Marialisa Mazzocchitti, Davide Quaglione, Alessandro Sarra

DEA approaches to handle environmental factors

(doi: 10.1430/85409)

L'industria (ISSN 0019-7416)
Fascicolo 3, luglio-settembre 2016

Ente di afferenza:
Università Luiss (luiss)

DEA approaches to handle environmental factors

Marialisa Mazzocchitti, Davide Quaglione, Alessandro Sarra

DEA approaches to handle environmental factors

Data envelopment analysis (DEA) is the most extensively applied method in comparative performance measurement. It is widely known that the basic DEA models do not ensure a precise efficiency analysis whenever some factors, which are beyond the managerial control, considerably impact firms' performance. Several new approaches have been proposed to overcome this limitation. The main purpose of this paper is to provide a brief review of the available DEA-based approaches that allow performing a careful efficiency analysis when the performances of the entities under investigation are strongly affected by the operational environment. A multi level classification of these approaches is also provided.

Keywords: Data Envelopment Analysis, Nonparametric Method, Environmental Factors, Continuous Variable, Categorical Variable
JEL Classification: C14; C61

1. INTRODUCTION

One of the main pitfalls in applying a basic DEA model to perform an accurate efficiency analysis concerns the non-homogeneity of the operating environment (Dyson *et al.*, 2001). Oftentimes certain factors – which are beyond the decision-making power of the DMUs under investigation – make uneven the context where the DMUs operate. In other words, though DMUs are homogeneous in terms of transformation process (i.e. they use the same inputs to produce the same outputs), some of them could, more than others, benefit from favourable contextual factors, or suffer unfavourable pressures from the environment. In all such cases, if a basic DEA model is used to assess DMUs performances, one cannot be sure that an entity obtains a high efficiency score because it has been efficiently managed; actually, it could have merely happened because it benefited from favourable contextual factors. Otherwise, an inefficient one may be badly managed or simply experienc-

Marialisa Mazzocchitti, Università degli Studi Gabriele d'Annunzio di Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, marialisa.mazzocchitti@unich.it
Davide Quaglione, Dipartimento di Economia, Università «G. d'Annunzio» di Chieti e Pescara, Viale Pindaro 42, 65127 Pescara, davide.quaglione@unich.it
Alessandro Sarra, Dipartimento di Economia, Università «G. d'Annunzio» di Chieti e Pescara, Viale Pindaro 42, 65127 Pescara, alessandro.sarra@unich.it

TAB. 1. *DEA approaches to treat environmental variables*

Type of Variable		Approach		DEA Model or DEA-based Procedure
Continuous		One-stage model		Banker, Morey 1986a Golany, Roll 1993 Ruggiero, 1996
		Multi-stage procedure	Two-Stage	Ray, 1991 McCarty, Yaisawarng 1993 Bhattacharyya *et al.*, 1997
			Three-Stage	Fried *et al.*, 1996 Ruggiero, 1998 Fried *et al.*, 2002 Muñiz, 2002
			Four-Stage	Fried *et al.*, 1999
Categorical	Hierarchical	Modified model		Banker, Morey 1986b Löber, Staat 2010
	Non-hierarchical	Separation		Charnes *et al.*, 1981 Grosskopf, Valdmanis 1987

ing unfavourable pressures from the environment. That is, comparing DMUs without controlling for the environmental factors might result in a misinterpretation of the results, with differences in the environment being wrongly interpreted as differences in the efficiency (Staat, 1999).

This is a well-known issue, to the extent that the need to make the DEA approaches suitable with the above circumstances has been underlined since when the prominent articles of Charnes *et al.* (1978) and Banker *et al.* (1984) introduced the basic DEA models. At present a wide range of alternative approaches to account for environmental variables is available.

Environmental variables are usually divided into continuous and categorical (Löber, Staat 2010), and accordingly the new approaches proposed for incorporating environmental influences can likewise be divided into two groups: *a*) those which allow handling continuous variables; and, *b*) those which allow handling categorical variables (see Table 1).

2. DEA APPROACHES FOR HANDLING CONTINUOUS ENVIRONMENTAL VARIABLES

Two options are available to deal with continuous variables: a1) single-stage models; a2) multi-stage procedures.

