Decision Support

Evaluation of system efficiency using the Monte Carlo DEA: The case of small health areas

Mercedes Torres-Jiménez a,⁎, Carlos R. García-Alonso b, Luis Salvador-Carulla c, Vicente Fernández-Rodríguez d

a Department of Mathematics and Engineering, Universidad Loyola Andalucía
b Department of Mathematics and Engineering, Universidad Loyola Andalucía, Escritor Castilla Aguayo, 4. 14004 Córdoba, Spain

c Brain and Mind Research Institute, and Centre for Disability Research and Policy Faculty of Health Sciences, University of Sydney, Sydney, Australia

d Department of Business Organization, Universidad Loyola Andalucía

ARTICLE INFO

Article history:
Received 6 February 2013
Accepted 9 October 2014
Available online 16 October 2014

Keywords:
Data Envelopment Analysis
Simulation
Expert knowledge
Operations research in medicine
Efficiency of health care areas

ABSTRACT

This paper uses Monte Carlo Data Envelopment Analysis (Monte Carlo DEA) to evaluate the relative technical efficiency of small health care areas in probabilistic terms with respect to both mental health care as well as the efficiency of the whole system. Taking into account that the number of areas did not permit maximum discrimination to be achieved, all the scenarios of non-correlated inputs and outputs of a specific size were designed using Monte Carlo Pearson to maximize the discrimination of Monte Carlo DEA and the information included in the models. A knowledge base was included in the simulation engine in order to guide the dynamic interpretation of non-standard inputs and outputs. Results show the probability that all DMU and the whole system have of being efficient, as well as the specific inputs and outputs that make the areas or the system efficient or inefficient, along with a classification of the areas into four groups according to their efficiency (k-means cluster analysis). This final classification was compared with an expert-based classification to validate both the knowledge base and the Monte Carlo DEA model. Both classifications showed results that were very similar although not exactly the same, basically due to the difficulty experts experience in recognizing “intermediately-inefficient” DMU. We propose this methodology as an instrument that could help health care managers to assess relative technical efficiency in complex systems under uncertainty.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Introduced by Charnes, Cooper, and Rhodes (1978), Data Envelopment Analysis (DEA) is a non-parametric method that evaluates the relative technical efficiency of a set of comparable Decision Making Units (DMU), each using multiple inputs to produce multiple outputs. DEA has been used in many decisional situations related to health care (Brandeu, Sainfort, & Pierskalla., 2004; Cheng & Zervopoulos, 2014; Du, Wang, & Chen, 2011; Kontodimopoulos, Bellai, Labiris, & Niakas, 2006; Ozcan, Lins, Da silva, Fiszman, & Pereira, 2010; Prior, 2006; Salvador-Carulla, Garcia-Alonso, Gonzalez-Caballer, & Garrido-Cumbra, 2007). However, in these complex stochastic systems, the application of standard DEA has some relevant drawbacks, such as that: (i) the results obtained do not agree with previous and well established expert opinions (Salvador-Carulla et al., 2007) if inputs and outputs (I/O) are not correctly interpreted (non-standard I/O); (ii) input and output values (variables in DEA models) are stochastic and the selection of appropriate statistical distributions to fit them is not a trivial matter; and (iii) the number of observations to be evaluated in the system (relative technical efficiency) is usually low compared to the number of I/O, which compromises the discriminating power of DEA models. For (i), weight control using specific constraints can be useful (DEA can have feasibility problems if these constraints are very restrictive) but only when I/O are correctly interpreted. In (ii), the statistical distribution selection for I/O values (all of them random variables) depends on the availability of enough real data (Kolmogorov-Smirnov-Lilliefors or maximum likelihood tests can be used) or, instead, on expert knowledge. This second option is very frequent and usually needs to check different statistical distributions before obtaining results. Finally, for (iii), other models like order-α, and order-m, can be used (Cazals, Florens, & Simar, 2002; Wheelock & Wilson, 2003, 2004a, 2004b), but DEA is also a nice option if discriminating enough scenarios (Dyson et al., 2001) are designed (combinations of non-correlated I/O). This approach (Salvador-Carulla et al., 2007) has another relevant advantage: each scenario offers a
different perspective of the relative technical efficiency of the system under study, and the analysis becomes multi-dimensional for experts and decision makers.

Recently, the integration of expert knowledge in operational models has been the object of significant scientific attention (Bose, 2003; Salvador-Carulla et al., 2007). Decision makers, who deal with the intrinsic risk of their decisions, need reliable and useful solutions from operational models like DEA, order-α, order-m, etc. Evaluating these solutions negatively can be because the initial formulation of the operational model does not match the real framework under analysis (it cannot represent the complexity of the environment). Incorporating a knowledge base in DEA models is a requirement to help them interpret I/O values and their relationships, especially when they are non-standard, undesirable or flexible I/O (Cook & Zhu, 2007; Gibert, García-Alonso, & Salvador-Carulla, 2010; Seiford & Zhu, 2002) that have to be managed.

A knowledge base designed by information transfer between experts and analysts in any DEA model includes: (i) I/O types—standard or non-standard—with different ranges of values; (ii) their corresponding statistical distribution; and, finally, (iii) their relative relevance (weights) in the system. Results are, therefore, guided by explicit expert knowledge and, in addition, can also be used to improve experts’ knowledge in a circular pursuit of excellence.

When DEA models evaluate complex systems, I/O may have flexible measures. Some variables can play I/O roles depending on the circumstances (Cook & Zhu, 2007) and can have different meanings for experts depending on their specific values. Non-standard, undesirable or flexible I/O are very frequent in healthcare systems and can be handled in DEA models using four different basic approaches (Liang, Yongjun, & Shubing, 2009): (i) the hyperbolic measure approach (Färe, Grosskopf, Lovell, & Pasurka, 1988); (ii) the transformation of non-standard outputs to inputs and vice-versa (Hailu & Veeman, 2001; Reinhard, Lovell, & Thijszen, 2000); (iii) the data transformation function approach (Athanasopoulos & Thanassoulis, 1995; Lovell, Pastor, & Turner, 1995; Scheel, 2001; Seiford & Zhu, 2002); and, finally, (iv) the directional distance-function approach (Chung, Färe, & Grosskopf, 1997; Färe & Grosskopf, 2004).

