October 2005

# Input Aggregation in Models of Data Envelopment Analysis: A Statistical Test with an Application to Indian Manufacturing

Subhash C. Ray
*University of Connecticut*

University of Connecticut

*Department of Economics Working Paper Series*

# Input Aggregation in Models of Data Envelopment Analysis: A Statistical Test with an Application to Indian Manufacturing

Subhash C. Ray
University of Connecticut

Working Paper 2005-54

October 2005

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

## Abstract

A problem frequently encountered in Data Envelopment Analysis (DEA) is that the total number of inputs and outputs included tend to be too many relative to the sample size. One way to counter this problem is to combine several inputs (or outputs) into (meaningful) aggregate variables reducing thereby the dimension of the input (or output) vector. A direct effect of input aggregation is to reduce the number of constraints. This, in its turn, alters the optimal value of the objective function. In this paper, we show how a statistical test proposed by Banker (1993) may be applied to test the validity of a specific way of aggregating several inputs. An empirical application using data from Indian manufacturing for the year 2002-03 is included as an example of the proposed test.

**Journal of Economic Literature Classification:** C61, C43

**Keywords:** Efficiency distribution, F Tests

The paper was written while the author was visiting the Indian Statistical Institute, Calcutta.

**INPUT AGGREGATION IN MODELS OF DATA ENVELOPMENT ANALYSIS: A STATISTICAL TEST WITH AN APPLICATION TO INDIAN MANUFACTURING**

The numbers of identifiably different inputs and outputs typically involved in a production process are quite large. In the interest of tractability, one has to aggregate various individual inputs (or outputs) into a smaller number of composite inputs (or outputs) that can be manageably incorporated in an appropriate specification of the technology through a production, cost, or profit function. In agriculture, for example, various kinds of equipment (like tractors, harvesters, grain elevators, etc.) are suitably combined into a single input called *machinery.*

In the context of an econometric model of a production function, such input aggregation amounts to imposing some prior restrictions on the coefficients of a regression model and appropriate tests (like the *F* test or a likelihood ratio test) may be employed to check the validity of such restrictions. The problem is different in the context of Data Envelopment Analysis (DEA) which is based on mathematical programming where the number of different inputs and outputs included in the model define the number of constraints in the relevant programming problem. There a direct effect of input aggregation is to reduce the number of constraints. This, in its turn, alters the optimal value of the objective function. In this paper, we show how a statistical test proposed by Banker (1993) may be applied to test the validity of aggregating the inputs.
The rest of the paper is organized as follows. In section 2, we present the basic DEA methodology for efficiency evaluation and use the dual or multiplier form of the relevant linear programming (LP) model to show the similarity between input aggregation in DEA and parameter restriction in regression models. Section 3 offers a brief description of Banker's interpretation of the DEA efficiency score as a maximum likelihood estimator and the "F" tests developed by him. Section 4 includes an empirical application of the test procedure in this paper using data from Indian manufacturing. Section 5 is the conclusion.

## 2. The DEA Methodology

In parametric models, one specifies an explicit functional form for the frontier and econometrically estimates the parameters using sample data for inputs and output. Hence the validity of the derived technical efficiency measures depends critically on the appropriateness of the functional form specified.

In contrast, the method of DEA introduced by Charnes, Cooper and Rhodes (CCR) (1978) and further generalized by Banker, Charnes, and Cooper (BCC) (1984) provides a nonparametric alternative to parametric frontier production function analysis. In DEA, one makes only a few fairly weak assumptions about the underlying production technology. In particular, no functional specification is necessary. Based on these assumptions a production frontier is empirically constructed using mathematical programming methods from observed input-output data of sample firms. Efficiency of firms is then measured in terms of how far they are from the frontier.