Single-stage models (named also one- or all-in-one- stage models) are extended versions of the basic linear programming model; in other words, they are modified DEA models. The first one was proposed by Banker & Morey (1986a); then, Golany and Roll (1993) and Ruggiero (1996) suggested some adjustments of such model to overcome some of its limitations.

Banker & Morey (1986a) proposed a reformulation of the basic DEA model with variable returns to scale to handle exogenously fixed inputs or outputs; but it has also been used with variables that may influence producer performance, even though they are neither inputs nor outputs of the production process itself. Thus, environmental variables are treated as non-discretionary inputs and outputs (see, for example, De Jaeger, Eyckmans, Rogge, Van Puyenbroeck 2011; and Lozano-Vivas, Pastor, Pastor 2002). The authors suggest inserting the exogenously fixed variable directly into the objective function, along with discretionary inputs and outputs (depending on the influence direction of the exogenous variable); but since the extent to which an exogenous variable may be reduced without requiring an outputs reduction (if it is treated as an input), or increased without requiring an inputs expansion (if it is treated as an output), would be a meaningless result – given that it is exogenously fixed, so the managers may not in any event reduce or increase it – a constraint has to be added to the problem, to specify that it is kept constant. This model is applicable if and only if the direction of each included environmental variable upon efficiency (i.e. whether a specific variable is favourable or detrimental to production) is known in advance.

Golany and Roll (1993) make the Banker and Morey (1986a) model suitable for handling non-discretionary inputs and outputs simultaneously; moreover, they propose an adaptation of the same model for the treatment of partially controlled inputs and outputs. In particular, when there is a level below which an input cannot be reduced, Golany and Roll suggest splitting the total amount of that input into the discretionary portion and the non-discretionary portion. The first portion will be entered into the objective function of the linear programming model, while the non-discretionary one will become an additional constraint.

Ruggiero (1996) proves that the Banker and Morey (1986a) model leads to biased results because, assuming convexity for all factors (both discretionary and non discretionary), it does not properly restrict the reference set; so, to overcome this issue, he developed an adjusted version of the model which includes more constraints to exclude from the comparison set the DMUs that face a more favourable environment. In this way, each DMU may have as a reference only the DMUs that are facing similar or worse environmental conditions. A subsequent paper by the same author (Ruggiero, 1998) shows that the new model cannot handle multiple non-discretionary factors; in order to overcome this weakness the author proposes a novel approach that will be described below, along with other three-stage procedures.

The main characteristic multi-stage procedures have in common is that they require estimating a traditional DEA model without taking into account the role of external factors, in the first stage, and then detect the impact of the environmental variables on efficiency DEA based measures, in one or more additional stages. These approaches can be classified into two-, three- and four-stage procedures.

In two-stage procedures, DEA technique is used in conjunction with other econometric or statistical techniques, like regression analysis and Stochastic Frontier Analysis (SFA). More in details, as in the first stage a traditional DEA model is run with only discretionary variables, while the second stage is used to factor out the effect that environment has on production, thus on DEA scores.

Ray (1991) and McCarty & Yaisawarng (1993) suggest performing a regression analysis, in the second stage, where DEA efficiency measures are regressed on a set of variables representing environmental factors believed to be affecting the performance of the DMUs under investigation. The estimated model is then used for the calculation of the maximum efficiency score achievable by a particular DMU given the specific level of the environmental variables; and the difference between this predicted value and the efficiency score calculated in the first stage (the residuals) will be taken as a measure of the managerial inefficiency which is not caused by external factors. Hence, according to these authors, the technical inefficiency of a DMU should be measured comparing the DEA score obtained with the maximum level achievable by the same DMU given the conditions of the operational environment, rather then measuring its distance from the best practice production frontier.

While Ray apply the Ordinary Least Square (OLS) method for estimating the value of parameters, McCarty & Yaisawarng (1993) adopt the censored regression model known as the tobit model. Which regression method is better or best when used in conjunction with the DEA analysis is a question that has been deeply investigated in several studies (Banker, Natarajan 2008; Hoff, 2007; McDonald, 2009; Ramalho, Ramalho, Henriques 2010; Simar, Wilson 2007); nevertheless, it remains still open (Liu, Lu, Lu 2016).