In stochastic DEA, the I/O interpretation (non-standard or flexible I/O) has to be guided by the knowledge base to apply the appropriate mathematical transformation.

Standard DEA models always assume that I/O values are known, but in real-life decisional situations (as in healthcare management), I/O are stochastic because they can be missing or based on expert judgment and/or predictions (Zhu, 2003). Stochastic DEA models explore data variation and there are many approaches to this operational problem (Dyson & Shale, 2010): sensitivity and stability analysis (Färe, Grosskopf, & Lovell, 1994; Neralic, 2004; Seiford & Zhu, 1998), stochastic frontiers (Banker, 1993; Banker & Nararajan, 2004; Ruggiero, 2004), chance-constrained DEA (Land, Lovell, & Thore, 1993; Olesen & Petersen, 1995), fuzzy DEA (León, Liern, & Ruiz, 2003; Lertworasirikul, Fang, Jonies, & Nuttle, 2003), Imprecise DEA (Zhu, 2003; Cook & Zhu, 2006) and Monte Carlo DEA (Kao & Liu, 2009; Krüger, 2012; Perelman & Santin, 2009).

According to Ingalls (2008), Monte Carlo simulation is a computational tool for modelling and analysing complex systems under uncertainty. This procedure lets operational modellers analyse complex and stochastic models with fewer assumptions (Kao & Liu, 2009) and therefore allows for the design of a more faithful representation of the real system. A Monte Carlo simulation engine can include almost any operational model made to design experiments to check DEA properties (Banker, Gadh, & Gorr, 1993; Perelman & Santin, 2009; Ruggiero, 1999), but has not yet been widely enough used to evaluate probabilistic relative technical efficiency.

In Monte Carlo DEA, the I/O values of each DMU are simulated from their statistical distribution to determine the distribution of each DMU’s relative technical efficiency. Therefore, Monte Carlo DEA understands that both I/O values and their resulting DMU relative efficiencies are stochastic. The selection of the statistical distribution for I/O in the knowledge base is always critical because simulation results are heavily dependent upon it.

On the other hand, the selection of the I/O for a DEA model is usually forced by the data available and can be subjective because it includes expert assumptions about system behaviour (Allen, Athanasopoulos, Dyson, & Thanassoulis, 1997; Cook & Zhu, 2007). Specifically for DEA models, the number of the I/O selected needs to be small compared to the total number of DMU, to reach an effective discrimination of the latter. Dyson et al. (2001) suggested that the number of DMU n should be at least two times the product of the number of inputs s and number of outputs r (n ≥ 2sr). If this does not happen, some of the I/O should be removed from the DEA model, but this selection is rarely obvious. Statistical correlation offers DEA modellers additional information that can be used for this selection (Farzipoor Saen, Memariani, & Hosseinzadeh Lotfi, 2005; Jenkins & Anderson, 2003). When I/O are stochastic, Monte Carlo simulation can also be used to carry out the correlation analysis (i.e. Monte Carlo Pearson) to design non-correlated I/O combinations of a specific size (scenarios).

The main objective of this paper is to demonstrate that it is possible to include random variables (inputs/outputs) in a relative technical efficiency-analysis to avoid problems associated with other approaches like Imprecise-DEA. For this, a Monte Carlo DEA model was applied to evaluate the relative technical efficiency of a system composed of 12 small health-care areas. This includes the evaluation of technical efficiency in probabilistic terms in both the DMUs and the system as a whole. Our model also includes a knowledge base for the algebraic interpretation of inputs/outputs. This evaluation could help health care managers to provide solutions for improving operational efficiency and to reduce the wasting of resources, while assuring a higher level of service quality.

This paper is structured as follows: Section 2 describes the methodology: an application to assess the efficiency of mental health care areas is carried out in Section 3; and, finally, some illustrative comments and conclusions are drawn in Sections 4–6.

2. Development of expert-oriented Monte Carlo DEA

2.1. The Monte Carlo DEA model

The Monte Carlo DEA methodology can integrate any DEA model into a simulation engine. The former evaluates the relative technical efficiency of each DMU once the latter has determined the I/O values for a specific simulation. This procedure generates a hybrid (statistical and operational) model that can be generalized by integrating other operational models like order-α and order-m ones (Wheelock & Wilson, 2003, 2004a, 2004b). According to this strategy, the process has four sections (Table 1): (i) I/O values are determined at random according to their specific statistical distribution (steps 1, 2, 17 and 18); (ii) original I/O values are interpreted—mathematically transformed—according to the knowledge base (steps 3–5); (iii) DEA models are designed automatically and solved (steps 6–10) and the solutions—relative technical efficiency—are saved in a file; finally, (iv) the statistical error is calculated (Nakayama, 2008) and if it is low enough, the simulation process ends (steps 11–15).

The number of simulations performed for each DMU and scenario depends on the statistical error that can make the simulation process terminate (Nakayama, 2008). Taking into account that our Monte Carlo DEA model is a steady-state process, the non-overlapping batch means the method is used to evaluate statistical error. The simulation process finishes once the error is lower than a small predefined value (the percentage over the mean of efficiency scores); once this situation is reached repetitively, the number of simulations does not greatly affect the efficiency mean (Kao & Liu, 2009).
Table 1
Structure of Monte Carlo DEA model (pseudo-code).

1. For \( i=1 \) to \( n^{i} \)
2. I/O values are determined at random (according to specific and expert-based statistical distributions) by the Monte Carlo simulation engine
3. If I/O values instantiate rules in the expert-based rulebase
4. The corresponding I/O are transformed (using, for example, a linear monotone decreasing transformation)
5. End if
6. For \( j=1 \) to \( n^{j} \)
7. The corresponding DEA model is designed
8. DEA model is resolved
9. Results are saved in a solution file (pool of solutions)
10. Next \( n^{i} \)
11. If \( n^{i} \geq n^{i} \min \)
12. The statistical error \( (\text{St}e) \) is calculated
13. If \( \text{St}e < \text{St}e_{\min} \)
14. Go to 18
15. End if
16. End if
17. Next \( n^{i} \)
18. End

\( ^{a} n^{i} \text{min} \): maximum number of simulations.
\( ^{b} n \): number of Decision Making Units (DMU).
\( ^{c} n^{i} \min \): minimum number of simulations to calculate statistical error.
\( ^{d} \text{St}e_{\min} \): minimum value of statistical error to stop the steady-state simulation process.

