Expert Systems with Applications 39 (2012) 10495–10499

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Uncover the predictive structure of healthcare efficiency applying a bootstrapped data envelopment analysis

Arianna De Nicola [a], Simone Gitto [b], Paolo Mancuso [c,*]

[a] De Nicola: Dipartimento di Ingegneria dell'Impresa. Università di Roma "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy
[b] Gitto: Dipartimento di Ingegneria dell'Impresa. Università di Roma "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy
[c] Mancuso: Dipartimento di Ingegneria dell'Impresa. Università di Roma "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy

ARTICLE INFO

Keywords:
Bootstrap
Data envelopment analysis (DEA)
Classification and regression trees
Environmental variables
Health policy
Efficiency
Patient mobility

ABSTRACT

One of the main problems in efficiency analysis is to determinate the environmental variables that have an impact on the production process. This paper shows that applying bootstrap to data envelopment analysis (DEA) before performing classification and regression trees (CART) increase the quality of the results. In particular, employing data on the Italian Health System, the paper highlights that bias corrected DEA allows to individuate variables affecting health efficiency which would remain undiscovered when the traditional DEA model is applied.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Many models have been developed to find an optimal solution to the problem to improve healthcare efficiency. In this paper, the concept of efficiency, measured by data envelopment analysis (DEA), is implemented together to classification and regression trees (CART) analysis to provide a set of rules that permit to identify on which environmental variables the governments should operate to improve healthcare efficiency. DEA is a well know non-parametric method developed by Charnes, Cooper, and Rhodes (1978) that identifies a production frontier and determines the efficiency scores of a set of decision making units (DMU), with the common set of inputs and outputs (Heidari & Mohammadi, 2012; Lin, Lee, & Chiu, 2009). In the other hand, one of the significant limits on applying this non-parametric technology is that the efficiency scores are an estimate of the true (and unknown) production frontier, conditional on observed data resulting from an underlying Data Generating Process (DGP) (Simar & Wilson, 1998, 2000). As a consequence, DEA efficiencies are biased by construction and are sensitive to the sampling variations of the obtained frontier. In order to overcome this problem, Simar and Wilson (1998) proposed a bootstrap procedure to approximate the sampling distribution of the efficiency scores and to make inference. See Halkos and Tzeremes (2012), Curi, Gitto and Mancuso (2011) and Gitto and Mancuso (2012) for recent applications of bootstrap-DEA methodology.

The CART methodology (Breiman, Friedman, Olshen, & Stone, 1984) which allows to identify some rules with the aim to classify a sample into two or more groups, has been applied in different fields (D'uva & De Siano, 2007; Li, Sun, & Wu, 2010; Sohn & Tae, 2004). Nowadays, to the best of our knowledge, it is never applied to support policy intervention in the health system.

In this paper, bootstrapped DEA and CART analysis are implemented in order to define policy intervention aimed to improve health efficiency. Moreover, this study discusses the importance to use DEA in an inferential setting by employing the bootstrap technique.

1.1. Research objectives

The main objectives of this study are:

1. To demonstrate the applicability of the CART methodology in the health sector.
2. To stress the importance of the bootstrap in DEA analysis.

2. The proposed methodology

The methodology used in the paper is illustrated in Fig. 1. It is composed by two different stages; the first deals with DEA technique while the second concerns the use of CART technique.

2.1. Data envelopment analysis

DEA is an efficiency evaluation approach entirely based on the observed data. The main concept is that the efficiency of a specific DMU is determined by its capability to obtain desirable outputs

* Corresponding author. Tel.: +39 0672597793; fax: +39 0672597305.
E-mail addresses: arianna.de.nicola@uniroma2.it (A. De Nicola), simone.gitto@uniroma2.it (S. Gitto), paolo.mancuso@uniroma2.it (P. Mancuso).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.02.074

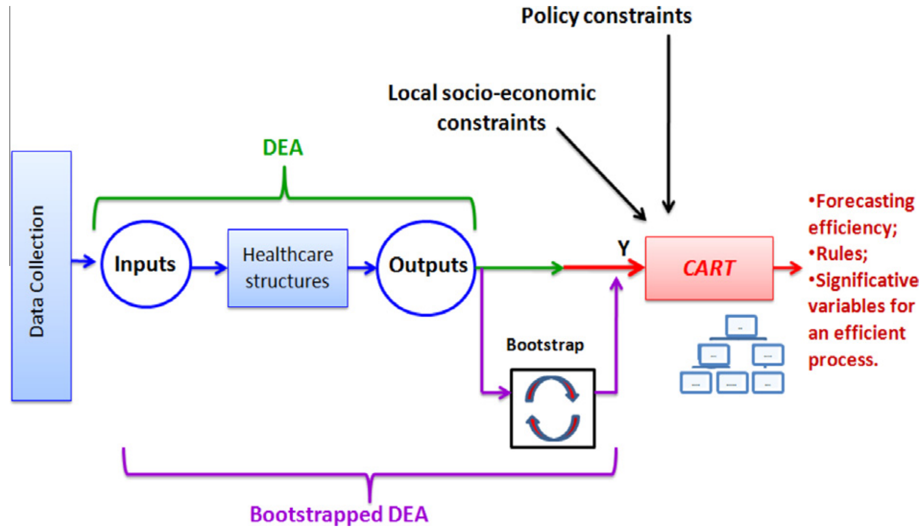

Fig. 1. The methodology.

