$\tilde{\eta}_2$	$\tilde{\eta}_3$	$\tilde{\eta}$
T1	-0.20	-0.30	-0.40	-0.90
T2	-0.80	-0.10	-0.20	-0.80
T3	1.00	2.00	3.00	2.50

Table 2. Mean Allen-Uzawa Partial Elasticities of Substitution (No Input and TE Correlation)

Input Pair	Technology		
	T1	T2	T3
1,2	1.1735	1.4196	0.4851
1,3	1.3690	1.5970	0.3639
2,3	1.5646	0.3549	0.2426

Table 3. Mean and Variance of Technical Inefficiency Error Terms

	$-u_i$ (inefficiency)			$\exp(-u_i)$ (efficiency)		
	EX	TN	HN	EX	TN	HN
Sample Mean	0.2964	0.5377	0.3746	0.7753	0.6132	0.7194
Sample Variance	0.1016	0.1052	0.1045	0.0375	0.0334	0.0378

Table 4. Descriptive Statistics on Rank Correlation Coefficients Partitioned by Dependent Variable

Dependent Variable	Mean	Std. Dev.	Minimum	Maximum	Observations ¹
TLML	0.51282	0.36537	-0.20616	0.94314	1800
TLCOLS	0.51337	0.35912	-0.13134	0.92868	1800
GLML	0.47218	0.47176	-0.96864	0.95154	1800
GLCOLS	0.39385	0.49043	-0.60922	0.93591	1800
CDML	0.57835	0.34741	-0.10973	0.96254	1800
CDCOLS	0.55516	0.34603	-0.29804	0.96268	1800
DEA	0.39917	0.39593	-0.49455	0.86347	1800

¹Number of observations on RCC.

Table 5. SUR Model Parameter Estimates And Regression Statistics¹

Equation 1: Translog Estimated by Maximum Likelihood (TLML)

Equation $R^2 = 0.89$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₁₁	0.51282	0.00291	176.20500	0
NC	B ₁₂	0.33796	0.00291	116.12300	0
CO	B ₁₃	-0.33796	0.00291	-116.12300	0
T1	B ₁₄	0.01193	0.00412	2.89900	0.0037
T2	B ₁₅	-0.07581	0.00412	-18.41900	0
T3	B ₁₆	0.06388	0.00412	15.52000	0
EX	B ₁₇	-0.03002	0.00412	-7.29500	0
TN	B ₁₈	-0.00152	0.00412	-0.36800	0.7126
HN	B ₁₉	0.03154	0.00412	7.66300	0

Equation 2: Generalized Leontief Estimated by Maximum Likelihood (GLML)

Equation $R^2 = 0.55$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₃₁	0.47218	0.00748	63.10100	0
NC	B ₃₂	0.31976	0.00748	42.71300	0
CO	B ₃₃	-0.31976	0.00748	-42.71300	0
T1	B ₃₄	-0.11505	0.01058	-10.87100	0
T2	B ₃₅	-0.06274	0.01058	-5.92800	0
T3	B ₃₆	0.17778	0.01058	16.80000	0
EX	B ₃₇	-0.03089	0.01058	-2.91900	0.0035
TN	B ₃₈	0.07856	0.01058	7.42300	0
HN	B ₃₉	-0.04767	0.01058	-4.50500	0

Equation 3: Cobb-Douglas Estimated by Maximum Likelihood (CDML)

Equation $R^2 = 0.88$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₅₁	0.57835	0.00327	176.95900	0
NC	B ₅₂	0.30761	0.00327	94.12100	0
CO	B ₅₃	-0.30761	0.00327	-94.12100	0
T1	B ₅₄	-0.05186	0.00462	-11.22000	0
T2	B ₅₅	-0.06335	0.00462	-13.70500	0
T3	B ₅₆	0.11520	0.00462	24.92500	0
EX	B ₅₇	-0.01462	0.00462	-3.16400	0.0016
TN	B ₅₈	0.01119	0.00462	2.42100	0.0155
HN	B ₅₉	0.00343	0.00462	0.74300	0.4574

Equation 4: Translog Estimated by Corrected Ordinary Least Squares (TLCOLS)

Equation $R^2 = 0.89$

Variable	Coefficient Value	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₂₁	0.51337	0.00279	183.78900	0
NC	B ₂₂	0.33152	0.00279	118.68500	0
CO	B ₂₃	-0.33152	0.00279	-118.68500	0
T1	B ₂₄	0.00494	0.00395	1.25100	0.2110
T2	B ₂₅	-0.07701	0.00395	-19.49400	0
T3	B ₂₆	0.07207	0.00395	18.24300	0
EX	B ₂₇	-0.03281	0.00395	-8.30500	0
TN	B ₂₈	-0.01593	0.00395	-4.03200	0.0001
HN	B ₂₉	0.04874	0.00395	12.33700	0