2.2. The I/O selection in non-correlated scenarios

For a number \( d \) of DMU, the number of all possible I/O mathematical combinations without repetition (scenarios) \( c \) can be very high:

\[
c = \binom{n}{i} \cdot \binom{p}{j}
\]  

(1)

where \( n \) is the total number of inputs, \( p \) the total number of outputs, \( i \) the number of inputs selected for the DEA model and, finally, \( j \) the number of outputs selected for the same DEA model.

Our decisional framework is composed of 12 DMU of mental health care, eight inputs and four outputs. This system cannot be analysed in a standard way (Dyson et al. 2001) because the rule \( 2 \times n \times p \leq d \) is not fulfilled \((2 \times 8 \times 4 > 12)\), and the resulting models will not be discriminating enough.

If scenarios of three inputs by two outputs are designed, the number of all possible combinations \( (1) \) is 336, great enough to make the problem computationally unfeasible. The number of scenarios can be reduced taking into account those combinations which are composed of non-correlated sets of \( i, j \) I/O. The selection of non-correlated I/O avoids the problem of a discretionary selection of variables in DEA models. These scenarios offer different technical perspectives for the same decisional situations and can be easily interpreted by experts because their structures are simpler. On the other hand, some I/O can be over-weighted because they might block others in the design process. If something like this happens, an appropriate selection of feasible scenarios is required.

Our decisional problem is stochastic and the correlation analysis cannot be standard. Monte Carlo correlation analysis (Pearson) is used to evaluate correlations between I/O according to their specific statistical distributions. This procedure generates thousands of 12 (DMU) by 12 (8I + 4O) matrixes and evaluates the correlation coefficient between I/O \((\alpha = 0.05)\). If the number of simulations where two I/O correlate significantly is greater than an expert-based predefined value (for example, greater than 95 percent of the simulations), and the mean of the correlation coefficients is greater than or equal to another predefined value (for example, 0.9; Farzipoor et al., 2005), then the two variables are highly correlated.

Once the scenarios are designed, Monte Carlo DEA systematically calculates the relative technical efficiency of each DMU in each scenario. Once the result sets are obtained, both the DMU probability of being efficient as well as its basic statistics can be calculated easily. For each DMU, in probabilistic terms, many different views of relative technical efficiency (one for each scenario) are obtained as well as the global behaviour of the system.

2.3. Incorporating expert knowledge in a Monte Carlo DEA model

Expert knowledge is absolutely necessary to guide the simulation process because of: (i) the selection of appropriate statistical distribution for all I/O, and (ii) the interpretation of non-standard I/O values in each simulation (Salvador-Carulla et al., 2007). These non-standard I/O need to be interpreted according to an expert-driven model because, if not, DEA will give incorrect results. For example, TR2, the available number of R2 type care units in a small health care area (Fig. 1), is a non-conventional variable because values close to 1 are considered very appropriate by health planners. Taking into account that this is an input, the best value that DEA can manage for it is \( \varepsilon \), a non-Archimedean number smaller than any real number—almost zero—(no input consumption). Values close to 1 need to be reduced and DMU with these original input values will prove to be better from a relative technical efficiency point of view.

The Monte Carlo DEA randomly generates a different DEA model for each DMU and simulation. The original I/O values (selected according to their statistical distributions) can instantiate corresponding rules in the knowledge base. These rules are structured in a rulebase—a standard relational database—included in the main knowledge base. The structure of the rulebase includes the information needed to transform the original I/O values algebraically (2) when specific conditions—I/O values—are fulfilled (Fig. 1).

\[
x_{i}^{t} = x_{i}^{o} \otimes f_{i}
\]  

(2)

where \( x_{i}^{t} \) is the transformed I/O value, \( x_{i}^{o} \) is the original value calculated by the simulation engine and, finally, \( f_{i} \) is the...
Table 2
Service description according to the European Service Mapping Schedule (ESMS), variable description and technical characteristics of the Expert-driven Model of Community Care (expert driven model).

<table>
<thead>
<tr>
<th>Grouping of services</th>
<th>Description of ESMS “Main Types of Care” (standard codes)</th>
<th>Variables</th>
<th>Technical characteristics of the expert driven model. Rates per 100,000 population except TR2, TR4R7, TRR8R13 and TD1 + D4 (number of main types of care).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute care Acute (R2)</td>
<td>Residential/Hospital</td>
<td>TR2, Input</td>
<td>High availability and utilisation by users from the area but avoiding over-use. Initial best practices: TR2 in a [1, 1.5] range; PR2 in [9.20] range and UR2 medium-high to high avoiding over-use in a [10.19] range.</td>
</tr>
<tr>
<td>Acute (R2)</td>
<td></td>
<td>UR2, Output</td>
<td></td>
</tr>
<tr>
<td>Non-acute hospital care R4 to R7</td>
<td>Residential/Hospital</td>
<td>TR4R7, Input</td>
<td>Low availability and utilisation but not 0. Initial best practices: TR4R7 low, between [1.31]; PR4R7 low, between [3.13] and UR4R7 low use, between [3.12].</td>
</tr>
<tr>
<td>R4 to R7</td>
<td></td>
<td>PR4R7, Input</td>
<td></td>
</tr>
<tr>
<td>Residential community care R8 to R13</td>
<td>Residential/Non-hospital</td>
<td>TRR8R13, Input</td>
<td>High availability and utilisation. Initial best practices: TRR8R13 high, greater than 3; PRR8R13 high, greater than 10 and URR8R13 high, avoiding over-use, between [10.40].</td>
</tr>
<tr>
<td>R8 to R13</td>
<td></td>
<td>PRR8R13, Input</td>
<td></td>
</tr>
<tr>
<td>Day care D1: Day care/Acute D4: Day care/Non-acute/</td>
<td>Residential/Non-hospital</td>
<td>TR1D4, Input</td>
<td>High availability and utilisation. Initial best practices: TD1 + D4 high, greater than 3; PD1 + D4 high, greater than 34 and UD1 + D4 high, greater than 33.</td>
</tr>
<tr>
<td>Residential/Non-hospital</td>
<td></td>
<td>D1 + D4, Output</td>
<td></td>
</tr>
<tr>
<td>Other structured activities</td>
<td></td>
<td>UD1 + D4, Output</td>
<td></td>
</tr>
</tbody>
</table>

