DATA ENVELOPMENT ANALYSIS (DEA)

Raymund R. Abara*

Abstract

This paper will introduce to readers an alternative approach to efficiency analysis called Data Envelopment Analysis (DEA). Because it is an advanced evaluation tool that incorporates both tangible and intangible attributes, DEA is now gaining popularity among economists, politicians, and all types of Decision-Makers (DM).

After a brief discussion of the model, this paper will demonstrate via examples how DEA can be used in aiding Decision-Makers evaluate their constituents.

I. Introduction

The top management of a company, which I will call ACME Inc, is conducting its annual company check-up. ACME Inc. operates 6 factories in six regions throughout the country. Moreover, since these factories are geographically disbursed, each General Manager has his own 'fiefdom'- operating his factory with total autonomy.

Manufacturing consumer goods is the primary business of ACME Inc., and its production volume has varied from factory to factory. Of course management does not expect the performance of each factory to be identical. With different factory sizes, capital machinery and employee make-up, it would be unreasonable for management to require that the factories churn the same production levels.

Despite these differences in production levels, top management is not alarmed. As a matter of fact, management knows that this diversity has been one of the factors of ACME's success. For every region where ACME has a factory, the General Manager adjusts the profile of his factory to match the demand level that his factory faces. ACME's overall strategy is that if expected demand in a region is high, the General manager can ask for more people to be deployed, more machinery to be installed and more space to be rented. If demand is low, the opposite measure is taken. In the back of management's head, this strategy will allow the company to maintain an optimal ratio between the production that is generated, with the company resources used to meet demand.

However, this annual review poses a problem for ACME Inc. Because of the same factory diversity, management is now scratching its head searching for a suitable evaluation technique. Management has to take into account the differences in company resources used by each factory and match it with the corresponding outputs.

* Assistant Professor at the University of the Philippines and holder of Faculty Grant.

II. Data Envelopment Analysis

It is always a difficult task to select the appropriate efficiency measure to be used to evaluate economic performance. ACME Inc. is but one of many examples where management asks: what benchmark can be utilized to measure regional performance? Even though the company may opt for conventional measures such as revenue, cost, or profit targets as its criteria, critics point out that, intangible factors such as manpower and capital are often overlooked in such customary efficiency measures.

A larger scale example that demonstrates this same dilemma would be a Decision Maker (DM)[1] that wants to evaluate the efficiency of an industry vis-à-vis its member companies (DMUs)[2]. The DM has to answer the questions; what measures should be included in our evaluation? Or more generally, which efficiency analysis tool should be used in the evaluation?

An alternative approach to solving this evaluation problem is to use Data Envelopment Analysis (DEA). Conceived by Charnes, Cooper and Rhodes, DEA is a more advanced efficiency approach based on the engineering ratio (outputs over inputs). The first DEA model was developed by Charnes *et al*, (1978) as a non-parametric way of testing the efficiency of DMUs through the use of mathematical programming.

Charnes *(1990)*, describes DEA as a "mathematical programming model applied to observed data that provides a new way of obtaining empirical estimates of extremal relationships such as the production functions and/or efficiency production possibility surfaces that are the cornerstones of modern economics".

Despite its usage of Mathematical Programming in the evaluation process, DEA is now gaining popularity among economists, politicians, and all types of Decision-Makers (DM). DEA has gained favor among these groups since, it is an advanced non-parametric tool and it does not require the diagnostic tests, which are usually associated with statistical models.

In essence, DEA compares several DMUs with respect to their input consumption and output production. It "pits" one DMU against another giving a gauge on whether a certain DMU is overusing inputs and/or if the DMU is under producing in its outputs.