Equation 5: Generalized Leontief Estimated by Corrected Ordinary Least Squares (GLCOLS) Equation $R^2 = 0.89$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₄₁	0.39385	0.00389	101.21100	0
NC	B ₄₂	0.44944	0.00389	115.49600	0
CO	B ₄₃	-0.44944	0.00389	-115.49600	0
T1	B ₄₄	0.02770	0.00550	5.03300	0
T2	B ₄₅	-0.12665	0.00550	-23.01400	0
T3	B ₄₆	0.09896	0.00550	17.98100	0
EX	B ₄₇	-0.03928	0.00550	-23.01400	0
TN	B ₄₈	-0.02775	0.00550	-5.04200	0
HN	B ₄₉	0.06703	0.00550	12.18000	0

Equation 6: Cobb-Douglas Estimated by Corrected Ordinary Least Squares (CDCOLS) Equation $R^2 = 0.88$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₆₁	0.55516	0.00286	194.37100	0
NC	B ₆₂	0.31759	0.00286	111.19300	0
CO	B ₆₃	-0.31759	0.00286	-111.19300	0
T1	B ₆₄	-0.03425	0.00404	-8.47900	0
T2	B ₆₅	-0.04821	0.00404	-11.93600	0
T3	B ₆₆	0.08246	0.00404	20.41500	0
EX	B ₆₇	-0.03524	0.00404	-8.72300	0
TN	B ₆₈	0.00716	0.00404	1.77200	0.0765
HN	B ₆₉	0.02808	0.00404	6.95200	0

Equation 7: Estimated by DEA

Equation $R^2 = 0.88$

Variable	Coefficient	Estimate	St. Er.	b/St.Er.	P[Z >z]
CON	B ₇₁	0.39917	0.00262	152.28200	0
NC	B ₇₂	0.37119	0.00262	141.61000	0
CO	B ₇₃	-0.37119	0.00262	-141.61000	0
T1	B ₇₄	0.05963	0.00371	16.08700	0
T2	B ₇₅	-0.10137	0.00371	-27.34600	0
T3	B ₇₆	0.04174	0.00371	11.25900	0
EX	B ₇₇	0.00594	0.00371	1.60100	0.1094
TN	B ₇₈	0.04139	0.00371	11.16500	0
HN	B ₇₉	-0.04732	0.00371	-12.76600	0

¹ St. Er. is the coefficients' estimated standard error.

b/St. Er. is the estimate divided by the estimated standard error, and is asymptotically distributed as a standard normal.

P[|Z|>z] is the probability level at which the null hypothesis that the coefficient is equal to zero fails to be rejected.

Table 6. Hypothesis Tests of Individual Independent Variables

Null Hyp.	Test Coefficients	Estimated Test Coefficient (b)	Std. Error (S.E.)	z=b/S.E.	P[Z >z]
Hyp. 1 NC	$\sum_{m=1}^7 B_{m2}/7 = 0$	0.34787	0.00281	123.627	0
Hyp. 2 CO	$\sum_{m=1}^7 B_{m3}/7 = 0$	-0.34787	0.00281	-123.627	0
Hyp. 3 T1	$\sum_{m=1}^7 B_{m4}/7 = 0$	-0.01385	0.00398	-3.480	0.0005
Hyp. 4 T2	$\sum_{m=1}^7 B_{m5}/7 = 0$	-0.07931	0.00398	-19.929	0
Hyp. 5 T3	$\sum_{m=1}^7 B_{m6}/7 = 0$	0.09316	0.00398	23.410	0
Hyp. 6 EX	$\sum_{m=1}^7 B_{m7}/7 = 0$	-0.02527	0.00398	-6.351	0
Hyp. 7 HN	$\sum_{m=1}^7 B_{m8}/7 = 0$	0.01330	0.00398	3.342	0.0008
Hyp. 8 TN	$\sum_{m=1}^7 B_{m9}/7 = 0$	0.01197	0.00398	3.009	0.0026

Table 7. Wald Tests for Simultaneous Significance of Hypotheses

Null Hypotheses	Wald Statistic	Degrees of Freedom	Probability From χ^2
Hyp. 1 and 2	15,284	1	0
Hyp. 3, 4, and 5	638	2	0
Hyp. 6, 7, and 8	588	3	0