- Types (T): Number of “main types of care” (R2, R4 to R7, R8 to R13 and D1 + D4) available in the DMU. Places (P): Places or beds available at the DMU. Utilization (U): Utilization of “main type of care” by patients from the area. Rates per 100,000 inhabitants.

a Types (T): Number of “main types of care” (R2, R4 to R7, R8 to R13 and D1 + D4) available in the DMU. Places (P): Places or beds available at the DMU. Utilization (U): Utilization of “main type of care” by patients from the area. Rates per 100,000 inhabitants.

b D1 includes day hospitals.

Table 3
Expert-based interpretation, based on the Expert-driven Model of Community Care (expert-driven model), of the inputs selected (rates per 100,000 population).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range</th>
<th>Interpretation based on expert driven model</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR2 [0,1]</td>
<td>$x_i = -4.32x_i + 4.32$ (Fig. 1)</td>
<td></td>
</tr>
<tr>
<td>TR2 [1,1.5]</td>
<td>$x_i = 3x_i - 3$ (Fig. 1)</td>
<td></td>
</tr>
<tr>
<td>TR2 [1.5,20]</td>
<td>Standard input: no changes ($x_i = x_i^t$). (Fig. 1)</td>
<td></td>
</tr>
<tr>
<td>PR2 [0,20]</td>
<td>$x_i = -0.93x_i + 18.59$</td>
<td></td>
</tr>
<tr>
<td>PR2 [20,100]</td>
<td>$x_i = x_i^t - 20$</td>
<td></td>
</tr>
<tr>
<td>TR4R7 [0,1,9]</td>
<td>$x_i = x_i^t + 1.9$</td>
<td></td>
</tr>
<tr>
<td>TR4R7 [1,9,3.1]</td>
<td>$x_i = 2.58x_i + 4.91$</td>
<td></td>
</tr>
<tr>
<td>TR4R7 [3,1,15]</td>
<td>Standard input: no changes ($x_i = x_i^t$).</td>
<td></td>
</tr>
<tr>
<td>PR4R7 [0,3]</td>
<td>$x_i = -x_i^t + 3$</td>
<td></td>
</tr>
<tr>
<td>PR4R7 [3,13]</td>
<td>$x_i = 1.33x_i + 3.9$</td>
<td></td>
</tr>
<tr>
<td>TRR8R13 [6,6]</td>
<td>Standard input: no changes ($x_i = x_i^t$)</td>
<td></td>
</tr>
<tr>
<td>TRR8R13 [0,10]</td>
<td>$x_i = 6 - x_i^t$</td>
<td></td>
</tr>
<tr>
<td>TRR8R13 [0,17]</td>
<td>$x_i = 17 - x_i^t$</td>
<td></td>
</tr>
<tr>
<td>TD1 + D4 [0,10]</td>
<td>$x_i = 10 - x_i^t$</td>
<td></td>
</tr>
<tr>
<td>PD1 + D4 [28,42]</td>
<td>$x_i = 42 - x_i^t$</td>
<td></td>
</tr>
</tbody>
</table>

- Original value of the variable in the database. $x_i^t$: Transformed value of the variable for DEA (input or output value used in the algebric model (3)).

Table 4
Expert-based interpretation, based on the Expert-driven Model of Community Care (expert-driven model), of the outputs selected (rates per 100,000 population).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>UR2 [0,6]</td>
<td>Values highly penalized (all of them close to 0)</td>
<td></td>
</tr>
<tr>
<td>UR2 [6,19]</td>
<td>Standard output</td>
<td></td>
</tr>
<tr>
<td>UR2 [19,100]</td>
<td>Values highly penalized (all of them close to 0)</td>
<td></td>
</tr>
<tr>
<td>UR4R7 [0,3]</td>
<td>Values highly penalized (all of them close to 0)</td>
<td></td>
</tr>
<tr>
<td>UR4R7 [3,13]</td>
<td>Standard output</td>
<td></td>
</tr>
<tr>
<td>UR4R7 [13,100]</td>
<td>Values highly penalized (all of them close to 0)</td>
<td></td>
</tr>
<tr>
<td>UR8R13 [0,40]</td>
<td>Standard output</td>
<td></td>
</tr>
<tr>
<td>UR8R13 [40,100]</td>
<td>Non-standard output</td>
<td></td>
</tr>
<tr>
<td>UD1 + D4 [0,15]</td>
<td>Values highly penalized (all of them close to 0)</td>
<td></td>
</tr>
<tr>
<td>UD1 + D4 [15,37]</td>
<td>Standard output</td>
<td></td>
</tr>
<tr>
<td>UD1 + D4 [37,100]</td>
<td>Non-standard output</td>
<td></td>
</tr>
</tbody>
</table>

In a steady-state simulation model where the system evolves in an infinite horizon, the evaluation of error is delicate basically due to the calculation of the variance (Nakayama, 2008). The existence of independent and identically distributed (normal distribution) data cannot be assumed so the estimation of the output variance is not a trivial issue. The method of multiple replications (Nakayama, 2008) is used to deal with this drawback. This method is based on designing (sufficiently large) independent and identically distributed replications, each of them with $\lambda$ simulations. Independence is achieved by generating non-overlapping series of random numbers in the simulation process. The replications are identically distributed if all of them commence under the same initial conditions. Taking this process into account, a sample variance of $r$ simulated values can be calculated, as well as their error and confidence intervals.