One of the strengths of DEA is that it easily identifies efficient and inefficient DMUs. Furthermore, DEA will indicate how much inefficient DMUs can still improve their performance, and at the same time not violate its technology mix. In the end, DEA will prescribe the necessary action to make the set of DMUs *pareto efficient.*

III. Notations[3]

The following notations will be employed in this paper:

The DM has a collection of n DMUs, which comprise the group to be analyzed. Each DMU j ($j=1,\dots,n$) will consume and produce varying amounts of inputs and outputs, which is common to the entire group of DMUs. DMU j uses x_{ij} ($x_{ij} > 0$) amount of input i and produces y_{rj} ($y_{rj} > 0$) of output r. Index i ranges from $1,\dots, m$, (where m is the total number of input types) and index r ranges from $1,\dots,s$ (where s is the total number of output types).

Each of the n DMUs will then solve a Linear Programming model that is formulated in terms of the x_{ij}s and y_{rj}s. For notational purposes, let the DMU that is being tested be denoted by DMUo. It should be carefully noted that x_{ij} and y_{rj} are constant and known values and they can be geometrically interpreted as points forming certain hyperplanes in the R^{n+m} space. Several surfaces are then constructed out of these hyper-planes which will be required to *envelop* all of these points, thus resulting in the use of the terminology *"Data Envelopment Analysis"*.

IV. Input-Output Oriented Model (IO-OM)[4]

The IO-OM is an extension of the original DEA model of Charnes, Cooper, and Rhodes. It is an Input-Output DEA model which in effect, prescribes the increase needed in the level of output (if production is below par), and/or the decrease needed in the input utilization (if inputs are not efficiently used).

This information profits both the DM and DMU. For the DM, he gains insight into which DMUs are inefficient or efficient through the usage of DEA; aiding him in the central planning of all the DMUs. For the underachieving DMU, DEA will prescribe either to decrease its inputs or to increase its output in order to achieve efficiency[5].

The following is the IO-OM DEA of DMUo :

maximize $\phi^o - \theta^o$
subject to:

(1) $\sum_{j=1}^{n} \lambda_j^o y_{rj} - S_r = \phi^o y_{ro}$ for every $r = 1,\ldots,s$

(2) $\sum_{j=1}^{n} \lambda_j^o x_{ij} + V_i = \theta^o x_{io}$ for every $i = 1,\ldots,m$

(3) $\sum_{j=1}^{n} \lambda_j^o = 1$

(4) $\lambda_j^o \geq 0$ for every $j = 1,\ldots,n$

(5) $\phi^o \geq 1$

(6) $0 \leq \theta^o \leq 1$

For those who are familiar with mathematical programming, it is noteworthy that constraints *(5)* and *(6)* are added for logical and practical reasons. Constraint *(5)* requires that any change in the outputs should be an increase in production, and constraint *(6)* recommends that any change in the inputs should be a decrease in usage.

As mentioned previously, each DMU will have to solve its own 'unique' mathematical program- a tailored made efficiency evaluation for each DMU.

V. Efficiency Criteria

Since we are interested in finding out what improvements can be made with respect to usage/production of inputs/outputs. A DMUo is considered to be IO-OM efficient if it is both *output and input efficient*.

Output Efficient

For the first criteria (output efficiency), let ϕ^{o*} be the output multiplier[6]. It gives DMUo the level that output production can be increased without violating its technological mix. If $\phi o^* > 1$, then DMUo is inefficient. And that for factor output r of DMUo, DEA tells us that it can be increased by the quantity $(\phi o^* - 1)\underline{Y}_o + S_r$. However, if $\phi o^* = 1$ and $S_r = 0$, IO-OM model is also telling us that DMUo is producing all that it should be producing. Hence in this scenario, DMUo is said to be *output efficient* and that no further increases in output production are warranted. All that remains is to verify whether DMUo is also *input efficient*.

Input Efficient

In the same vein to the analysis above, let θo^* be the input multiplier. θo^* indicates whether the inputs are fully utilized by DMUo, which would signal that DMUo is *input efficient*. If $\theta o^* = 1$ and $V_i = 0$, then DMUo is *input efficient* and it does not waste any inputs. Otherwise if $\theta o^* < 1$, then DMUo is inefficient and an improvement can be made. The input reduction needed for DMUo can be retrieved by computing. $(1 - \theta o^*)\underline{X}_o + V_i$.