Consider an industry producing a scalar output, $y$, from a bundle of m inputs, $x=(x_1,x_2,\ldots,x_m)$. Let $(x^j, y^j)$ be the observed input-output bundle of firm $j$ $(j= 1,2,\ldots, N)$. The technology is defined by the production possibility set

$$T=\{(x, y): y \text{ can be produced from } x\}. \qquad (1)$$

An input-output combination $(x^0, y^0)$ is feasible if and only if $(x^0, y^0) \in T$.

We make the following assumptions about the technology:

- All observed input-output combinations are feasible. Thus, $(x^j, y^j) \in T$ $(j = 1,2,\ldots,N)$.
- The production possibility set, $T$, is convex. Hence, if $(x^1, y^1) \in T$ and $(x^2, y^2) \in T$, then $(\lambda x^1+(1-\lambda)x^2, \lambda y^1+(1-\lambda)y^2) \in T$, $0 \le \lambda \le 1$.

  In other words, weighted averages of feasible input-output combinations are also feasible.
- Inputs are freely disposable. Hence, if $(x^0, y^0) \in T$ and $x^1 \geq x^0$, then $(x^1, y^0) \in T$. This rules out negative marginal productivity of inputs.
- Output is freely disposable. Hence, if $(x^0, y^0) \in T$ and $y^1 \leq y^0$, then $(x^0, y^1) \in T$

  Varian (1984) pointed out that the smallest set satisfying the above assumptions is;

$$S=\left\{(x,y): x \geq \sum_{1}^{N} \lambda_j x^j; y \leq \sum_{i}^{N} \lambda_j y^j; \sum_{1}^{N} \lambda_j = 1; \lambda_j \geq 0; j=1,2,\ldots,,N\right\}. \qquad (2)$$

Let $\bar{x}=\sum_{1}^{N} \lambda_j x^j; \bar{y}=\sum_{1}^{N} \lambda_j y^j; \sum_{1}^{N} \lambda_j = 1; \lambda_j \geq 0.$ By virtue of convexity, $(\bar{x}, \bar{y})$ is feasible. Then, for any $x \geq \bar{x}, (x, \bar{y})$ is feasible. Finally, for any $y \leq \bar{y}, (x, y)$ is also feasible. If we assume constant returns to scale (CRS), for any $(x,y) \in T, (kx, ky) \in T$ for any $k \geq 0$. In that case, the $\lambda_j$ s will be only restricted to be non-negative and would not have to add up to unity. The CRS production possibility set would then be

$$S^C=\left\{(x,y): x \geq \sum_{1}^{N} \lambda_j x^j; y \leq \sum_{i}^{N} \lambda_j y^j; \lambda_j \geq 0; j=1,2,\ldots,,N\right\}. \qquad (3)$$

Under the CRS assumption, the output-oriented technical efficiency of any firm producing output $y^0$ from input $x^0$ is $1/\phi^*$ , where

$$\phi^* = \max\ \phi : (x^0, \phi y^0) \in S^C.$$

To compute technical efficiency[1] one solves the following linear programming problem:

$$\phi_k = max \quad \phi \qquad\qquad k \in (1,\dots,N)$$

$$\begin{aligned} \text{s.t} \quad & \sum_i^N \lambda_j y^j \geq \varphi y^k; \\ & \sum_i^N \lambda_j x^j \leq x^k; \qquad\qquad (4) \\ & \lambda_j \geq 0; (j = 1,2,\dots,N). \end{aligned}$$

The dual of the LP problem (4) is

$$\begin{aligned} & \min w' x^k \\ \text{s.t.} \quad & w' x^j - p' y^j \geq 0; \qquad\qquad (5) \\ & w' x^k = 1; \\ & w \geq 0; p \geq 0. \end{aligned}$$

For a simple example, consider the 3-input 1-output case. Thus, the input-output bundle of firm *j* is $(x_{1j}, x_{2j}, x_{3j}; y_j)$ $(j=1,2,\dots,N)$. For this example the explicit form of problem (4) above is

$$\begin{aligned} & \varphi_k = \max \varphi \\ \text{s.t.} \quad & \sum_i^N \lambda_j y_j \geq \varphi y_k; \\ & \sum_i^N \lambda_j x_{1j} \leq x_{1k}; \\ & \sum_i^N \lambda_j x_{2j} \leq x_{2k}; \\ & \sum_i^N \lambda_j x_{3j} \leq x_{3k}; \qquad\qquad (6) \\ & \lambda_j \geq 0; (j = 1,2,\dots,N). \end{aligned}$$

[1] Under constant returns to scale the output- and input-oriented technical efficiency measures coincide.

