TE Estimation with Multiple Inputs and Multiple Outputs using Regression Analysis

Dr. S. Joseph Robin

Associate Professor, Department of Mathematics, Scott Christian College (Autonomous), Nagercoil – 629 003, Kanyakumari District, Tamil Nadu, India.

sjosephrobin@yahoo.com

***Abstract* – A huge amount of work has already taken place for the estimation of Technical Efficiency of farm/firm. In this regression analysis is used along with Data Envelopment Analysis to measure the Technical Efficiency. A linear multiple input-multiple output function is used. The properties of Isoquants in the production analysis are used in the estimation. An illustration with real data collected is given at the end.**

***Keywords* – Isoquants, Stochastic distance function, Data envelopment analysis.**

I INTRODUCTION

Measuring the performance relative to an estimator so far the regression and the L.P techniques have been used Farell's (1957) is the starting point of this type of work. It gives a conceptual frame work for the measurement of efficiency. He has shown that the technical and allocative inefficiency can be measured relative to the observed isoquants with equi-proportional measures. He has also illustrated efficiency using piecewise linear isoquants.

Farell's work is extended by Aigner and Chu(1968) by applying the programming model to measure the production in the deterministic models where all deviations from the frontier are one-sided due to inefficiency. Winsten(1957) and Greene (1980) have showed that the OLS can be used to estimate inefficiency in the case of one sided-deviations. This is done by adding the largest residual to the intercept in the production function and this is referred as the corrected OLS (COLS). COLS is limited due to the one output in the production function. Lovel *et al.* (1994) proposed a solution in the multiple output case by specifying a distance function, exploiting homogeneity and also rearranging the terms to specify the production process with one output used as the dependent variable treating all other outputs as independent variables. This method is called the stochastic distance function (SDF) approach and this is popularized by Grosskopf *et al.* (1997) and (Coelli and Perelman (1999, 2000). However this has an endogeneity problem as pointed out by Vinod (1969). However as pointed out by Coelli and Perelman (2000), a maintained advantage of SDF is the ability to estimate non-separable production. The advantage with regression-based approach is the goodness of fit and the other statistics to evaluate the overall model. Here we consider COLS to handle multiple out-puts. Unlike the SDF, this treats outputs symmetrically and it also satisfies proper curvature in output space.

Another limitation of COLS and its associates is the inability to properly account for measurement error. It is to be noted that deviations from the frontier can occur not only from inefficient behavior but also from measurement error and the statistical noise. A good majority of the related papers assumed that the deviation from the frontier consisted of an overall error composed of inefficiency and statistical noise. The panel data models are efficiently able to separate the effects of noise and inefficiency, creating a distinct advantage over the cross-sectional SFA model which cannot separate these components effectively or DEA which assumes no noise in the data. DEA easily handles multiple inputs and multiple outputs and allows direct comparisons of production possibilities without any additional input price data.

This study applies a DEA based method in the first stage to provide measure of aggregate output which is then incorporated into a second-stage regression. The regression-based approaches to measure efficiency in input and multiple output technologies is introduced.

II METHODOLOGY

Assume that each of the n DMU's employ a vector x of s inputs to produce a vector y of m outputs according to the technology

$$T = \{(x,y) : x \in R^S,\ y \in R^m.\ x \text{ can produce } y\}$$

Here we define the output set as $P(x) = \{y : (x,y) \in T\}$. The standard properties on $P(x)$ in Fare et. al (1994) are assumed. Following him now define the isoquant

$$\text{Isoq } P(x) = \{y \in P(x) : 0y \notin P(x) \text{ for } 0 > 1\}$$

This boundary is used to compare observed production possibilities to the boundary of the output set. DEA uses a piecewise linear approximation to the estimation of the output set (and the input set) Fare et. al (1994) prove that the piece wise linear technology P(X) is closed and bounded, sufficient conditions for the existence of the efficiency measure. Bonker *et al.* (1984) output-oriented DEA model to evaluate the Technical Efficiency (TE) of DMU "0" under the assumption of variable returns to scale (VRS) is given by

$F_o(x_o, y_o) = \max 0_o$

Such that $\sum_{j=1}^{n} \lambda_j y_{kj} \geq 0_o y_{ko} \;\; \forall \;\; k = 1. \ldots. s.$

$\sum \lambda_j \boldsymbol{x}_{lj} \leq \boldsymbol{x}_{\ell o \circ} \;\; \forall \, l=1,2\ldots,m \qquad (1)$