The two-stage procedure proposed by Bhattacharyya, Lovell, & Sahay (1997) is different from the previous two because these authors specify the second-stage explanatory regression as a stochastic frontier regression model, rather than an OLS or tobit model. They argue that when using an OLS or tobit model, the risk exists that a part of variation in the calculated efficiencies remains unaccounted for, ending up mixing with the white noise error term and contaminating the estimated regression coefficients. Instead, with the stochastic frontier regression model the entire variation in calculated efficiencies is decomposed into three parts: one is associated to exogenous factors (that the authors call «systematic source of variation»); the second is the white noise error term; the remainder one is a «one-sided component» that captures the part of efficiency variation which is not associated with the explanatory variables included in the model.

In three-stage procedures, a DEA model is run at least in the first and in the third stage, in the latter case with the data adjusted in the second stage.

Fried & Lovell (1996) developed the first three-stage procedure totally based on the use of the DEA technique; namely, three DEA sequential stages are designed to obtain an evaluation of the managerial performance. In particular, a DEA model (let's suppose it be an input oriented one) with only discretional variables is performed in the first stage. Total input slacks calculated at the end of this stage become inputs of a radial input oriented second-stage model, the outputs of which will be environmental variables. For those DMUs that are not situated on this second estimated frontier, their distance to the frontier is used to increase the original value of their discretional inputs. The adjusted input data and the original outputs are used to estimate a third DEA model; this last one allows obtaining an assessment of the efficiency and the corresponding classification of DMUs once the effect of environmental variables on their performances has been compensated.

Muñiz (2002) criticises the way in which the original input data are adjusted between the second- and the third-stage and proposes a modification for the just mentioned procedure. More in detail, he suggests *a*) to identify, at the end of the second stage and for each DMU, the coordinates of the point on the efficient frontier used as a benchmark for it (target) – according to Muñiz, this is the part of each slack attributable to the environmental factors, that can be interpreted as the portion of the inefficiency that an entity inevitably suffer given the (fixed) values assumed by environmental variables – and *b*) to subtract these values from original input data.

The three-stage procedure proposed by Fried, Lovell, Schmidt & Yaisawarng (2002) is only partially DEA-based; in fact, each total input slack calculated at the end of the first stage become a dependent variable in a second stage SFA regression model, with the observable environmental variables as the independent ones. The authors claim that the inefficiency detected within the first stage integrates managerial inefficiency, environmental influences, and statistical noise arising from measurement errors; and the SFA regression model allows decomposing each total slack calculated into these three components. Once the parameters and the statistical noise have been obtained, all data required to adjust original input data are available. In order to level the playing field, the authors prefer to adjust upward the inputs of producers who have been advantaged by their relatively favourable operating environments or by their relatively good luck. At the third stage, the same DEA model estimated in the first stage is run, using input adjusted values and original output data. Thus, at the end of the third stage, the evaluation of the solely managerial efficiency is obtained.

Ruggiero (1998) proposes a three-stage procedure where the efficiency measures given by the DEA model estimated (using only discretionary variables) in the first stage are regressed on a set of variables representing the environmental factors in the second stage. The estimated coefficients are then used to build an index of overall environmental harshness that is included in the third stage linear programming model (which is the modified DEA model proposed by Ruggiero, 1996).

A four-stage procedure for obtaining a measure of the managerial efficiency, that controls for the exogenous feature of the operating environment, is introduced by Fried, Schmidt & Yaisawarng (1999). A DEA model (again, let's suppose it be an input oriented one) with controllable factors is estimated in the first stage to calculate the total slack for each input. To quantify the effect of external conditions on the excessive use of inputs, in the second stage, each total input slack becomes the dependent variable of a regression equation, where the independent variables are measures of the environmental conditions applicable to the particular input. Using the estimated coefficients and real data on external variables, in the third stage, predicted total input slack for each input and for each unit are calculated. These predictions are used to adjust for the influence of environmental conditions input data used in the first stage. In the fourth stage, the DEA model estimated in the first stage is used, but with the adjusted values. Eventually, an efficiency measure is obtained which is not affected by environmental factors.

3. DEA APPROACHES FOR HANDLING CATEGORICAL ENVIRONMENTAL VARIABLES

Within the set of categorical variables, we can distinguish those that allow establishing a specific order among categories (hierarchical or ordinal variables) from those that do not allow ranking categories according to any specific order (non-hierarchical variables).

To deal with a hierarchical variable, one can select the modified model proposed by Banker & Morey (1986b) or that of Löber & Staat (2010).