The Peer Set of DMU*o*

The other information that might be also useful is $\lambda_j o^*$, which identifies the peer set of DMUo in the IO-OM DEA. It was alluded to earlier that DEA 'pits' DMU against DMU. It might be to the interest of a particular DMUo to see what DMUs were used to base DMUo's performance on. To interpret this, for any $\lambda_j o^* > 0$ (for $j \neq$ DMUo), that j^{th} DMU serves as one of the peers of DMUo[7]. So for inefficient DMUs it can be expected that $\lambda_j o^* > 0$ in the end of the mathematical programming process.

VI. ACME Example

Going back to our illustrative company, ACME, the following summarizes the annual performance of each factory, together with the company resources used.

A close inspection of the data reveals that Factory 5 is clearly dominated by Factory 4. In all of the critical factors enumerated by the company, production, factory space, capital machinery and employees, Factory 4 has produced more than Factory 5 while at the same time using less inputs. Hence, it is a trivial matter to say that Factory 5's performance has much to improve on.

To demonstrate how DEA can be used to uncover inefficient Factories, the following LP model was formulated to analyze Factory 5's performance:

$$\text{maximize } \phi o - \theta o$$
$$\text{subject to :}$$

$$5\lambda_1 o + 2\lambda_2 o + 9\lambda_3 o + 3\lambda_4 o + 15\lambda_5 o + 6\lambda_6 o + V_1 = 15\theta o$$
$$3\lambda_1 o + 1\lambda_2 o + 2\lambda_3 o + 2\lambda_4 o + 9\lambda_5 o + 5\lambda_6 o + V_2 = 9\theta o$$
$$7\lambda_1 o + 6\lambda_2 o + 8\lambda_3 o + 4\lambda_4 o + 6\lambda_5 o + 6\lambda_6 o + V_3 = 6\theta o$$
$$10\lambda_1 o + 5\lambda_2 o + 11\lambda_3 o + 5\lambda_4 o + 3\lambda_5 o + 8\lambda_6 o - S_1 = 3\phi o$$
$$\sum_{j=1}^{6} \lambda_j o = 1$$
$$\phi o \geq 1$$
$$0 \leq \theta o \leq 1$$
$$\lambda_j o \geq 0 \qquad \text{for every } j = 1, \ldots, 6$$

Having run this LP formulation, the following are the optimal solutions for ϕo, θo, $\lambda_j o$:

$$\phi o^* = 2.778,\ \theta o^* = 1.0,\ \lambda_1 o^* = .667,\ \lambda_4 o^* = .333$$

FACTORY	PRODUCTION [1]	FACTORY SPACE[2]	CAPITAL MACHINERY	EMPLOYEES[3]
1	10	5	3	7
2	5	2	1	6
3	11	9	2	8
4	5	3	2	4
5	3	15	9	6
6	8	6	5	6

[1] Production is in 10,000 units annually
[2] Factory space is in 1000 sq. ft. rented annually
[3] Employees is in number of people employed in the factory for a given year

The above results show the extent of Factory 5's inefficiency and just confirms our previous observation that Factory 5 is indeed inefficient and is in need of improving its output by *278%* to be considered efficient. Although this figure may shock some people, if you again inspect the amount of inputs used by Factory 5, it is no surprise that a drastic result was given by the IO-OM model.

Other useful information is the peer set of Factory 5, which are Factories 1 and 4 since $\lambda_1 o^* > 0$ and $\lambda_4 o^* > 0$.