This error is critical because it determines when the simulation engine stops (it defines the number of simulations $n_{sim}$, Table 1). When the error (in percentage over the mean) is repeatedly lower than a specific number (usually, 2.5 percent), the steady state model stops.

2.5. Validating the model
Monte Carlo simulation models are completely blind, which implies that the validation of the results they give is one of the main challenges to be dealt with. In our case, the Monte Carlo DEA model gives us the probability that each DMU has of being efficient (or inefficient), taking into account their I/O statistical distribution and the
expert-driven model. These probabilities are therefore biased by the expert knowledge integrated into our model and, without any additional information, cannot be validated at all. For validation purposes, the panel of experts that designed the B-MHCC model also classified the 12 DMU into four groups: efficient, nearly efficient, inefficient and doubtful. This expert-based classification is used to validate both the expert-driven model and the Monte Carlo DEA models. Once the probabilities that each DMU and scenario has of being efficient have been calculated, a k-means cluster analysis is carried out to classify the DMU into four groups \((k = 4)\) in order to match it to expert-based classification. An intra-class correlation analysis is carried out to evaluate the degree of agreement among them (McGrav & Wong, 1996).

3. Evaluation of the relative technical efficiency of a complex system: The case of small health areas in mental health care

3.1. Structure of the database and scenario design

The PSICOST-12 database used to check the Monte Carlo DEA model includes 12 widely different DMU in Spain, described by their health care structure (Table 2) classified according to the European Service Mapping Schedule model. All existing health and social services for mental health care were evaluated in every DMU by an external expert. The data for each catchment area were aggregated into residential use, structured day activities, continuous out-patient care and emergency out-patient care, as shown in Table 2. The variables have been described previously (Salvador-Carulla et al., 2009). This in-bables chosen are standard according to the eDESDE-LTC instrument (http://www.edesdeproject.eu/). The selection of specific variables in each scenario was done according to two criteria:

- finding sufficiently discriminating DEA models (Dyson et al., 2001). The number of DMU (12) is very small: in this situation the problem of dimension must be taken into account (Alirezaee, Howland, Van de Panne, 1998; Staat, 2001) The Monte Carlo DEA model completely analyses the statistical distributions selected for inputs/outputs. This process shows if there are relevant changes in the relative technical efficiency distribution when input/output values vary. If this occurs, Monte Carlo DEA can identify the specific input/output ranges where the size of the example becomes critical.
- To avoid correlated I/O. All of the selected 3 \(\times\) 2 combinations (scenarios) have non-correlated I/O.

All of the variables have been described previously (Salvador-Carulla et al., 2007) and all have a specific meaning for decision makers (health planners).

According to the expert driven model (Table 2), all DMU inputs and outputs were interpreted numerically (Tables 3 and 4) and their values were fitted to triangular statistical distributions (some examples in Table 5, data available on request) through modal estimators (Table 4 shows an example for D1, D6 and D9). The knowledge base includes 144 statistical distributions—12 \(\times\) (8 + 4)—in addition to the rules generated by the expert-driven model.

3.2. The DEA model

The DEA model selected (3) was the standard input-oriented BCC model—variable returns to scale (steps 7 and 8 in Table 1). The BCC model was selected because there is no evidence that inputs/outputs in small health areas can show constant returns to scale behaviour (Salvador-Carulla et al., 2007). In the assessment of health areas, input consumption is always the real problem once the outputs are known: this is why the input-oriented model was selected for the example.

Once the simulation engine has generated I/O values, they can be interpreted according to the expert-oriented rulebase structure. This process randomly generates expert-interpreted BCC-DEA models. The rules in the rulebase are automatically instantiated once I/O values have been generated. All the rules are semantically expressed and formalised in a standard relational database.

\[
\begin{align*}
\text{Min} \theta - \epsilon & \left( \sum_{h=1}^{H} S_h + \sum_{j=1}^{J} S_j \right) \\
\text{s.t.} & \sum_{m=1}^{D} \lambda_m \lambda_m + S_h = \theta \lambda_m, h = 1, 2, \ldots, i \\
& \sum_{m=1}^{D} y_{rm} \lambda_m - S_j = y_{ro}, r = 1, 2, \ldots, j \\
& \sum_{m=1}^{D} \lambda_m = 1 \\
& \lambda_m \geq 0; m = 1, 2, \ldots, d
\end{align*}
\]

Where \(D\) is the number of DMU, \(i\) the number of inputs and \(j\) the number of outputs. The DMU \(m\) consumes \(x_{hm}\) of input \(h\) and produces \(y_{rm}\) of output \(r\). \(\theta\) is the efficiency score and \(S_h\) and \(S_j\) are the slacks.

The existence of outliers is usually a big problem in DEA when the number of DMU is low. Other methods like order-\(\alpha\) and order-\(\rho\) are less sensitive to this situation and can also be included in the simulation engine (Wheelock & Wilson, 2003, 2004a and 2004b).

3.3. Determining scenarios

According to Dyson et al. (2001), 12 DMU do not make the BCC-DEA model discriminating enough with eight inputs and four outputs. In this case, a scenario analysis was designed to improve the global design. These scenarios of non-correlated I/O were designed automatically once the Monte Carlo Pearson model (two-tailed value 0.708 with 10 degrees of freedom, \(a = 0.05\) and number of simulations = 10,000) had shown the existing correlations between I/O. In

---

**Table 5**

Some examples of the statistical distributions chosen for selected DMU: D1, D6 and D9.

