A method to consider Non-Discretionary factors in Data Envelopment Analysis

Mehdi Fallah Jelodar [1]

[1]Department of Mathematics, Ayatollah Amoli branch, Islamic Azad University, Amol, Iran.

ABSTRACT. The technique for efficiency measurement known as Data Envelopment Analysis (DEA) has been developed to use non-discretionary inputs that affect measurement. Some methods proposed for measuring efficiency scores to control these factors in production. This paper review these approaches, providing a discussion of strengths and weaknesses and highlighting potential limitations. In addition, a new approach is developed that overcomes existing weaknesses and it is based on relative importance of non-discretionary inputs. To facilitate comparison, a numerical example is used. The results show that the new approach improve pervious models and performs relatively well.

Keywords: Data Envelopment Analysis, Efficiency, Non-discretionary Factors.

2010 Mathematics subject classification: 90Bxx; Secondary 90Cxx.

1. Introduction

Based on the work of Farrell [7], a body of literatures has developed analyzing the technical efficiency of individual decision making units (DMUs). The pioneering work on technical inefficiency of Farrell has received renewed interest in the last decade with the development of Data Envelopment Analysis (DEA), a mathematical programming approach pioneered by Charnes et.al [4] and extended by Banker et.al [1] and Fare et.al [6]. DEA extends the theoretical discussion of technical efficiency

of Farrell into direct measurability by enveloping the observed data to determine a best-practice frontier.

The applicability of DEA, however, depends crucially on the underlying production process. It is essential that the model be consistent with known properties of the production process. Production in the public sector has been modelled by Bradford et al. [3] as a two stage process where final outcomes are determined not only by discretionary inputs but also by environmental (i.e., non-discretionary) factors. Empirical studies of public sector production support this theory; environmental variables have substantial impact on the output produced. A good example the provision of fire services by local communities, measured perhaps by the number of lives saved and/or the dollar amount of property damaged prevented. We would expect that a community with a high proportion of brick houses would be able to provide a higher level of services than a community consisting primarily of old wood houses, given the same discretionary input usage. In this case residential structure is exogenous to the decision of the first department.

Because environmental variables are exogenous to the production process it is necessary to modify the existing DEA models to properly control for these fixed factors. There are three models that incorporate non-discretionary input into DEA models. Banker and Morey [2] provided the first model by modifying the constraints on the fixed factors within the the DEA models. This model differs from the original DEA model by breaking the link between non-discretionary inputs and efficiency. However, as shown by Ruggiero [9], this model does not properly restrict the reference set. In essence, the presence of non-discretionary inputs leads to different frontiers; to control these fixed factors, Ruggiero added constraints to exclude DMUs with a more favorable production environment. Using simulated data, Ruggiero showed the superiority of his model over the Banker/Morey model.

The third method presented by Ray [8], uses a two stage method to control for fixed factors. In the first stage DEA is performed using only discretionary inputs. In the second stage, the efficiency index obtained from the first stage is regressed on the exogenous factors to disentangle inefficiency environmental effect. Adjusting the error term provides a measure of technical efficiency. This approach requires a prior functional form specification for the second stage regression; mis-specification leads to distorted measurement. The purpose of this paper is to compare the last three approaches, highlighting potential strengths and weaknesses. A new method is developed that overcome identified weaknesses.

The rest of this article is organized as follows. In section 2, the technique of DEA is presented. Also, the existing models for controlling

non-discretionary inputs are discussed and potential strengths and weaknesses are identified. Based on this discussion, a new method is developed that handles multiple exogenous factors. Finally, a numerical example is illustrated for this purpose. The last section is conclusion.

2. DEA BACKGROUND

2.1. **DEA models without exogenous inputs.** In this section we are going to summarize some DEA models. In this way consider $n, DMUs$ with m inputs and s outputs. The input and output vectors of DMU_j $(j = 1, \ldots, n)$ are $x_j = (x_{1j}, \ldots, x_{mj})^t$, $y_j = (y_{1j}, \ldots, ysj)^t$ where $x_j \geq 0$, $x_j \neq 0$, $y_j \geq 0$, $y_j \neq 0$.