$\sum_{j=1}^{N} \lambda_j = 1,$

$\lambda_j \geq 0 \;\; \forall \;\; j$

The general set-up is shown in Fig.1 where two output sets are shown. $P(\boldsymbol{x}) \subset P(\boldsymbol{x}')$ with $\boldsymbol{x}' \geq \boldsymbol{x}$. Five observed production possibilities A, B, C, D and F are shown. It is assumed that $A,C,F \in P(\boldsymbol{x})$ but, $A,C,F \notin P(\boldsymbol{x}')$

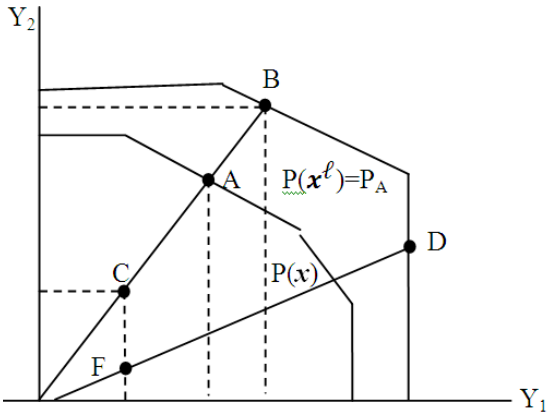

Fig. 1 Representation of Technology

Production possibilities C and F however are technically inefficient. Based on the definition of output-oriented efficiency and the solution of (1), we have

$F_o(\boldsymbol{x}_C, y_C) = y_{1A} \,|\, y_{1C}$ and $F_o(\boldsymbol{x}_A, y_A) = F_o(\boldsymbol{x}_B, y_B) = 1$

As mentioned in the beginning, the purported advantage of DEA over regression – based approaches has the ability to estimate the production technology characterized by multiple inputs and multiple outputs without relying on input prices.

Now we define the aggregate output set.

The aggregate output set P_A is defined as

$P_A = \bigcup_{j=1}^{N} P(\boldsymbol{x}_j)$

Since union of compact sets is compact, P_A is compact and it is closed and bounded, thus guarenteeing the existence of a distance function from an element in P_A to the boundary of P_A. The boundary is defined by the isoquant:

$\text{Isoq } P_A = \{ Y \in P_A \;:\; 0y \notin P_A \text{ for } 0 > 1\}$

We can use a piecewise linear approximation to generate Isoq P_A · Given the assumptions on each output set $P(\boldsymbol{x})$, the aggregate output set P_A can be thought of as the output set associated with the highest Isoq $P(\boldsymbol{x})$. The relevant properties on the production technology given in Fare et. al (1994) hold for aggregate output set P_A.

The LP model to measure the distance F_A for DMU "0" to the aggregate. Output set is given by.

$F_A(y_0) = \text{Max } \theta_0$

Such that $\sum_{j=1}^{N} A_j\, y_{kj} \geq \theta_0 y_{k0} \;\; \forall \;\; k = 1,2,\ldots s$

$\sum_{j=1}^{N} A_j = 1 \qquad (2)$

$A_j \geq 0 \;\; \forall \;\; j$

Now this model is similar to the output oriented DEA model assuming variable returns to scale with the exclusion of the input constraints. Model (2) produces an estimated output isoquant Isoq P_A. This model has been previously used to compare observations based on their multi-criteria output vector; in this case if the convexity constraint is included then the input constraint is redundant. Here (2) is used to aggregate outputs since the separability is assumed. The output aggregate proposed is the measure of output relative to the estimated isoquant Iso P_A.

Based on Fig.1, the solution to (2) leads to

$F_A(\boldsymbol{x}_B, y_B) = F_A(\boldsymbol{x}_D, y_D) = 1,$

$F_A(\boldsymbol{x}_C, y_C) = y_{1B} / y_{1C}$

$F_A(\boldsymbol{x}_A, y_A) = y_{1B} / y_{1A}$

and

$F_A(\boldsymbol{x}_F, y_F) = y_{1D} / y_{1F}$

Production unit F poses a special problem; the output constant for y_2 does not hold with equality: excess slack exists after radial projection leading to a shadow price of zero, a well known problem of DEA. Now Farrell measure adequately measures preference even in the presence of slack, if the underlying technology is everywhere substitutable. The measures can be decomposed into products of efficiency and distances between isoquant.

For example

$F_A(\boldsymbol{x}_C, y_C) = y_{1B} \,|\, y_{1C} = F_0(\boldsymbol{x}_C, y_C) \times F_A(\boldsymbol{x}_A, y_A)$

This distance function captures inefficiency (comparing C to A) and the distance between Frontiers (Comparing A to B).