The corresponding dual problem is

$$\min w_1 x_{1k} + w_2 x_{2k} + w_3 x_{3k}$$

s.t. $\quad w_1 x_{1j} + w_2 x_{2j} + w_3 x_{3j} - p y_j \geq 0; \qquad (j=1,2,\ldots,k,\ldots,N); \qquad (7)$

$$p y_k = 1;$$

$$w_1, w_2, w_3, p \geq 0.$$

Now impose an additional constraint $aw_1 - bw_2 = 0$. That is, $w_2 = \frac{a}{b} w_1$.

The restricted version of (6) would then be

$$\min w_1 (x_{1k} + \tfrac{a}{b} x_{2k}) + w_3 x_{3k}$$

s.t. $\quad w_1 (x_{1j} + \frac{a}{b} x_{2j}) + w_3 x_{3j} - p y_j \geq 0; \qquad (j=1,2,\ldots,k,\ldots,N); \qquad (7a)$

$$p y_k = 1;$$

$$w_1, w_3, p \geq 0.$$

Define, now, the aggregated input

$$X_{1j} = x_{1j} + \tfrac{a}{b} x_{2j}.$$

Problem (7a) would then become

$$\min w_1 X_{1k} + w_3 x_{3k}$$

s.t. $\quad w_1 X_{1j} + w_3 x_{3j} - p y_j \geq 0; \qquad (j=1,2,\ldots,k,\ldots,N); \qquad (7b)$

$$p y_k = 1;$$

$$w_1, w_3, p \geq 0.$$

The dual of this problem is

$$\hat{\varphi}_k = \max \varphi$$

$$\text{s.t.} \sum_i^N \lambda_j y_j \geq \varphi y_k;$$

$$\sum_i^N \lambda_j X_{1j} \leq X_{1k};$$

$$\sum_i^N \lambda_j x_{3j} \leq x_{3k}; \qquad (8)$$

$$\lambda_j \geq 0; (j = 1,2,\ldots,N).$$

Obviously, when $a = b$, $X_1$ is simply the sum of the quantities of the inputs $x_1$ and $x_2$. In that case, the two inputs are treated as perfect substitutes. Further, because (7b) is a restricted version of (7) the minimum value of the objective function at the optimal solution of (7b) will be no lower than what is obtained at the optimal solution of (6). Therefore, by standard duality results, $\varphi^k \leq \hat{\varphi}_k$ for every $k$. The test of validity of the aggregation amounts to a decision as to whether the distributions of efficiency with and without the restriction are significantly different[2]. In regression models, a commonly used test of significance compares the restricted and unrestricted residual sums of squares. We now consider a comparable F test developed by Banker in the context of DEA.

### 3. DEA as Maximum Likelihood Estimation and Banker's F Test

We start with $N$ observed input-output bundles. The pair $(x^j, y_j)$ represents the input bundle $x^j$ used by firm $j$ to produce the scalar output $y_j$. Next, following Banker (1993), consider the production function mapping from the n-element input bundle $x^0 \in X \subseteq R^n_+$ onto the non-negative scalar output $y_0$:

$$y_0 = g(x^0). \qquad (9)$$

We assume that the production function satisfies the following postulates:.

(P1) $g(x)$ is monotonic in $x$. That is if $x^{1'} \geq x^2$, then $g(x^1) \geq g(x^2)$.