Banker & Morey (1986b) suggest a way to carry out correctly an efficiency analysis through DEA when the DMUs under investigation operate in different competitive environments. Since the comparison should be made among entities with fairly similar general conditions, though a traditional DEA model does not guarantee that this happens, Banker and Morey propose a modified model that ensure that the reference group of each DMU is composed of DMUs operating in an even more difficult or unfavourable situation. More in details, one of the constraints of the traditional DEA model has been replaced by three new constraints created to exclude from the reference set of a given DMU the ones that benefit from a more favourable situation (i.e. that belong to higher categories). The case study they use to illustrate an application of the proposed modified model concerns pharmacies operating in communities with different sizes. The size of communities is a continuous variable transformed into the categorical hierarchical variable. They categorize the market size into 11 categories and classify pharmacies according to this categorization; then, in order to avoid that the ones located in smaller communities be compared to peers located in larger communities, they adopt the proposed modified DEA model to evaluate each pharmacy.

Löber & Staat (2010) observe that the Banker & Morey (1986b) model has rarely been used for empirical studies because it requires extensive data manipulations and numerous reformulations of new program codes; so they propose a more flexible solution to handle with categorical hierarchical variables (but extensible also to non-hierarchical data) that require neither programming nor repeated DEA runs. Their proposal relates to the construction and the inclusion into the DEA model of «indicator variables» that allow similar DMUs to be grouped together.

Charnes *et al.* (1981) propose a procedure to handle non-hierarchical categorical variables and apply it to a case where the categorical variable is dichotomous (or binary). The proposed procedure consists in splitting the set of DMUs into two subsets and run two separate DEA models, one for each subset. Once all the DEA scores are obtained, the inefficient entities must be projected on their frontier. Projected points will be used to estimate a new unique DEA model. This allows the assessment of any difference in the mean efficiency of the two sub-samples.

A procedure quite similar to that proposed by Charnes *et al.* (1981) has been used by Grosskopf & Valdmanis (1987) for assessing the relative performance of public and not-for-profit Californian hospitals. They first run a DEA model for all DMUs to obtain an «overall efficiency index» for each of them. Then, they partition the sample by ownership and calculate the efficiency of each observation relative to its separate ownership group's frontier, named «within-group efficiency index». Calculating efficiency for the pooled as well as the partitioned samples allows them to calculate also the «between-group efficiency». At the end, they compare the best practice frontiers of public and not-for-profit hospitals and decompose the efficiency based on the pooled sample into a within group and between group component.

References

Banker R.D., Charnes A., Cooper W.W. (1984), *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis*, in «Management Science», 30, 9, pp. 1078-1092, http://doi.org/10.1287/mnsc.30.9.1078.

Banker R.D., Morey R.C. (1986a), *Efficiency Analysis for Exogenously Fixed Inputs and Outputs*, in «Operations Research», 34, 4, pp. 513-521.

Banker R.D., Morey R.C. (1986b), *The Use of Categorical Variables in Data Envelopment Analysis*, in «Management Science», 32, 12, pp. 1613-1627.

Banker R.D., Natarajan R. (2008), *Evaluating Contextual Using Data Variables Envelopment*

Affecting Analysis, in «Operations Research», 56, 1, pp. 48-58, http://doi.org/10.1287/opre.l070.0460.

Bhattacharyya A., Lovell C.A.K., Sahay P. (1997), *The Impact of Liberalization on the Productive Efficiency of Indian Commercial Banks*, in «European Journal of Operational Research», 98, 2, pp. 332-345, http://doi.org/10.1016/S0377-2217(96)00351-7.

Charnes A., Cooper W.W., Rhodes E. (1978), *Measuring the Efficiency of Decision Making Units*, in «European Journal of Operational Research», 2, pp. 429-444, http://doi.org/10.1016/0377-2217(78)90138-8.

Charnes A., Cooper W.W., Rhodes E. (1981), *Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow*, in «Management Science», 27, 6, pp. 668-697.

De Jaeger S., Eyckmans J., Rogge N., Van Puyenbroeck T. (2011), *Wasteful Waste-reducing Policies? The Impact of Waste Reduction Policy Instruments on Collection and Processing Costs of Municipal Solid Waste*, in «Waste Management», 31, 7, pp. 1429-1440, http://doi.org/10.1016/j.wasman.2011.02.021.