Production units farther from the aggregate output set produce lower output aggregates; hence $S = F_A^{-1} = 1|F_A$ provides an index of aggregate observed output. This measure can be used in a second stage regression wherein aggregate production is regressed on observed inputs. This second stage approach, like all regression based models, requires a priori specification of the production function. A translog model can also be used for the flexibility.

III Estimation – via Regression

A multiple input, multiple output, production function is specified as

$h(y_i) = f(x_i) + \in_i, \;\; i = 1, 2\ldots, N \qquad (3)$

Here y_i is the output vector of the i^{th} firm, x_i is the i^{th} firm's input vector, f is an input aggregate function, h is an output function and $\in_i = v_i - u_i$ is the composite error which captures all deviations from the production frontier. (We assume that production function is seperable) v_i is a random disturbance term which includes the effects of omitted factors, measurement errors, and the stochastic noise. v is assumed to be a truncated normal variable with zero mean and F_v is a probability density function consistent with this specification. $u_i \geq o$ is the random inefficiency of the i^{th} firm. The existence of a well-behaved probability density function f_u with left-truncation at zero is assumed. u_i and v_i are assumed to be independently distributed random variables that are uncorrelated with the input variables x_i and with each other. Variables X a randomly sampled from the domain D_x. The Joint density function of the three random variables is denoted by $f_d(x, u, v)$.

Statement: A needed property of any estimator is the consistency. Now for (2) we show that it consistently estimates P_A.

The needed assumptions are

- the boundary of T is monotonic and concave function in X,
- the underlying production function $h(y_i) = f(x_i)$, should be separable,
- the sequence $\{(Y_i, X_i), i=1,2,\ldots n\}$ is a random sample of independent observations,
- the noise term v_i have a truncated distribution: $|v| \leq V^M$, $f_v(V^M) > 0$,
- the Joint density f_d satisfies $f_d(X,0,V^M) > 0\ \forall\ X \in Dx$,

In this case the estimator (2) is a consistent estimator for the boundary of P_A, in the following sense

$\lim_{n \to \infty}$ Isoq(X_i) = Isoq(P_A)+V^M for all i = 1,2…, n

Proof: By the first assumption isoquants are nested and the second gives output sets can be analyzed for a given aggregate input level. Now consider an arbitrary randomly drawn observation (y_i, x_i). Now for any input level x_i, there is a positive probability $p_i > 0$ of randomly drawing from the sample an observation k such that $f(X_k)=W$; $v_k=V^M$. For this observation $y_k = W + V^M$. Since the boundary of T is globally concave, it is not possible to achieve a better output level than y_k by using x_k. Thus, if an observation k characterized by the equations above is randomly drawn, then y_k is a member of the set Iso q (x_i). Otherwise, if the observation k is not drawn to sample, y_k is not a member of P_A. Consistency requires that the probability of drawing unit k approach unity as the sample size tends to infinity.

The probability that unit k is not observed in a sequence of n independent random draws is equal to $(1-p_i)^n$. Asymptotically this converges to zero. Thus, observation k is almost surely observed as the sample size tends to infinity. Hence

$\lim_{n \to \infty}$ Isoq(x_i) = Isoq(P_a) + V^M

Since x_i is arbitrary the same is true for all i=1, …n. This implies the consistently estimability of the nise component. Now the true isoquant can be recovered by subtracting V^M.

Now given that (2) is a consistent estimator, with a sufficiently large sample our measure of aggregate output $S = F_A^{-1}$ can be used in a subsequent regression. The first two examples are deterministic an assume no measurement error. For both examples, we adopted method in [2]. In particular technology is represented by a two-input, two-output transformation function with a Constant Elasticity of Transformation (CET) output aggregate and a Cobb-Douglas input aggregate. We used the OLS for model parameter estimation and COLS to estimate the technical efficiency. OLS in the first stage and the COLS in the second stage to get consistent estimator for the production function.

Example 1: After applying model(2) and estimating the aggregate output S, OLS is applied to estimate the productions. With the same data reported in Table 1, the efficient production function is given by h(y)=f(x), where

$$h(y) = (0.5\, y_1^2 + 0.5\, y_2^2)^{0.5}$$

and

$$f(x) = (x_1^{0.5}\ x_2^{0.5})^{\delta} \quad (4)$$

δ is used to account for the Variable Returns to Scale (VRS) $\in = 0 \Longrightarrow$ that the efficiency is one.