(P2) $g(x)$ is concave. Hence, if $x^1, x^2 \in X$ and $x^* = \lambda x^1 + (1-\lambda)x^2$, $0 < \lambda < 1$, then

$g(x^*) \geq \lambda\, g(x^1) + (1-\lambda)\, g(x^2)$.

(P3) For each observation $(x^j, y_j)$, $g(x^j) \geq y_j$ ; $(j = 1,2,\ldots, N)$.

(P4) For *any* other function $\tilde{g}(x)$ also satisfying (P1-P3), $\tilde{g}(x) \geq g(x)$ for all $x \in X$.

Now consider the set $X^* = \{x : x \geq \sum_{j=1}^{N} \lambda_J x^j ; \sum_{j=1}^{N} \lambda_J = 1; \lambda_j \geq 0\} \subseteq X$. Clearly, $X^*$ is the free disposal convex hull of the observed input bundles. Banker has shown that the unique function $y = g(x)$ determined for $x \in X^*$ by the postulates (P1-P4) corresponds to that estimated by DEA.

---

[2] Pastor, Ruiz, and Sirvent (1995) performed a nonparametric statistical test of nested radial DEA models to determine the optimal choice of inputs and outputs.

We first note that if the function $y = \hat{g}(x)$ satisfies properties (P1-P4) and if $\hat{y}_0 = \hat{g}(x^0)$ for $x^0 \in X^*$, then $\hat{y}_0 = g^*(x^0)$, where

$$g^*(x^0) = y_0^* = \max \sum_{j=1}^{N} \lambda_j y_j$$

$$\text{s.t.} \quad \sum_{j=1}^{N} \lambda_j x^j \leq x^0; \tag{10}$$

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

$$\lambda_j \geq 0.$$

It is easy to see that $g^*(\,.\,)$ satisfies (P1-P3). First, consider the input bundle $\tilde{x} \geq x^0$. Obviously, the optimal solution for the DEA problem for $x^0$ is a feasible solution of the DEA problem for $\tilde{x}$. Thus, clearly, $g^*(\tilde{x}) \geq y_0^* = g^*(x^0)$. Next we show that $g^*(x)$ is concave. Suppose that $\lambda' = (\lambda_1', \lambda_2', ..., \lambda_N')$ and $g^*(x')$ is the optimal solution of the DEA LP problem for the input bundle $x' \in X^*$. Similarly, $\lambda'' = (\lambda_1'', \lambda_2'', ..., \lambda_N'')$ and $g^*(x'')$ is the optimal solution for $x'' \in X^*$. For any arbitrary $\theta \in [0,1]$ define $\bar{\lambda} = \theta\lambda' + (1-\theta)\lambda''$ and $\bar{x} = \theta x' + (1-\theta)x''$. Clearly, $\bar{\lambda}$ is a feasible solution for the DEA LP for $\bar{x}$ leading to the objective function value $\theta g^*(x') + (1-\theta)g^*(x'')$. Obviously, the optimal solution $g^*(\bar{x})$ satisfies $g^*(\bar{x}) \geq \theta g^*(x') + (1-\theta)g^*(x'')$. This verifies that $g^*(x)$ is a concave function.

Let $y_0^* = g^*(x^0) = \sum_{j=1}^{N} \lambda_j^* y_j$ be the optimal solution of the DEA LP for $x^0$. Next suppose that some other function $\hat{g}(x)$ satisfies the postulates (P1-P3). Then,

$$\hat{g}\left(\sum_{j=1}^{N} \lambda_j^* x^j\right) \geq \sum_{j=1}^{N} \lambda_j^* \hat{g}(x^j) \geq \sum_{j=1}^{N} \lambda_j^* y_j = g^*(x^0).$$

Further, because $x^0 \geq \sum_{j=1}^{N} \lambda_j^* x^j$, $\hat{g}(x^0) \geq \hat{g}\left(\sum_{j=1}^{N} \lambda_j^* x^j\right) \geq g^*(x^0)$.