<table>
<thead>
<tr>
<th>Variable</th>
<th>D1*</th>
<th>D6</th>
<th>D9</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR2</td>
<td>[1.071,1.19,1.309]</td>
<td>[1.764,1.96,2.156]</td>
<td>[1.476,1.64,1.804]</td>
</tr>
<tr>
<td>PR2</td>
<td>[9.072,10.08,11.088]</td>
<td>[41.976,46.64,51.304]</td>
<td>[5.958,6.62,7.282]</td>
</tr>
<tr>
<td>TR4R7</td>
<td>[1.071,1.19,1.309]</td>
<td>[1.793,1.31,1.441]</td>
<td>[7.052,2.28,3.608]</td>
</tr>
<tr>
<td>PR4R7</td>
<td>[39.564,43.96,48.356]</td>
<td>[32.121,35.69,39.259]</td>
<td>[1.41,69.46,33.50,963]</td>
</tr>
<tr>
<td>TR8R13</td>
<td>[1.611,1.79,1.969]</td>
<td>[1.764,1.96,2.156]</td>
<td>[1.476,1.64,1.804]</td>
</tr>
<tr>
<td>PR8R13</td>
<td>[9.819,10.91,12.001]</td>
<td>[14.13,15.77,17.27]</td>
<td>[11.799,13.11,14.421]</td>
</tr>
<tr>
<td>TD1 + D4</td>
<td>[1.071,1.19,1.309]</td>
<td>[1.793,1.31,1.441]</td>
<td>[1.476,1.64,1.804]</td>
</tr>
<tr>
<td>TD1 + D4</td>
<td>[3.726,4.14,4.554]</td>
<td>[21.465,23.85,26.235]</td>
<td>[36.873,40.97,45.067]</td>
</tr>
<tr>
<td>TD1 + D4</td>
<td>[7.4,8.22,9.04]</td>
<td>[21.74,24.16,26.58]</td>
<td>[7.48,8.31,9.14]</td>
</tr>
<tr>
<td>UR2</td>
<td>[7.4,8.22,9.04]</td>
<td>[21.74,24.16,26.58]</td>
<td>[7.48,8.31,9.14]</td>
</tr>
<tr>
<td>UR4R7</td>
<td>[10.79,11.99,13.19]</td>
<td>[5.89,6.54,7.19]</td>
<td>[10.97,12.19,13.41]</td>
</tr>
<tr>
<td>UR8R13</td>
<td>[17.72,19.69,21.66]</td>
<td>[2.35,2.61,2.87]</td>
<td>[4.43,4.92,5.41]</td>
</tr>
<tr>
<td>UD1 + D4</td>
<td>[13.72,15.24,16.76]</td>
<td>[53.59,54.35,65.49]</td>
<td>[17.42,19.355,21.29]</td>
</tr>
</tbody>
</table>

* T: standard triangular distribution [minimum, modal estimator, maximum].
order to reach a compromise between the maximum discrimination of BCC-DEA models and the maximum information to be included, scenarios with \( i = 3 \) inputs and \( j = 2 \) outputs were designed.

3.4. The Monte Carlo DEA model

As stated before, our simulation model is steady state. The number of simulations, \( n_{sim} \) (Table 1), was determined by the statistical error; the simulation engine stopped when this error (percentage over the mean) was repeatedly (20 times) lower than 2.5 percent. The error was calculated using 10 batches (\( r = 10 \)) after the first 500 simulations. For safety reasons, a maximum number of simulations of 10,000 was selected for stopping the process if the error did not converge (Table 1).

4. Results

Our procedure generated 336 scenarios, but only 27 included non-correlated I/O (Table 6); these final scenarios were analysed using Monte Carlo DEA. The total number of simulations performed was 629,420; the number of simulations in each scenario was \([22,222; 24,737]\); and, finally, the average number of simulations per DMU in each scenario was \([1,852; 2,061]\). The software was run in a 6-processor computer.

Table 7 shows their probabilities of being efficient for the 12 DMU (D1 to D12) (relative technical efficiency obtained from the BCC-DEA model equal to one, and all the slacks equal to zero). These DMU, none completely efficient, show very different patterns depending on the scenario analysed (Fig. 2 shows the results for four DMU). Not only can we evaluate if a DMU is more or less efficient, but we can also determine specific non-correlated I/O combinations (Table 6) that make them more or less efficient (27 scenarios of a specific size with three inputs and two outputs, Table 7). The probability of being efficient is a stochastic estimator of relative technical efficiency that could be complementary to other central tendency statistics (like mean or median) because the representation of these statistics could be questionable due to statistical distribution of efficiency scores (which can show non-symmetric patterns and can also be multimodal) (Fig. 3). This probability was compared to order-\( \alpha \) and order-m results for identifying potential outliers. For these latter procedures, only I/O central estimators were taken into consideration. Due to this, the analysis cannot be considered probabilistic (the relative technical efficiency cannot be calculated as a probability as in Monte Carlo DEA). As expected (all DMU were studied in depth), there are not any outliers and results were similar to those obtained using the Monte Carlo DEA (other methodologies like order-\( \alpha \) and order-m methods can also be incorporated into the Monte Carlo simulation engine).

The mean probability of being efficient in all the scenarios (Table 7) is far from 1, between 0.238 (scenario 16) and 0.614 (scenario 26). As expected (Nakajima, 2008), the standard deviation is very high according to a steady-state simulation model. The mean is lower than 0.5 in most of the scenarios, so their probability of being efficient is under 0.5. Scenarios 16 and 18 are the worst; their means become especially inefficient DMU in combinations of the number of the European Service Mapping Schedule codes available in acute care (TR2), the number of the European Service Mapping Schedule codes available in residential community care (TR8R13), the number of European Service Mapping Schedule codes available in day care (TD1 + D4), utilization per 100,000 inhabitants in non-acute hospital care (UR4R7) and utilization per 100,000 inhabitants in residential community care (UR8R13) or utilization per 100,000 inhabitants in day care (UD1 + UD4). These are the variables that should probably be the first to be improved from a system-efficiency point of view. Scenario 23 is the best: in this case TR2, TD1 + D4, bed-places per 100,000 inhabitants in day care (PD1 + D4), UR8R13 and UD1 + D4 variables make DMU efficient. We can readily observe that D1 + D4 services