For f(x) and h(y) were assumed (4) hypothetical values with values for δ to be 1.000, 0.889 and 0.931. Using this the output aggregate values for S are estimated and presented in the same table 1. The correlation between S and h(y) is 0.989. Due to VRS, as suggested the translog equation was used and the OLS estimation of the S are also presented in Table 1 under $\hat{s}$. Here all the parameters are significant at one percent level of probability and the R-square is close to one. This is an indication for the use of regression for the multiple input, multiple output case.

IV Conclusion

The study reveals that the assumption of a proper truncated density function for the error along with the COLS estimation leads to better estimates for the parameters in the case of multiple inputs and multiple outputs for the firms/farms.

TABLE 1
DATA TAKEN FROM 20 FISHNET FIRMS USING TWO INPUTS AND TWO OUTPUTS

DMU	x_1	X_2	y_1	y_2	**h(y)**	**f(x)**	**δ**	**S**	$\hat{s}$
1	28.80	28.80	20.00	20.00	25	35	0.889	0.25	0.26
2	23.04	36.00	24.00	14.97	25	35	0.889	0.25	0.25
3	17.28	48.00	16.00	23.32	25	35	0.889	0.26	0.25
4	16.00	52.00	12.00	25.62	25	35	0.889	0.26	0.26
5	34.56	24.00	9.60	26.61	25	35	0.889	0.25	0.25
6	40.00	40.00	40.00	4.00	50	50	1.000	0.50	0.50
7	36.00	44.45	32.00	46.65	50	50	1.000	0.51	0.50
8	48.00	33.34	48.00	29.94	50	50	1.000	0.50	0.51
9	24.00	66.66	24.00	51.22	50	50	1.000	0.52	0.51
10	56.00	28.57	16.00	54.26	50	50	1.000	0.51	0.52
11	60.00	60.00	60.00	60.00	75	75	1.000	0.74	0.73
12	36.00	100.00	24.00	81.38	75	75	1.000	0.76	0.75
13	48.00	75.00	40.00	74.83	75	75	1.000	0.75	0.76
14	84.00	42.86	64.00	55.71	75	75	1.000	0.77	0.77
15	68.00	52.94	72.00	44.90	75	75	1.000	0.75	0.76
16	115.20	115.20	80.00	80.00	100	145	0.931	1.000	1.000
17	92.16	144.00	92.00	65.85	100	145	0.931	1.000	1.001
18	69.12	192.00	64.00	93.30	100	145	0.931	1.000	1.001
19	138.24	96.00	52.00	100.48	100	145	0.931	1.000	1.000
20	161.27	82.29	48.00	102.45	100	145	0.931	1.000	1.000
Mean	59.78	68.10	41.89	54.58	62.5	76.25	0.955	0.63	0.63
S.D	40.53	43.31	24.25	27.32	31.43	41.93	0.06	0.31	0.31
Min	16.00	24.00	9.60	14.97	25.00	35.00	0.89	0.25	0.25
Max	161.27	192.00	92.00	102.45	100.00	145.00	1.000	1.00	1.01

REFERENCES

[1] Aigner, D; Chu, S.F, (1968). On estimating the industry production function. American Economic Review 58, 826-839.

[2] Coelli, T.Perelman, S.(1999). A comparison of parametric and non-parametric distance functions: with application to European railways, European Journal of operational Research 117, 326-339.

[3] Coelli, T. Perelman S, (2000). Technical efficiency of European railways: A distance function approach. Applied Economics 32, 1967-1976.

[4] Coreene, W; (1980) Maximum Likelihood estimation of econometric frontier productions. Journal of Econometrics, 13, 27-56.

[5] Frell, M.J, (1957). The measurement of productive efficiency. Journal of Royal Statistical society series A.120, 253-281.

[6] Gross Kopf. S, Hayes, K.J., Taylor, L.L.; Weber, W.L., (1997) Budget – Constrained frontier measures of fiscal equality and efficiency in schooling. Review of Economics and statistics 79 (1) 116-124 Vinod, H.D., (1969). Econometrics of Joint production.

[7] Kumbhakar, S. Lovell, C.A.K. 2000. Stochastic Frontier Analysis. Cambridge university Press, New York, N.Y.

[8] Lovell, C.A.K. Richardson, S. Travers, P.Wood, L.L. (1994). Resources and functionings: A new view of inequality in Australia. In: Eichon, W.(Ed), Models and Measurement of Welfare and Inequality. Springler –Verlag, Berlin, p.787-807.

[9] Winston, C.B. (1957) . Discussion Mr. Farrell's Paper. Journal of Royal statistical Society Series A: Statistics in Society 120 (B) ; 282-284.