Thus, the function $g^*(x) \le \tilde{g}(x)$ for any function $\tilde{g}(x)$ satisfying (P1-P3) for any function $\tilde{g}(x)$ satisfying (P1-P3 over the set $X^*$. An implication of this is that the deviation $\varepsilon_j = \tilde{g}(x^j) - y_j$ is minimized for each observation $j$ by the function $g^*(x)$.

Now consider the frontier production function

$$y = g(x) - \varepsilon; \varepsilon \ge 0. \qquad (11)$$

Here, the non-negative deviation of the observed output $y$ from the frontier $g(x)$ has some one-sided probability distribution $f(\varepsilon)$. Then the likelihood maximization problem can be specified as:

$$\underset{f(.),g(.)}{\text{maximize}}\ L = \prod_{j=1}^{N} f\left(\varepsilon_j = g(x^j) - y_i\right)$$

$$\text{subject to} \quad g(x^j) - y_j \ge 0; \qquad (12)$$

$g(.)$ is a monotone increasing and concave function.

It may be noted that the DEA efficiency residuals $\varepsilon_j$ are obtained independently of each other. This is in contrast with the frontier production function model proposed by Aigner and Chu (1968). In their case, a single parametric function is fitted to the entire data set and the efficiency residuals are jointly derived and, therefore, are not independent of one another. Now suppose that we choose a probability density function $f(\,.\,)$ such that $f\,\varepsilon_j)$ is monotone decreasing in the efficiency residuals. In that case, because the DEA estimate of the production function minimizes each $\varepsilon_j$, it thereby maximizes each $f(\varepsilon_j)$. Hence, the DEA frontier $g^*(x)$ maximizes the likelihood function subject to the constraints specified above.

Banker specifies the deterministic frontier where the random inefficiency component of $y$ appears in an additive manner. One may directly link the one-sided econometric frontier with the DEA frontier by specifying (11) differently as

$$y = g(x)e^{-\varepsilon}; \varepsilon \ge 0 \qquad (11a)$$

leading to

$$g(x) = \varphi y; \varphi \geq 1. \qquad (11b)$$

Thus,

$$\varepsilon = \ln(\varphi) \geq 0. \qquad (11c)$$

Note that all the preceding arguments about the DEA frontier $g^*(x)$ as a maximum likelihood estimator of the unknown frontier $g(x)$ remains valid.

Banker has proposed a number of statistical tests for comparing two groups of firms to assess whether one group is more efficient than the other. Assume that there are $N$ firms in the sample of which $m_1$ are in group 1 and $m_2$ are in group 2. Firms in group1 have the exponential distribution of (in)efficiency $\varepsilon_j$ with parameter $\sigma_1$ and those in group 2 also have the exponential distribution but with parameter $\sigma_2$ . Designate the first group of firms as $M_1$ and the second group as $M_2$. Consider the residuals $\varepsilon_j^*(j = 1,2,..,N)$ obtained from DEA. Under the maintained hypothesis, the sample statistic

$\sum_{j \in M_i} \frac{\varepsilon_j^*}{\sigma_i}$ has the $\chi^2$ distribution with $2m_i$ $(i = 1, 2)$ degrees of freedom.

Under the null hypothesis $\sigma_1 = \sigma_2$, the test statistic

$$F = \frac{\sum_{j \in M_1} \varepsilon_j^* \Big/ m_1}{\sum_{j \in M_2} \varepsilon_j^* \Big/ m_2} \qquad (13)$$

has the F distribution with $(2m_1, 2m_2)$ degrees of freedom.