For the remaining Factories it is still uncertain to which are inefficient and efficient. Again is a simple matter to use the IO-OM to test the remaining Factories. The summary of the efficiency evaluations is given in the following table:

Factory	PHI	THETA	LAMBDA
1	$\phi o^*=1$	$\theta o^*=1$	$\lambda_1 o^*=1,\ \lambda_2 o^*=\lambda_3 o^*=\lambda_4 o^*=\lambda_5 o^*=\lambda_6 o^*=0$
2	$\phi o^*=1$	$\theta o^*=1$	$\lambda_2 o^*=1,\ \lambda_1 o^*=\lambda_3 o^*=\lambda_4 o^*=\lambda_5 o^*=\lambda_6 o^*=0$
3	$\phi o^*=1$	$\theta o^*=1$	$\lambda_3 o^*=1,\ \lambda_1 o^*=\lambda_2 o^*=\lambda_4 o^*=\lambda_5 o^*=\lambda_6 o^*=0$
4	$\phi o^*=1$	$\theta o^*=1$	$\lambda_4 o^*=1,\ \lambda_1 o^*=\lambda_2 o^*=\lambda_3 o^*=\lambda_5 o^*=\lambda_6 o^*=0$
5	$\phi o^*=2.778$	$\theta o^*=1$	$\lambda_1 o^*=.667,\ \lambda_4 o^*=.333,$ $\lambda_2 o^*=\lambda_3 o^*=\lambda_5 o^*=\lambda_6 o^*=0$
6	$\phi o^*=1.042$	$\theta o^*=1$	$\lambda_1 o^*=.667,\ \lambda_4 o^*=.333,$ $\lambda_2 o^*=\lambda_3 o^*=\lambda_5 o^*=\lambda_6 o^*=0$

Using simple inspection will not identify Factory 6 as being inefficient. However by the results of the IO-OM model, it indicates that Factory 6 needs some improvement in its performance as well. Although Factory 6 needs to improve on its output production (approximately 4%) its desired improvement is marginal and is not as drastic as Factory 5's.

Finally, the above results also show that the first 4 Factories are *pareto efficient* and that no further improvements are justified.

Chinese Cities Illustration[8]

Certain selected cities in China will demonstrate the effectiveness of DEA when the number of DMUs increases in scale. In this scenario, the concern of the DM is to find the inefficient cities in China

The inputs and outputs of the 20 selected cities are shown in the table below:

		Output Levels			Input Levels		
DMU No.	DMU Name	Labor	W. Fund	Inv.	GIOV	P&T	R.S.
1	Shanghai	206.73	1577603	959226	6743346	880041	18316
2	Beijing	110.22	794509	724255	2374342	80119	12790
3	Tianjin	98.38	628243	541587	2252611	538202	6895
4	Shenyang	68.62	419358	198494	1017454	195987	5072
5	Wuhan	64.87	442813	183811	1218527	295199	5090
6	Guangzhou	50.92	389641	354879	1154147	275588	8362
7	Harbin	48.2	356752	138972	672427	124508	3856
8	Chongqing	77.39	443862	210947	1037584	189878	5233
9	Nanjing	42.18	269246	222623	836544	208006	3779
10	Xi'an	46.2	294539	140906	635575	101261	3292
11	Chendu	46.86	304726	173655	734613	150853	4381
12	Changehun	31.34	183319	101556	473369	118062	3460
13	Taiyuan	38.34	221065	170776	513907	84812	1896
14	Dalian	37.96	282373	198278	753961	194512	3708
15	Lanzhou	35.89	235416	107328	580669	140557	2151
16	Jinan	26.39	152034	78312	441724	109039	2441
17	Kunming	27.59	168224	112871	439756	117719	2148
18	Hangzhou	34.32	212524	120028	726172	159354	4106
19	Ningboo	13.52	69895	72845	320516	74492	2866
20	Changsha	18.12	99307	83395	255540	50355	2652

Note:

Labor (unit=10,000 people) is the amount of staff and workers engaged in industries; W. Fund (unit=RMB 10,000) is the amount of circulating capital plus the total annual wage bill of staff and workers in industries; Inv (unit=RMB 10,000) is the annual amount of additional capital supplied for capital construction of state-owned units and technical acquisitions for machinery and fixed assets of collective units; GIOV (unit=RMB 10,000) is the gross industrial output value measured in terms of the total monetary value of outputs produced; P&T (unit=RMB 10,000) is the net contribution to central government in terms of the amount of profit and taxes generated by state-owned enterprises; R.S. (unit=RMB 1,000,000) is the total production value of the final products in the local market

Again by mere inspection, it would be difficult to ascertain which cities are performing below par. The IO-OM was first implemented to determine which cities are inefficient. Assigning Giov to be Output1, P&T to be Output 2 and R.S. to be Output 3, (a similar procedure is done to the Inputs) we can tabulate the results of the IO-OM.