Table 8. Joint Hypothesis Tests

Null Hyp.	Test Coefficient	Estimated Test Coefficient (b)	Std. Error (S.E.)	b/S.E.	P[Z >z]
Hyp. 9 ML-COLS	$(B_{11}+B_{31}+B_{51})-(B_{21}+B_{41}+B_{61}) = 0$	0.03366	0.00260	12.927	0
Hyp. 10 ML-DEA	$((B_{11}+B_{31}+B_{51})/3)-B_{71} = 0$	0.12195	0.00383	31.878	0
Hyp. 11 COLS-DEA	$((B_{21}+B_{41}+B_{61})/3)-B_{71} = 0$	0.08829	0.00318	27.739	0
Hyp. 12 TL-GL	$(B_{11}+B_{21})-(B_{31}+B_{41}) = 0$	0.08008	0.00365	21.930	0
Hyp. 13 TL-CD	$(B_{11}+B_{21})-(B_{51}+B_{61}) = 0$	-0.05366	0.00162	-33.049	0
Hyp. 14 GL-CD	$(B_{31}+B_{41})-(B_{51}+B_{61}) = 0$	-0.13374	0.00399	-33.534	0
Hyp. 15 T1-T2	$\sum_{m=1}^7 B_{m4}/7 - \sum_{m=1}^7 B_{m5}/7 = 0$	0.06546	0.00689	9.494	0
Hyp. 16 T1-T3	$\sum_{m=1}^7 B_{m4}/7 - \sum_{m=1}^7 B_{m6}/7 = 0$	-0.10701	0.00689	-15.525	0
Hyp. 17 T2-T3	$\sum_{m=1}^7 B_{m5}/7 - \sum_{m=1}^7 B_{m6}/7 = 0$	-0.17246	0.00689	-25.022	0
Hyp. 18 EX-HN	$\sum_{m=1}^7 B_{m7}/7 - \sum_{m=1}^7 B_{m8}/7 = 0$	-0.03857	0.00689	-5.596	0
Hyp. 19 EX-TN	$\sum_{m=1}^7 B_{m7}/7 - \sum_{m=1}^7 B_{m9}/7 = 0$	-0.03725	0.00689	-5.404	0
Hyp. 20 HN-TN	$\sum_{m=1}^7 B_{m8}/7 - \sum_{m=1}^7 B_{m9}/7 = 0$	0.00132	0.00689	0.192	0.8476

References

- Bauer, P., 1990, 'Recent Developments in the Econometric Estimation of Frontiers,' *Journal of Econometrics*, v. 46, pp. 39-56.
- Charnes, A., W.W. Cooper, A.Y. Lewin, and L.M. Seiford, 1994, *Data Envelopment Analysis: Theory, Methodology, and Application*. Kluwer Academic Publishers, Boston.
- Coelli, T., 1995, 'Estimator and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis,' *Journal of Productivity Analysis*, v. 6, no. 3, pp. 247-268.
- Dixon, B.L., P. Garcia, and M. Anderson, 1987, 'Usefulness of Pretests for Estimating Underlying Technologies Using Dual Profit Functions,' *International Economic Review*, v. 28 no. 3, pp. 623-633.
- Driscoll, P.J., and R. Boisvert, 1991, 'Dual Second-and Third-Order Translog Models of Production,' *American Journal of Agricultural Economics*, v. 73 no. 4, pp. 1146-1160.
- Førsund, F., K. Lovell, and P. Schmidt, 1980, 'A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement,' *Journal of Econometrics*, v. 13, pp. 5-25.
- Gong, B., and R.C. Sickles, 1989, 'Finite Sample Evidence on the Performance of Stochastic Frontier Models Using Panel Data,' *Journal of Productivity Analysis*, v. 1, no. 3, pp. 229-261.
- Gong, B., and R.C. Sickles, 1992, 'Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data,' *Journal of Econometrics*, v. 51, pp. 259-284.
- Greene, W., 1995, 'LIMDEP computer program and manual: Version 7.0,' (Economic Software Inc., Bellport, New York).
- Greene, W., (H. Pesaran and P. Schmidt, eds.), 1997, *Handbook of Applied Econometrics: Microeconomics Vol 2*, Blackwell Publishing.
- Guilkey, D.K., K. Lovell, and R.C. Sickles, 1983, 'A Comparison of the Performance of Three Flexible Functional Forms,' *International Economics Review*, v. 24, no. 3, pp. 591-616.
- Hanoch, G., 1971, 'CRESH Production Functions,' *Econometrica*, v. 39, no. 5, pp. 695-712.
- Hjalmarsson, L., S.C. Kumbhakar, and A. Heshmati, 1996, 'DEA, DFA and SFA: A Comparison,' *Journal of Productivity Analysis*, v. 7, no. 2/3, pp. 303-327.
- Jondrow, J., K. Lovell, I.S. Materov, and P. Schmidt, 1982, 'On the Estimation of Technical Efficiency in the Stochastic Frontier Production Function Model,' *Journal of Econometrics*, v. 19, no. 2/3, pp. 233-238.

- Khoju, M., and B. Dixon, 1994, 'Sensitivity of Rankings of Inter-Firm Efficiency Measurement to Model Specification.' Paper Presented at the 1994 Southern Agricultural Economists Association Annual Meeting.
- Mood, A., F. Graybill, and D. Boes, 1974, *Introduction to the Theory of Statistics*. McGraw-Hill, New York, New York.
- Mukerji, V., 1963, 'Generalized SMAC Function with Constant Ratios of Elasticities of Substitution,' *Review of Economic Studies*, v. 30, pp. 233-236.
- Olsen, J.A., P. Schmidt, and D.M. Waldman, 1980, 'A Monte Carlo Study of Estimators of the Stochastic Frontier Production Function,' *Journal of Econometrics*, v. 13, pp. 67-82.