On the other hand, if the $\varepsilon_j$ s have the half Normal distribution, (i.e., the Normal distribution with mean 0 and variance $\sigma^2$ truncated from below at 0), then $\sum_{j \in M_1} \left( \frac{\varepsilon_j^*}{\sigma_1} \right)^2$ has the $\chi^2$ distribution with $m_1$ degrees of freedom. Similarly, $\sum_{j \in M_2} \left( \frac{\varepsilon_j^*}{\sigma_2} \right)^2$ has the $\chi^2$ distribution with $m_2$ degrees of freedom. Hence, in this case, under the null hypothesis $\sigma_1 = \sigma_2$,, the statistic

$$F = \frac{\sum_{j \in M_1} (\varepsilon_j^*)^2 \Big/ m_1}{\sum_{j \in M_2} (\varepsilon_j^*)^2 \Big/ m_2} \qquad (14)$$

has the F distribution with *($m_1$, $m_2$)* degrees of freedom.

One would use $\varepsilon_k^* \equiv \ln(\hat{\varphi}_k)$ for the aggregated (i.e., the restricted) model and $\varepsilon_k^* \equiv \ln(\varphi_k)$ for the disaggregated (i.e., the unrestricted) model in an empirical application. For appropriate distributional assumption the aggregated model would be rejected only when the test statistic exceeds the critical value for the relevant degrees of freedom.

## 4. Application to Indian Manufacturing

In the empirical application we use state level aggregate data on output and inputs in total manufacturing for the different states (and union territories) of India from the Annual Survey of Industries (ASI) for the year 2002-03. A single output *(y)* measured by the value of production at current prices is considered. Because it is a single cross section data set and state level indexes of output price are not available, the quantity of output is treated as proportional to its value. Inputs considered were (a) production workers *($L_1$)*, non-production workers *($L_2$)*, fixed capital *(K)*, fuels *(F),* and materials *(M).* The two labor inputs are measured in numbers of persons employed. All other inputs are expressed in value terms. The data used are reported in Table 1.

Two different DEA models were considered. In one the two labor inputs are treated separately. In the other they are combined into a single labor input *(L)*. The optimal DEA objective function values and the corresponding (inefficiency) residuals are reported in Table 2. The columns labeled $\varphi_k$ and $\varepsilon_k$ relate to the disaggregated model where the two kinds of labor are treated as two distinct inputs. Similarly, $\hat{\varphi}_k$ and $\hat{\varepsilon}_k$ relate to the model where the total employment is treated as a single input. As expected, for many states, DEA inefficiency residuals are larger in the restricted model with aggregated labor. The summary statistics relevant for the F tests are reported in Table 3. Under the assumption that ε has an exponential distribution, the test statistic is

$$F = \frac{1.7834756}{0.9749294} = 1.835\,.$$

This exceeds the critical value (1.610) of the F distribution with 46 degrees of freedom for both the numerator and the denominator at the 5% level of significance in a 1-tailed test. Thus the model using total labor as one input is rejected and a disaggregated model with production and non-production workers treated as distinct inputs is chosen.

For the alternative assumption that ε has a half Normal distribution, the test statistic is

$$F = \frac{0.2740236}{0.1005418} = 2.725.$$

The critical value of $F_{23,23}$ at the 5% significance level is 2.01 for a 1-tailed test. Thus, the aggregated model is rejected under the half-Normal distributional assumption as well.

In this application, we treated production and non-production workers as perfect substitutes and measured the aggregate labor input by total employment. In light of skill differences in the two kinds of labor, such simple aggregation is questionable and some kind of differential weighting is called for. We tried an alternative aggregation procedure using the relative wage rates of production and non-production workers as weights. Using all-India data, we found that the annual earnings of a non-production worker was 2.9787 times the earnings of a production worker. We, therefore, counted one non-production worker as equivalent to 2.9787 production workers and created total labor in production worker equivalents accordingly. This time the respective F statistics for the exponential and the half-Normal assumptions were 1.6283 and 2.26. Thus, even when we use salary-weighted employment as a single labor input, the aggregated model is clearly rejected for the half-Normal distribution. Under the exponential distributional assumption, though, the *p*-value barely exceeds 5% and the aggregated model is marginally rejected.