Nearly half (9 out of 20) of the DMUs needs some form of improvement in their performances. The worst rating, as far as output is concerned, is the city *Taiyuan,* which needs to improve its output production by at least 86%. On the other hand of the spectrum, when it comes to resource management of inputs, the city *Chendu* got the worst rating since it needs to cut back on its inputs by at least 18%.

Hence by using the IO-OM model, the DM has a guide on how to improve the performance of the inefficient cities so that in performs on par with the efficient cities.

DMU No.	DMU Name	ϕ^*	θ^*	λo^*
1	Shanghai	1	1	$\lambda_1{}^* = 1$
2	Beijing	1	1	$\lambda_2{}^* = 1$
3	Tianjin	1	1	$\lambda_3{}^* = 1$
4	Shenyang	1	.9952	$\lambda_1{}^* =.0064$ $\lambda_5{}^* = .1134$ $\lambda_6{}^* = .0217$ $\lambda_7{}^* = .1454$ $\lambda_8{}^* = .6275$ $\lambda_{18}{}^* = .\ 0856$
5	Wuhan	1	1	$\lambda_5{}^* = 1$
6	Guangzhou	1	1	$\lambda_6{}^* = 1$
7	Harbin	1.048	.8455	$\lambda_{18}{}^* = .9464$ $\lambda_{19}{}^* = .0536$
8	Chongqing	1	1	$\lambda_8{}^* = 1$
9	Nanjing	1.0879	1	$\lambda_3{}^* = .2615$ $\lambda_5{}^* = .0393$ $\lambda_{14}{}^* = .1819$ $\lambda_{19}{}^* = .5172$
10	Xi'an	1.335	.9729	$\lambda_1{}^* = .0203$ $\lambda_{18}{}^* = .9797$
11	Chendu	1	.8169	$\lambda_6{}^* = .0859$ $\lambda_7{}^* = .1626$ $\lambda_{12}{}^* = .0775$ $\lambda_{18}{}^* = .674$
12	Changehun	1	1	$\lambda_{12}{}^* = 1$
13	Taiyuan	1.8626	1	$\lambda_1{}^* = .0942$ $\lambda_3{}^* = .0164$ $\lambda_{19}{}^* = .8894$
14	Dalian	1	1	$\lambda_{14}{}^* = 1$
15	Lanzhou	1.1188	1	$\lambda_5{}^* = .2478$ $\lambda_{16}{}^* = .6277$ $\lambda_{18}{}^* = .0752$ $\lambda_{19}{}^* = .0492$
16	Jinan	1	1	$\lambda_{16}{}^* = 1$
17	Kunming	1.1674	1	$\lambda_3{}^* = .0356$ $\lambda_5{}^* = .2104$ $\lambda_{19}{}^* = .754$
18	Hangzhou	1	1	$\lambda_{18}{}^* = 1$
19	Ningboo	1	1	$\lambda_{19}{}^* = 1$
20	Changsha	1.0807	.8735	$\lambda_{19}{}^* = 1$

VII. Summary/Suggestions

Clearly, DEA is a tool which management can use to evaluate units vis-à-vis performance. This paper has just demonstrated how DEA can be used in a variety of ways. Whether the DMU is a city, company or division, DEA can ascertain which DMUs are inefficient without having to resort to using parametric measures in the evaluation. Although the process can be complex due to the dependency of DEA on mathematical programming, the information that is gathered from using DEA outweighs the difficulties that are involved.