from a set of inputs. So, this methodology constructs an efficient production frontier based on the best practice, applying a linear programming technique to the established sample. In order to facilitate the interpretation of the results in the next sections, it is useful to recall that in the output orientated DEA model, under the hypothesis of variable return to scale (VRS), an efficiency score $\hat{D}_{it}$ is calculated for each DMU i ($i = 1,2,\ldots,n$) at time t (t=1,2,...,T), by solving the following linear program:

$$\begin{aligned} &\hat{\theta}_{it} = [\hat{D}_{it}]^{-1} = \max_{\theta\lambda} \theta \\ &\quad s.t. x_{it} \geqslant X_t \lambda \\ &\quad \theta y_{it} \leqslant Y_t \lambda \quad i = 1,2\ldots,n; \quad t = 1,2\ldots,T \\ &\quad 1'\lambda = 1 \\ &\quad \lambda \geqslant 0 \end{aligned} \tag{1}$$

where $\hat{\theta}_{it}$ and $\hat{D}_{it}$ are the Farrell (1957) and Shepard (1970) distance functions, n is the number of DMUs and T is the number of time periods; $\mathbf{Y}_t$ is a sxn matrix of s outputs, $\mathbf{X}_t$ is a rxn matrix of r inputs, λ represent a $nx1$ vector of weights which allows to obtain a convex combination between inputs and outputs and **1'** is a vector of ones.

Now, $\hat{\theta}_{it}$ is an inefficiency measure and always assumes values equal to or greater than one. Consequently, $\hat{D}_{it}$ is an efficiency measure and it assumes values between zero and one. DMU with an efficiency score equal to one are located on the frontier and as consequence their outputs cannot be further expanded without a corresponding increase in inputs.

In the first stage of this analysis, we assume an output-orientated model with variable return to scale to maximize the outputs that could be produced given the inputs (Ancarani, Di Mauro, & Giammanco, 2008; Barbetta, Turati, & Zago, 2007; Ferrier, Rosko, & Valdmanis, 2006).

The DEA approach offers many strengths: minimum assumptions about the structure of production, flexibility and direct relationship to the economic theory (Coelli et al.,1998).

2.2. The bootstrap DEA

Nevertheless, as discussed by Simar and Wilson (2000), DEA estimator is biased toward unity. In fact, relation (1) does not allow us to determine whether the efficiency values are real, or merely an artifact of the fact that we do not know the true production frontier and must estimate them from a finite sample (Simar & Wilson, 2000). In a context of two-stages procedure as proposed in this paper, the use of biased scores can lead to misleading results as discussed by Simar and Wilson (2007). Consequently bootstrapping techniques, based on the idea that the DGP can be estimated by using the given sample to generate a set of bootstrap samples from which parameters of interest can be calculated, must be used to obtain unbiased results. In the research results, we show what happens in a case study, when the bias is not taken into account.

Following Simar and Wilson (1998), we employ a consistent bootstrap estimation procedure to obtain the sampling distribution of the efficiency scores, and so to correct for the bias. The idea underlying the bootstrap is to approximate the sampling distributions of $\hat{\theta}$, by simulating their DGP. In other terms, given the estimates $\hat{\theta}_{it}$ of the unknown true values of θ_{it} we generate through the DGP process a series of bootstrap estimates $\hat{\theta}^*_{it}$. Thus, for the generic unit i, compute the bias term:

$$BIAS(\hat{\theta}_i) = B^{-1}\sum_{b=1}^{N}\hat{\theta}^*_{i,b} - \hat{\theta}_i, \quad \forall\, i = 1,\ldots,n \tag{2}$$

where $\hat{\theta}^*_{i,b}$ is the bootstrapped technical efficiency and B is the number of bootstrap replications. The bias-corrected estimator of $\hat{\theta}_i$ is:

$$\hat{\theta}^c_i = \hat{\theta}_i - BIAS(\hat{\theta}_i) = 2\hat{\theta}_i - B^{-1}\sum_{b=1}^{B}\hat{\theta}^*_{i,b} \tag{3}$$

The quality of the bootstrap depends on both the number of replications and the sample size (Simar & Wilson, 2000). If the bootstrap is consistent, then:

$$(\hat{\theta}_{it} - \theta_{it})\,\Big|\hat{s} \quad \sim^{approx} \quad (\hat{\theta}^*_{it} - \hat{\theta}_{it})\Big|S^* \quad i = 1,2,\ldots,n \quad t = 1,2,\ldots,T \tag{4}$$

where, $\hat{s}$ and S^* denotes the observed and the bootstrap sample.

In the case study, the results of the model are obtained from 2000 iterations.

2.3. Classification and regression tree (CART)

In the second step, we use a regression type CART where the explanatory variables represent the characteristics of population and health services provided. CART, is a non-parametric statistical procedure for predicting a dependent variable using some explanatory variables (predictors). In particular, the major goal of this methodology is to uncover the predictive structure of the health efficiency in the Italian provinces, creating an accurate dataset. So CART algorithm permits, by binary recursive partitioning, to find through all value of predictors, those minimizes the weighted variance (Razi & Athappilly, 2005). The final tree consists of a root node that includes all the observations, some parent nodes which can be

Download English Version:

https://daneshyari.com/en/article/387643

Download Persian Version:

https://daneshyari.com/article/387643

Daneshyari.com