By using the variable return to scale, convexity and possibility postulates, the non-empty production possibility set (PPS) is defined as follows:

$$T_v = \{(x, y) : \ x \geq \sum_{j=1}^{n} \lambda_j x_j, \ y \leq \sum_{j=1}^{n} \lambda_j y_j, \ \sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \geq 0, j = 1 \ldots, n\}$$

By the above definition the BCC model proposed by Banker et al [1] and based on the work of Farrell is as follows:

$$
\begin{aligned}
F(x_o, y_o) = \min \quad & \theta \\
S.t \quad \sum_{j=1}^{n} \lambda_j x_{ij} &\leq \theta x_{io}, \qquad i = 1, \ldots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} &\geq y_{ro}, \qquad r = 1, \ldots, s \\
\sum_{j=1}^{n} \lambda_j &= 1 \\
\lambda_j \geq 0, \qquad & j = 1, \ldots, n
\end{aligned}
$$

(2.1)

and its dual is as follows:

$$
\begin{aligned}
DF(x_o, y_o) = \max \quad & \sum_{r=1}^{s} u_r y_{ro} - u_0 \\
S.t \quad \sum_{i=1}^{m} v_i x_{io} &= 1 \\
\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - u_0 &\leq 0, \qquad j = 1, \ldots, n \\
u_r \geq 0, \quad & r = 1, \ldots, s \\
v_i \geq 0, \quad & i = 1 \ldots, m
\end{aligned}
$$

(2.2)

Following the work of Lovell [11], the above poduction technology transform inputs $x = (x_1, ..., x_m)$ into output $y = (y_1, ..., y_s)$ for $j = 1, ..., n$ firms can be represented with the input set:
$L(y) = \{x : (x,y) \ is \ feasible\}$. For every output vector y, $L(y)$ has isoquant as follows:
$IsoqL(y) = \{x : x \in L(y), \ \lambda x \notin L(y), \ \lambda \leq 1\}$
and efficient subset
$Eff(y) = \{x : x \in L(y), \ x' \notin L(y), \ x' \leq x\}$

It is important to note that the radial Farrell measure does not require comparison of a given input vector to an input vector that belong to the estimated efficient subset.

2.2. **DEA models with exogenous inputs and a new model of technical efficiency.** The original DEA models assumed that all in puts were discretionary. Assume instead that each DMU uses a vector x of inputs to produce a vector y of outputs given x vector non-discretionary inputs $z = (z_1, ..., z_k)$. These non-discretionary inputs affect on the transformation of discretionary inputs into outputs. For convenience, the vector z is defined so that increases in any component leads to a more favorable environment, ceteris paribus. The production technology transforms input vector X into output vector Y can be represented by the conditional input set:
$L(y|z) = \{x : (x,y) \ is \ feasible \ given \ z\}$
For every output vector y, L(y) has isoquant:
$IsoqL(y|z) = \{x : x \in L(y|z), \ \lambda x \notin L(y|z), \ \lambda \leq 1\}$
and efficient subset
$Eff(y|z) = \{x : x \in L(y|z), \ x' \notin L(y|z), \ x' \leq x\}$
It is assumed that $L(y|z) \subseteq L(y|z')$ implies that z' is more favorable environment than z. Given multiple non-discretionary inputs, it is necessary to identify the importance of each exogenous factors in production process.
The first DEA model to allow continuous exogenous variables was developed by Banker and Morey [2]. Recognizing the inappropriateness of treating fixed factors a discretionary, the authors modified the constraints on the fixed inputs. The BM input oriented (variable return to scale) efficiency measure for production possibility (x_o, y_o) is as follows:

$$BM(x_o, y_o) = \min \quad \theta$$

$$S.t \quad \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, \qquad i = 1, \ldots, m$$

$$\sum_{j=1}^{n} \lambda_j z_{ij} \leq z_{io}, \qquad i = 1, \ldots, k$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, \qquad r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0, \qquad j = 1, \ldots, n$$

(2.3)

and its dual is as follows:

$$DBM(x_o, y_o) = \max \quad \sum_{r=1}^{s} u_r y_{ro} - \sum_{i=1}^{k} w_i z_{io} - u_0$$

$$S.t \quad \sum_{i=1}^{m} v_i x_{io} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{i=1}^{k} w_i z_{ij} - u_0 \leq 0, \qquad j = 1, \ldots, n$$

$$u_r \geq 0, \quad r = 1, \ldots, s$$

$$v_i \geq 0, \quad i = 1 \ldots, m$$

$$w_i \geq 0, \quad i = 1 \ldots, k$$

(2.4)

The constraints on fixed factors are similar to the constraints on discretionary inputs; thy are modified, however, to break the link between efficiency and fixed factors. This modification purportedly controls for fixed factors of production by requiring a convex combination of the referent production possibilities to have an environment no better than the DMU under analysis. Ruggiero [9], however, showed that the referent production possibility may not be feasible, because return to scale should be defined relatively only to discretionary inputs. Enforcing convexity with respect to the non-discretionary inputs leads to improper restriction of the production possibility sets and distorted efficiency measurement.