Even though the concentration of this paper was limited to evaluating the performance of DMUs in previous periods, DEA is not limited to the past. DEA can also be used both for present evaluation or future planning of DMUs. (In the latter case, the target levels of input usage and output production can be evaluated whether it is at par with the current technology level).

Finally, although this paper concentrated on the basic model of DEA and its application to industry, more advanced DEA models have been built throughout the years. Some of the new research in DEA incorporate situations wherein certain inputs/outputs are fixed, constrained or free. Fixed inputs/outputs are variables, which are exogenous in nature (i.e. they are beyond the DM's control). Furthermore, these models allow the inputs/outputs to be changed, but only to a certain degree.

In summary, DEA is an evaluation technique that has very promising future. And as more people are becoming familiar with DEA, it is hoped that its application will be more widespread in industry and government.

REFERENCES

Abara, R., *An Input-Output DEA model employing Fixed, Constrained, and Free Inputs/Outputs,* Unpublished Thesis, 1996.

Ali, A., and Seiford, L., *The Mathematical Programming Approach to Efficiency Analysis,* The Measurement of Productive Efficiency: Techniques and Applications, 1992.

Banker, R.., Conrad, R., and Strauss R. *A Comparative Application of Data Envelopment Analysis and Translog Methods; An Illustration Study of Hospital Production,* Management Science Vol.32, No.1, January 1986.

Banker, R., and Morey, R. *Efficiency Analysis for Exogenously Fixed Inputs and Outputs,* Operations Research Vol. 34, No.4, July-August 1986.

Charnes, A., Cooper, W.W., Huang, Z.M., and Sun, D.B., *DEA Cone-Ratio Approaches for use in Developing Decision Support Systems to Monitor Performance in a Collection of Banks,* Center for Cybernetic Studies Report, September 1989.

Charnes, A., and Cooper, W.W., *DEA Usages and Interpretations,* Graduate School of Business, University of Texas at Austin, May 1990.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L., and Stutz, J., *Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions,* Center for Cybernetic Studies Research Report, November 1984.

Charnes, A., Cooper, W.W., and Rhodes, E., *Measuring the Efficiency of Decision Making Units,* European Journal of Operational Research 2, 1978.

Charnes, A., Cooper, W.W., and Li, S., *Using Data Envelopment Analysis to Evaluate Efficiency in the Economic Performance of Chinese Cities,* Socio-Economic Planning Science Vol.23, No.6, 1989.

Taha, H., *Operations Research: An Introduction,* 5th Edition, Macmillan Publishing Company, 1992.

Tapia, C., *Centralized Data Envelopment for Efficiency Analysis,* Asia Pacific Journal of Operations Research 11, 1994.

Tapia, C., *Data Envelopment Analysis: Basic Theory and Mathematical Programming Models,* Matimyas Matematika, January 1994.

ENDNOTES

[1] We use the term **Decision Maker** broadly here. A DM is any person, group or central planner that initiates the DEA procedure and is in charge of the efficiency analysis.

[2] From this point on we will use the term **DMUs** to generalize any economic entity that is a member of the group that the Decision Maker wishes to examine.

[3] Although understanding the mathematical notation is an arduous process, it has to be used to present the model in compact form. Examples will be provided later to enrich the model.

[4] For more details on the IO-OM model you can refer to Abara (1998)

[5] Applying mathematical programming to the DMUs will assure the DM that the current technological mix of inputs and outputs will be maintained.

[6] Whenever you encounter a variable with a * superscript, it is to mean that that variable is the optimal solution of the mathematical program.

[7] The λ_j^{o*} that are greater than zero form a convex combination of DMUs, forming a facet against which DMU*o* is compared. Any DMU that is efficient should lie on the facet or surface. And any inefficient DMU will be under the production (output) facet or above the utilization (input) facet thus, it is said to be enveloped by the convex combination of efficient DMUs.

[8] The following data was reprinted from Tapia (1994)