## 5. Conclusion

A problem frequently encountered in DEA is that the total number of inputs and outputs included tend to be too many relative to the sample size. Because each additional input or output included in the analysis imposes another constraint, the set of feasible solutions tends to become smaller and more and more firms tend to lie on or close to the frontier. One way to counter this problem is to combine several inputs (or outputs) into (meaningful) aggregate variables reducing thereby the dimension of the input (or output) vector. In this paper we show how an F test developed by Banker may be employed under appropriate distributional assumptions of the efficiency component to test the validity of any specific aggregation procedure. The empirical application using data from Indian manufacturing suggests that using total employment as a single labor input is not a valid aggregation of production and non-production workers.

**Table1. ASI State-level Input-output Data in Manufacturing 2002-03**

| State | Y | $L_1$ | L | K | F | M |
|---|---|---|---|---|---|---|
| JK | 175631 | 19636 | 24881 | 37863 | 8361 | 108479 |
| HP | 611899 | 25375 | 34023 | 358278 | 52533 | 323166 |
| PU | 4057079 | 276677 | 351102 | 1119761 | 322346 | 2614087 |
| CH | 114081 | 5180 | 8243 | 28005 | 5072 | 61798 |
| UT | 603559 | 27815 | 41485 | 204586 | 49348 | 342505 |
| HA | 5261740 | 223831 | 299765 | 1410886 | 209354 | 3381965 |
| DE | 1724437 | 85693 | 127935 | 249193 | 44935 | 979492 |
| RA | 3311696 | 190971 | 244265 | 1355028 | 362936 | 1868534 |
| UP | 8052063 | 409116 | 542160 | 2847631 | 475443 | 4930123 |
| BI | 807680 | 44280 | 54184 | 317048 | 23045 | 551546 |
| AS | 1250610 | 93129 | 110879 | 596413 | 64932 | 766956 |
| WB | 4873244 | 428096 | 538858 | 2458806 | 320238 | 3041783 |
| JH | 2476429 | 121427 | 156497 | 1576865 | 264896 | 1082276 |
| OR | 1486235 | 92686 | 118187 | 1061308 | 210580 | 747136 |
| CT | 1449707 | 63771 | 93794 | 728551 | 209379 | 669932 |
| MP | 3805750 | 156565 | 208874 | 1389786 | 304492 | 2228977 |
| GU | 18269979 | 528217 | 717055 | 8235801 | 981847 | 12488798 |
| MH | 21759551 | 829305 | 1170461 | 7697089 | 970795 | 13006687 |
| AP | 8228642 | 864822 | 1007463 | 3206186 | 432844 | 5466125 |
| KA | 6568082 | 370217 | 485917 | 3307273 | 349655 | 3664688 |
| GO | 895740 | 24318 | 35061 | 342420 | 43042 | 544433 |
| KE | 2665085 | 227347 | 270548 | 679483 | 132151 | 1885033 |
| TN | 10807543 | 920127 | 1125497 | 4347893 | 708413 | 6697103 |

Notes:
(i) Variable Definition: Y = output; $L_1$ = production workers;
L= total labor; K = fixed capital; F = fuel; M = materials.
Labor inputs are measured by number of persons employed;
all other inputs are in *lakhs* of Rupees at current prices ( 1 *lakh*=0.1 million).
(ii) State Names:
JK: Jammu & Kashmir; HP: Himachal Pradesh; PU: Punjub; CH: Chandigarh;
UT: Uttaranchal; HA: Haryana; DE: Delhi; RA: Rajasthan; UP: Uttar Pradesh;
BI: Bihar; AS: Assam; WB: West Bengal; JH: Jharkhand; OR: Orissa;
CT: Chhattisgarh; MP: Madhya Pradesh; GU: Gujarat; MH: Maharashthra;
AP: Andhra Pradesh; KA: Karnataka; GO: Goa; KE: Kerala; TN: Tamilnadu.