To evaluate a given DMU, it is necessary to exclude DMUs with a more favorable environment. This was achieved with the public sector model of Ruggiero [9]. The Ruggiero input oriented (variable return to scale) efficiency measure for production possibility (x_o, y_o) is as follows:

$$R1(x_o, y_o) = \min \quad \theta$$
$$S.t \quad \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, \qquad i = 1, \ldots, m$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, \qquad r = 1, \ldots, s \qquad (2.5)$$
$$\lambda_j = 0, \ if \ \exists z_{ij} > z_{io}, \ i = 1, ..., k$$
$$\sum_{j=1}^{n} \lambda_j = 1$$
$$\lambda_j \geq 0, \qquad j = 1, \ldots, n$$

This model explicitly restrict the comparison set to exclude DMUs that face more favorable environment. Similar to the BM model, this model requires a priori specification of the continuous non-discretionary variables. Importantly, as the numbers of continuous fixed factors increases, the probability of identifying a DMU s efficient by default increases. This ignores comparisons between a given DMU and another DMU that overall, has the same or worse environment even though it has a more favorable level of at least one non-discretionary input. This fact suggests an inherent weakness of the ruggerieo model.

To remove these weaknesses Ruggerio [10], modified his model and proposed the following linear programming to measure technical efficiency in the presence of non-discretionary factors:

$$R2(x_o, y_o) = \min \quad \theta$$
$$S.t \quad \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, \qquad i = 1, \ldots, m$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, \qquad r = 1, \ldots, s \qquad (2.6)$$
$$\lambda_j = 0, \ if \ Z_j > Z_o, \ \forall j = 1, ..., n$$
$$\sum_{j=1}^{n} \lambda_j = 1$$
$$\lambda_j \geq 0, \qquad j = 1, \ldots, n$$

Model (6) prevents DMUs with a higher level of the non-discretionary input into reference set. One key assumption in model (6) is that true efficiency is not correlated with non-discretionary factors. As shown in Ruggiero [10], model (6) has some weaknesses. The problem arises because non-discretionary factors has two effects on production: it simultaneously determines the location of the true frontier and effects the distance from the frontier. The efficiency measure $R2(x_o, y_o)$ of model

(6) is unable to disentangle the two effects, attributing both effects to the location of the frontier.

   To remove the difficulties of pervious models, we propose a two stage model. In first stage use model (4) for evaluating all DMUs. Then $\sum_{i=1}^{k} w_i^* z_{ij}$ is the "relative importance" of non-discretionary factors obtained by DBM model. Consider the following linear programming to obtain true efficiency of decision making units:

$$
\begin{aligned}
\theta^{ND} = \min \quad & \theta \\
S.t \quad & \sum_{j=1}^{n} \lambda_j x_{ij} \le \theta x_{io}, \qquad i = 1, \ldots, m \\
& \sum_{j=1}^{n} \lambda_j y_{rj} \ge y_{ro}, \qquad r = 1, \ldots, s \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
\lambda_j = 0, \; & if \; \sum_{i=1}^{k} \overline{w}_i^* z_{ij} < \sum_{i=1}^{k} w_i^* z_{io}, \; \forall j = 1, ..., n \\
& \lambda_j \ge 0, \qquad j = 1, \ldots, n
\end{aligned}
\tag{2.7}
$$

Problem may be occurred when problem 7 has alternative optimal solutions. In this case one may use especial measures to choose one of the solutions. For example, one may use lexico minima of the vector of optimal solutions.
The last model may be used in constant return to scale case. We apply this case for the following example.

3. Example

   Consider 20 iranian bank branches with 2 non-discretionary inputs and 3 discretionary inputs and 4 outputs. The first non-discretionary input is the area of branch and the second is score of staff's education. Normalized data is used to illustration. We added Ruggiero's model (model 6) and proposed model (model 7) through these data. These data and results are summarized in table 1:

DMUs	N-D inputs Z_1 Z_2	D inputs I_1 I_2 I_3	Outputs O_1 O_2 O_3 O_4	Ruggiero	$\sum_{i=1}^{2} w_i^* z_{ij}$	New model
1	0.9333 0.7006	0.8276 0.4404 0.1820	0.0556 0.1077 0.0026 0.5803	0.5804	16E-6	0.7105
2	0.8667 0.9641	0.6810 0.3793 0.1541	0.0518 0.0684 0.0107 0.6871	0.9166	18E-6	1
3	1.0000 0.8862	0.7974 0.3301 0.1030	0.0857 0.1394 0.0039 0.7942	0.8357	19E-6	0.9341
4	1.0000 0.4371	1.0000 0.2058 0.3468	0.2213 0.1611 0.0064 1.0000	0.8635	0.680576	0.9543
5	0.9333 0.7964	0.7888 0.3601 0.1966	0.0949 0.0817 0.0102 0.7922	0.89166	17E-6	0/9687
6	0.9333 1.0000	0.7759 0.4424 0.0000	0.0644 0.0330 0.0038 0.5137	0.7852	19E-6	0.8496
7	0.7333 0.5329	0.6250 0.2579 0.0000	0.0454 0.0890 0.0023 0.5449	1	13E-6	1
8	0.6667 0.7365	0.5733 0.1102 0.0005	0.0253 0.0514 0.0050 0.3229	1	14E-6	1
9	0.6667 0.8563	0.5086 0.2552 0.3636	0.0698 0.1010 0.0007 0.6304	1	15E-6	1
10	0.5333 0.3413	0.5086 0.1698 0.0000	0.0778 0.1401 0.0172 0.4081	1	9E-7	1
11	0.8000 0.6467	0.6940 0.2675 0.0355	0.0687 0.0544 0.0011 0.8005	0.9345	14E-6	1
12	0.6000 0.6228	0.5043 0.0514 0.0011	0.0710 0.0743 0.0031 0.4810	1	12E-6	1
13	0.6000 0.6168	0.5129 0.0617 0.0206	0.0724 0.0447 0.0010 0.3620	0.7221	12E-6	0.7221
14	0.7333 0.6347	0.6207 0.1260 0.1383	0.0591 0.0723 0.0011 0.4404	0.6165	14E-6	0.6971
15	0.6667 0.7725	0.5431 0.0674 0.0044	0.0537 0.0663 0.0124 0.5679	1	14E-6	1
16	0.6000 0.4850	0.5302 0.1461 0.1000	0.0395 0.0532 0.0068 0.4320	0.7811	13E-6	0.8412
17	0.8000 0.4731	0.7672 0.1828 0.0028	0.4035 0.1294 0.0026 0.9467	0.9321	13E-6	1
18	0.4000 0.2575	0.3707 0.0823 0.0043	0.0323 0.0174 0.0007 0.4249	1	7E-7	1
19	0.9333 0.6287	0.8405 0.3395 0.5039	0.1402 0.3263 0.0207 0.6414	1	16E-6	1
20	0.6667 0.3892	0.6336 0.1749 0.0023	0.0911 0.2292 0.0015 0.7404	1	11E-6	1

Table 1: Data and Results

Following to this example all non-discretionary factors should be consider in evaluating efficiency score of each DMU by using Ruggiero's model. But by using the proposed model the efficiency score of DMUs are improved, therefore, this example show that some of the non-discretionary factors have not any correlation with efficiency score but they considered in ruggiero's model. Also the tradition model did not obtain true efficiency score. For example consider DMU 4. It is clear that all non-discretionary inputs are considered for evaluating its efficiency score in Ruggiero's model, but by using the new model the non-discretionary inputs of other DMUs dose not have any effect on efficiency score of this unit, therefore, they did not consider in evaluation and its efficiency score is improved from 0.8635 to 1.

4. Conclusion

This paper has focused on the presence of non-discretionary inputs in production process and in the programming models used to measure inefficiency. The existing models were discussed and their strength and weaknesses were identified. Finally, a new model was developed to overcome these weaknesses. This model based on the "relative importance"

of units obtained by the weights of non-discretionary factors. A numerical example was presented for illustration.

References

[1] Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078-1092.

[2] Banker, R.D, Morey, R.C, 1986, Efficiency analysis for exogenously fixed inputs and outputs. European Journal of Operational Research 553-565.

[3] Bradford, D., Malt, R., Oates, W., 1969. The rising cost of local public services.Some evidences and rreflections, National tax Journal 22 185-202.

[4] Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2, 429-444.

[5] Fare, R., Grosskof, S., Lovell, C.A.K., 1994. Production frontiers, Cambridge University Press, New York.

[6] Fare, R., Grosskof, S., Lovell, C.A.K., 1985. The management of efficiency of production. Klower-Nijhoff , Boston, MA.

[7] Farell, M.J, 1957. The measurement of productive efficiency. Journal of the royal statistical science series A, General 120, 253-281.

[8] Ray, S.c., 1991. Resource use eficiency in public schools. A study of connecticut data, Management Science 37, 1620-1628.

[9] Ruggiero, J., 1996. On the measurement of technical efficiency in the public sector, European Journal of Operational Research553-565.

[10] Ruggiero, J.,2004. Performance evaluation when non-discretionary factors correlate with technical efficiency. European Journal of Operational Research, 159, 250-257.

[11] C.A.K Lovell, 1993. Production frontiers and productive efficiency. Oxford university press, New York, pages 3-67.