Dyson R.G., Allen R., Camanho A.S., Podinovski V.V., Sarrico C.S., Shale E.A. (2001), *Pitfalls and Protocols in DEA*, in «European Journal of Operational Research», 132, 2, pp. 245-259. http://doi.org/10.1016/S0377-2217(00)00149-1.

Fried H.O., Lovell, C.A.K. (1996), *Searching the Zeds*, Working paper presented at II Georgia productivity workshop.

Fried H.O., Lovell C.A.K., Schmidt S.S., Yaisawarng S. (2002), *Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis*, in «Journal of Productivity Analysis», 17, 1-2, pp. 157-174, http://doi.org/10.1023/A:1013548723393.

Fried H.O., Schmidt S.S., Yaisawarng S. (1999), *Incorporating the Operating Environment Into a Nonparametric Measure of Technical Efficiency*, in «Journal of Productivity Analysis», 12, pp. 249-267, http://doi.org/10.1023/a:1007800306752.

Golany B., Roll Y. (1993), *Some Extensions of Techniques to Handle Non-Discretionary Factors in Data Envelopment Analysis*, in «Journal of Productivity Analysis», 4, 4, pp. 419-432.

Grosskopf S., Valdmanis V. (1987), *Measuring Hospital Performance. A Non-parametric Approach*, in «Journal of Health Economics», 6, 2, pp. 89-107, http://doi.org/10.1016/0167-6296(87)90001-4.

Hoff A. (2007), *Second Stage DEA: Comparison of Approaches for Modelling the DEA Score*, in «European Journal of Operational Research», 181, 1, pp. 425-435, http://doi.org/10.1016/j.ejor.2006.05.019.

Liu J.S., Lu L.Y.Y., Lu, W.-M. (2016), *Research Fronts in Data Envelopment Analysis*, in «Omega», 58, pp. 33-45, http://doi.org/10.1016/j.omega.2015.04.004.

Löber G., Staat M. (2010), *Integrating Categorical Variables in Data Envelopment Analysis Models: A Simple Solution Technique*, in «European Journal of Operational Research», 202, 3, pp. 810-818, http://doi.org/10.1016/j.ejor.2009.05.032.

Lozano-Vivas A., Pastor J.T., Pastor J.M. (2002), *An Efficiency Comparison of European Banking Systems Operating under Different Environmental Conditions*, in «Journal of Productivity Analysis», 18, 1, pp. 59-77, http://doi.org/10.1023/A:1015704510270.

McCarty T.A., Yaisawarng S. (1993), *Technical Efficiency in New Jersey School Districts*, in Fried H.O., Lovell C.A.K., Schmidt S.S. (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, pp. 271-287.

McDonald J. (2009), *Using Least Squares and Tobit in Second Stage DEA Efficiency Analyses*, in «European Journal of Operational Research», 197, 2, pp. 792-798, http://doi.org/10.1016/j.ejor.2008.07.039.

Muñiz M. (2002), *Separating Managerial Inefficiency and External Conditions in Data Envelopment Analysis*, in «European Journal of Operational Research», 143, pp. 625-643.

Ramalho E.A., Ramalho J.J.S., Henriques P.D. (2010), *Fractional Regression Models for Second Stage DEA Efficiency Analyses*, in «Journal of Productivity Analysis», 34, 3, pp. 239-255, http://doi.org/10.1007/s11123-010-0184-0.
Ray S.C. (1991), *Resource-Use Efficiency in Public Schools: A Study of Connecticut*, in «Management Science», 37, 12, pp. 1620-1628.
Ruggiero J. (1996), *On the Measurement of Technical Efficiency in the Public Sector*, in «European Journal of Operational Research», 90, 3, pp. 553-565, http://doi.org/10.1016/0377-2217(94)00346-7.
Ruggiero J. (1998), *Non-discretionary Inputs in Data Envelopment Analysis*, in «European Journal of Operational Research», 111, pp. 461-469.
Simar L., Wilson P.W. (2007), *Estimation and Inference in Two-stage, Semi-parametric Models of Production Processes*, in «Journal of Economic», 136, 1, pp. 31-64.
Staat M. (1999), *Treating Non-discretionary Variables One Way or the Other: Implications for Efficiency Scores and Their Interpretation*, in Westermann G. (Ed.), *Data Envelopment Analysis in the Service Sector*, pp. 23-49.