**Table 2. DEA Results from Disaggregated and Aggregated Models**

| State | $\varphi_k$ | $\hat{\varphi}_k$ | $\varepsilon_k$ | $\hat{\varepsilon}_k$ |
|---|---|---|---|---|
| JK | 1.11571 | 1.11571 | 0.10949 | 0.10949 |
| HP | 1.04076 | 1.04076 | 0.03995 | 0.03995 |
| PU | 1.06577 | 1.18864 | 0.0637 | 0.17281 |
| CH | 1.00146 | 1.00146 | 0.00146 | 0.00146 |
| UT | 1.06137 | 1.06137 | 0.05956 | 0.05956 |
| HA | 1 | 1.01053 | 0 | 0.01047 |
| DE | 1 | 1 | 0 | 0 |
| RA | 1.03272 | 1.10962 | 0.03219 | 0.10402 |
| UP | 1.05402 | 1.10688 | 0.05261 | 0.10154 |
| BI | 1 | 1 | 0 | 0 |
| AS | 1.03953 | 1.14278 | 0.03877 | 0.13347 |
| WB | 1.17665 | 1.19726 | 0.16267 | 0.18004 |
| JH | 1 | 1 | 0 | 0 |
| OR | 1.1441 | 1.1441 | 0.13462 | 0.13462 |
| CT | 1 | 1 | 0 | 0 |
| MP | 1 | 1.03452 | 0 | 0.03393 |
| GU | 1 | 1.0027 | 0 | 0.0027 |
| MH | 1.01165 | 1.02236 | 0.01158 | 0.02212 |
| AP | 1.13296 | 1.2282 | 0.12484 | 0.20555 |
| KA | 1.00898 | 1.05314 | 0.00894 | 0.05178 |
| GO | 1 | 1 | 0 | 0 |
| KE | 1.01945 | 1.27955 | 0.01927 | 0.24651 |
| TN | 1.12219 | 1.18943 | 0.11528 | 0.17348 |

State Names:
JK: Jammu & Kashmir; HP: Himachal Pradesh; PU: Punjub; CH: Chandigarh; UT: Uttaranchal; HA: Haryana; DE: Delhi; RA: Rajasthan; UP: Uttar Pradesh; BI: Bihar; AS: Assam; WB: West Bengal; JH: Jharkhand; OR: Orissa; CT: Chhattisgarh; MP: Madhya Pradesh; GU: Gujarat; MH: Maharashthra; AP: Andhra Pradesh; KA: Karnataka; GO: Goa; KE: Kerala; TN: Tamilnadu.

**Table3. Summary Statistics of DEA Residuals from Alternative Models**

| Distributional Assumption | $\sum \varepsilon_k$ | $\sum \hat{\varepsilon}_k$ | $\sum \varepsilon_k^2$ | $\sum \hat{\varepsilon}_k^2$ |
|---|---|---|---|---|
| | | | | |
| Exponential | 0.9749294 | 1.7834756 | | |
| Half-Normal | | | 0.1005418 | 0.2740236 |

## References


Aigner, D.J. and S. F.Chu (1968), “On Estimating the Industry Production Function,” *American Economic Review* 58:4 (September) 826-39.

Banker, R. D. (1993), “Maximum Likelihood, Consistency, and Data Envelopment Analysis: A Statistical Foundation”, *Management Science,* 39, 1265-1273.

Banker, R.D., A. Charnes, and W.W. Cooper (1984), “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis,” *Management Science,* 30:9 (September), 1078-92.

Charnes, A., W.W. Cooper, and E. Rhodes (1978) “Measuring the Efficiency of Decision Making Units,” *European Journal of Operational Research* 2:6 (November), 429-44.

Government of India (2004) *Annual Survey of Industries 2002-03.*

Pastor, J.T., J.J. Ruiz, and I. Sirvent (1995) “A Statistical Test for Nested Radial DEA Models”; *Working Paper,* Departmento de Estadistical e Investigacion Operativa; Universidad de Alicante, Spain.

Varian, H. R.(1984), “The Nonparametric Approach to Production Analysis,” *Econometrica* 52:3 (May) 579-97.