Table 7

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Input 1</th>
<th>Input 2</th>
<th>Input 3</th>
<th>Output 1</th>
<th>Output 2</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>TR2</td>
<td>TR2</td>
<td>TD1</td>
<td>D4</td>
<td>UR4R7</td>
<td>UR8R13</td>
</tr>
<tr>
<td>D2</td>
<td>.99</td>
<td>.99</td>
<td>.74</td>
<td>.97</td>
<td>.94</td>
<td>.95</td>
</tr>
<tr>
<td>D3</td>
<td>.90</td>
<td>.00</td>
<td>.90</td>
<td>.79</td>
<td>.79</td>
<td>.47</td>
</tr>
<tr>
<td>D4</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>D6</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>D7</td>
<td>.24</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>D8</td>
<td>.06</td>
<td>.18</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>D9</td>
<td>.52</td>
<td>.71</td>
<td>.00</td>
<td>.79</td>
<td>.80</td>
<td>.81</td>
</tr>
<tr>
<td>D10</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>D11</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>D12</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Mean</td>
<td>.33</td>
<td>.42</td>
<td>.46</td>
<td>.46</td>
<td>.46</td>
<td>.46</td>
</tr>
<tr>
<td>Median</td>
<td>.13</td>
<td>.36</td>
<td>.21</td>
<td>.55</td>
<td>.48</td>
<td>.37</td>
</tr>
</tbody>
</table>

* SD: Standard Deviation.
dominate this last scenario and are probably responsible for the increased probability of being efficient. Variable codes are described in more detail in Table 2.

On analysing specific DMU, we can see that, for example, D1 is an efficient DMU in all the scenarios analysed except number 4 and 27 (Table 7 and Fig. 2). In both cases, the presence of UR8R13 instead of UR4R7 (S4) or UD1 + D4 (S27) makes the difference: D1’s probability of being efficient falls substantially. On the other hand, D6 is an inefficient DMU (Table 7 and Fig. 2) except in scenarios 22 and 23. In this case, the combination of TR2, TD1 + D4, PD1 + D4 and UD1 + D4 is especially favourable when incorporating UR4R7 and/or UR8R13. Some of these I/O also appear in scenario 24 where D6 has a relatively high probability of being efficient.

Taking into account the scenarios designed, the probability of being efficient for all DMU can be calculated individually (in each scenario), along with the corresponding probability for the whole system (all scenarios and DMU as a whole), which was 0.3956 (Table 8). This probability shows an important margin for improvement and can be
used as an initial reference to evaluate potential DMU improvements because of the likely existence of trade-offs between individual DMU and system efficiency improvements.

The efficiency mean, its standard deviation and the corresponding variation coefficient, have been also included in Table 8 for each DMU considering all the scenarios. These basic statistics can be also used to evaluate the situation for the DMU and the whole system, taking into account the special characteristics of the efficiency statistical distributions (Fig. 3). Comparing DMU efficiency mean and their probability of being efficient both are very similar, especially in the most efficient and inefficient cases. The probability of being efficient is more restrictive than the efficiency mean. Due to this, some DMU with a relatively high efficiency mean can have a low probability of being efficient (D2, Table 8).

Some of the I/O (especially TR2) are more relevant in the analysis of the whole system because they dominate scenario design (based on non-correlated I/O, Table 6). These global probabilities highlight that D1 and D10 are efficient DMU while D6 is completely inefficient and the rest of the DMU are difficult to classify.

Taking the matrix of probabilities into account (Table 7), a cluster analysis (k-means) was carried out. Expert classification identified four DMU groups: efficient areas, inefficient areas, nearly efficient areas and, finally, doubtful areas. For this reason, k was initially fixed at 4. According to global efficiency (the probability of being efficient taking into consideration all the scenario results), the cluster analysis (Table 8) showed that: cluster 1 groups the efficient areas (global probability of being efficient greater than 0.9); cluster 2—the inefficient ones (global probability of being efficient ranged between [0.28, 0.41]); cluster 3—the very inefficient ones (global probability of being efficient ranged between [0.09, 0.31]); and, finally, cluster 4—the intermediately-inefficient ones (global probability of being efficient ranged between [0.43, 0.62]: in this group the variability was very high). The similarities between expert and cluster classification are evident when considering:

- The areas considered as “nearly efficient” for experts are really inefficient ones for the cluster analysis (except D1, which is efficient for the cluster analysis). They are inefficient but the experts cannot explain this situation in a precise way.
- The inefficient areas for the experts are “very inefficient” for the cluster analysis (except D3 and D11 that are simply inefficient for the cluster analysis).
- The “doubtful” areas for the experts are intermediately-inefficient for the cluster analysis (due to its variability, this group can also be called “doubtful”). In this situation, experts have the same identification problem that appears in the “nearly efficient” areas.

So we can say that the experts identified very inefficient areas very precisely (they call them “inefficient”), intermediately-inefficient areas are more complex for them. Taking into account that this classification is ordinal, the differences between the two are logical (neighbouring categories).

To confirm the existence of four clusters, an ANOVA test was carried out to check the existence of significant differences between the cluster means (p-value = 0.000, means are different). To confirm if all the clusters show significant differences in their means, a post-hoc range test was carried out. Table 9 shows the results of the Tuckey Test: all the means showed significant differences in their means, a post-hoc range test was carried out. Table 9 shows the results of the Tuckey Test: all the means showed significant differences, which confirms the existence of four groups. A k-means analysis where k = 3 was also carried out in the group denominated as “doubtful”, composed of D9 and D12, disappeared; the former was considered intermediately-inefficient and the latter efficient. Taking into account the great
variability of these two DMU, we considered that it was not appropriate to consider them one group. Due to this and the power of discrimination in the $k$-means analysis with $k = 4$, this was the number of clusters finally considered.

5. Discussion

The software that carried out the Monte Carlo DEA model has been designed to run on a multi-processor computer. The process is very computer-demanding and needs careful testing before the analysis begins, mainly for the checking of DEA model sizes—see (2). There were no additional problems related to the feasibility of linear DEA models unless we tried to be very strict about constraining I/O weights. In these cases, the number of simulations with non-feasible DEA models increased a great deal. The number of simulations per DMU and scenario was not so high; all were stopped by statistical error.

The results showed that the whole system was globally inefficient. Only two areas (17 percent) could be considered efficient while 50 percent of the areas studied were inefficient. These results showed an important margin of improvement in the majority of the DMU studied.

We also found almost perfect agreement (Landis & Koch, 1977) between cluster analysis and the expert-based classification provided by a panel of experts (Intraclass Correlation Coefficient, 0.856; individual cases and 0.922: mean scores, $p$-values = 0 in both cases). The existence of two efficient DMU (D1 and D10) was confirmed by the cluster analysis (group 1) although experts considered D1 to be nearly efficient, and they included this area in group 2. In this case, taking into account Monte Carlo DEA results (Table 7), we can conclude that D1 was technically efficient, although the panel of experts explicitly considered scenarios 24 and 27 as more relevant when classified by DMU. Group 2 obtained from the cluster analysis included one area that was classified as “nearly efficient” by the experts and another that was considered inefficient. This could be motivated by the explicit representation of the experts’ mental framework; they distinguished these DMU as a little bit more efficient than the rest but not enough to consider them clearly efficient. Cluster analysis (groups 3 and 4) agreed with the panel of experts when classifying both inefficient and doubtful DMU (Table 8). The doubtful DMU group was characterised by the chaotic behaviour identified graphically in Fig. 2 in the case of area D12. In any case, these areas showed some nearly inefficient patterns according to expert criteria, or intermediate-inefficient, found in cluster analysis, for example, in units like D3.

Therefore, classifications based on expert knowledge and cluster analysis showed very similar results, although they were not exactly the same, due basically to the difficulty that the experts had in recognizing “intermediately-inefficient” DMU. At this time it is important to try to answer a key question: Is it necessary for the DEA model to coincide with experts’ classification? Even though expert opinion could be wrong, in this specific example we consider that the structure of their knowledge was basically correct (Salvador-Carulla et al., 2007), so the coincidence between the two classifications (the Monte Carlo DEA and the experts’ opinion) could be a good thing. However, the differences in classification could not be ignored because experts cannot combine all of the variables in a global model as DEA does, so the method can be better than the experts’ opinion. In our problem, expert opinion belonged to a Community Care model that assessed us on the “appropriateness” of each variable value. It is important to note that these experts had no relationship whatsoever with the areas analysed, so they could be considered unbiased. The objective of comparing Monte Carlo DEA results to the classification based on expert opinion is to validate the proposed Expert-driven Model for Community Care (rulebase). Our results showed that this model can be incorporated in a Monte Carlo DEA as a knowledge base (rulebase). This knowledge base is useful because:

1. Not always do decision makers have a panel of experts when deciding on a policy. Our proposal can help them in these circumstances because expert knowledge is included in the operational analysis through the use of a rule-base.
2. Relative technical efficiency models cannot be used without expert-based data interpretation. Thus, results can be completely useless because decision makers do not recognise the real picture that they know exists.
3. Decision makers can learn from probabilistic relative technical efficiency as well as from results that coincide with their opinions and also, on the other hand, from results that show a different picture. They can analyse the reasons for the agreements and disagreements variable by variable (I/O) and can calculate their potential improvements in a stochastic way.

So we propose this methodology as an instrument that could help health care managers to evaluate and improve their operational efficiency as well as reduce the wasting of resources, while assuring a higher level in service quality.

6. Conclusions

It is absolutely possible to include an operational model and a knowledge base in a Monte Carlo simulation engine. This procedure lets researchers and potential users analyse relative technical efficiency in complex systems where some or all variable values are statistical distributions.

Health systems and health areas are complex phenomena which require data analysis methods that consider non-linearity
De Savigny & Adam, 2009). This study demonstrates that Monte Carlo DEA can be successfully used in complex probabilistic DEA models, whatever their structure might be, to determine the statistical distribution of technical efficiency of either any DMU—small health area—or the whole system. The main advantage of this method is that explicit expert knowledge can be included once it is formalized in a standard way (for example using a relational database). Expert knowledge guides operational analysis and, on the other hand, results obtained from the analysis can improve expert knowledge in a positive iterative process. Expert knowledge structures the knowledge-base that includes the statistical distribution selected for I/O, the Expert-driven Model of Community Care and the management of I/O weights.

Relative technical efficiency depends upon the statistical distribution selected for each DMU and I/O. The existence of any kind of empirical evidence with respect to its behaviour is really critical for the design of the model. The design of the Expert-driven Model of Community Care needed a panel of management experts in health care systems who, in the end, could validate the results obtained by the Monte Carlo DEA model. Finally, I/O weights need to be carefully managed because they can control the relevance of specific I/O in all or some of the scenarios designed and can make many algebraic models unfeasible.

The Expert-driven Model of Community Care is composed of rules that are instantiated once the I/O original values have been determined by the simulation engine. These rules can be more or less complicated, from linear monotone transformations (as in this paper) to fuzzy rules. The process of evaluating rule outputs is completely automatic and can be easily included in a Monte Carlo simulation engine. The design of non-correlated scenarios offers experts the opportunity to analyse various technical perspectives of the same problem. The aggregation of all scenarios to evaluate the efficiency of the whole system must be done very carefully to avoid an overweighting of some I/O that could dominate the scenario design process.

The efficiency of each DMU is expressed in terms of probability and lets the operational designer evaluate potential improvements in its I/O. Taking into account the existence of multiple scenarios designed by sets of non-correlated I/O, I/O improvements in one scenario for a specific DMU can have unexpected results in other scenarios (positive or negative) and on the efficiency of the whole system. The study of these potential trade-offs is the next challenge facing Monte Carlo DEA.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Spanish Inter-Ministerial Commission of Science and Technology (MICYT) (TIN2011-22794), FEDER funds and the Andalusian Research Plan (PAI: P06-CTS-01765), and P08-TIC-3745 project of the “Junta de Andalucía” (Spain).

The Monte Carlo DEA model has been developed by the authors with the collaboration of service researchers from PSICOST (Salvador-Carulla et al., 2007). This new model has been tested on the PSICOST-12 database, a reliable and comprehensive database of service availability and use in various small mental health areas using the eDESDE-LTC coding system of services (Salvador-Carulla et al., 2009) based on the European Service Mapping Schedule (Johnson & Kuhlmann, 2000). This work is part of the P11/02008 project (Spatial Analysis and Ordinal Classification of the geographical distribution of mental illness in Andalusia) of the “Instituto de Salud Carlos III” (Spain).

References


