Randomizing Efficiency Scores in DEA Using Beta Distribution: An Alternative View of Stochastic DEA and Fuzzy DEA

Parakramaweera Sunil Dharmapala, Department of Operations Management and Business Statistics, Sultan Qaboos University, Muscat, Oman

ABSTRACT

Data Envelopment Analysis (DEA) has come under criticism that it is capable of handling only the deterministic input/output data, and therefore, efficiency scores reported by DEA may not be realistic when the data contain random error. Several researchers in the past have addressed this issue by proposing Stochastic DEA models. Some others, citing imprecise data, have proposed Fuzzy DEA models. This paper proposes a method to randomize efficiency scores in DEA by treating each score as an ‘order statistic’ that follows a Beta distribution, and it uses Thompson et al.’s (1996) DEA model appended with Assurance Regions (AR) randomized by our “uniform sampling”. In an application to a set of banks, the work demonstrates the randomization and derives some statistical results.

Keywords: Assurance Regions (AR), Beta Distribution, Data Envelopment Analysis (DEA), Fuzzy DEA, Order Statistics, Stochastic DEA, Uniform Sampling

1. INTRODUCTION

DEA measures relative efficiency of decision making units (DMUs), assuming that input/output data are deterministic, but in reality, these data may be contaminated with statistical noise. While highlighting this as a drawback in DEA methodology, some researchers in the past have incorporated statistical noise into DEA modeling, where they treated input/output data as random variables that follow probability distributions, thus giving birth to Stochastic DEA (SDEA). Yet some other researchers took a different approach by citing imprecise data used in DEA and formulated Fuzzy DEA (FDEA) models. We were inspired by both SDEA and FDEA models reported in the literature. In this paper, we propose a method to randomize efficiency scores by treating each score as an ‘order statistic’ of an underlying Beta distribution. Here, we use Thompson et al.’s (1996) DEA model with assurance regions, randomized by “uniform sampling” that we introduce. Then, we apply this method to a set of banks from India and show the randomization process along with some statistical results and ranking the banks.

DOI: 10.4018/ijban.2014100101
To our knowledge, to date, such a study on randomizing DEA efficiency scores has rarely been reported in the literature.

2. LITERATURE REVIEW

A multitude of research publications have appeared, since the original publication on Data Envelopment Analysis (DEA) by Charnes et al. (1978) in measuring the efficiency of decision making units, and a significant portion of them has been devoted to DEA applications of efficiency in the banking sector. A comprehensive survey of literature on bank efficiency could be found in Fethi and Pasiouras (2010). They have examined bank branch efficiencies in more than 30 studies over the period 1998-2009. All these studies used DEA to estimate bank efficiency. In this paper, we narrow down our literature survey to SDEA and FDEA modeling, as it suits the theme of our discussion.

Several authors who formulated SDEA models treated the input/output vectors as independent and jointly “multivariate normal” random vectors, whose components are expected values of input/outputs. Thus, each input/output of each DMU was treated as a “normally distributed” random variable. Then, the constraints on inputs and outputs (in the traditional DEA model) were expressed in probabilistic terms with the noise parameter (degree of uncertainty) attached to them. Some others treated inputs/outputs as “means” of series of observations and constructed “confidence intervals” for the means.

Banker (1986), Cooper et al. (1998, 2002a, 2004), Land et al. (1993), Despotis and Smirlis (2002), Gstat (1998), Huang and Li (1996, 2001), Olesen and Petersen (1995), Olesen (2006), Kao (2006), and Sengupta (1982, 1987, 1998), are among them. Simar and Wilson (1998) used the “bootstrap method” in Statistics in the sensitivity analysis of efficiency scores. This method goes by creating many samples (bootstrap samples) from an original sample of efficiency scores and constructing interval estimates for the efficiency scores from those samples, and they carry “bootstrap percentiles” reflecting the degree of confidence through high probability. In contrast to all above, within the framework of multivariate probability distributions, Bruni et al. (2009) proposed probabilistically constrained models in DEA with the key assumption that the random variables representing the uncertain data follow a discrete distribution or a discrete approximation of continuous distribution is available.

Some other authors were critical on DEA modelling that the input/output data used in DEA are crisp, but in reality they are imprecise. They have proposed FDEA models that fall into four categories; (i) fuzzy ranking approach, (ii) defuzzification approach, (iii) tolerance approach, and (iv) α-cuts based approach. Sengupta (1992), Meada et al. (1998), Guo and Tanaka (2001), Lertworasirikul et al. (2003), Mohtadi et al. (2002), Jahanshahloo (2009), Zhou et al. (2010) and Zhou et al. (2012a) are among them. We focus our attention on α-cuts based approach and its most recent “generalized FDEA/AR” model proposed by Zhou et al. (2012b), as it is more relevant to our “uniform sampling” of assurance regions.

3. METHODOLOGY

Thompson et al. (1993) presented a new DEA theory, which did not require the use of non-Archimedean principle used in the original DEA theory developed by Charnes et al. (1978), and it allowed zero data entries. In this paper, we use Thompson et al.’s methodology as described below.

A DEA data domain consists of n decision-making units (DMUs), n input vectors (each with m inputs), and n output vectors (each with r outputs). The selected DMU \((c = 1, 2, \ldots, n)\) is characterized by an input vector \(X^c (x^{1c}, x^{2c}, \ldots, x_{mc})\) and an output vector \(Y^c (y^{1c}, y^{2c}, \ldots, y_{rc})\). U-output multiplier of r unknowns \((u_k^c; k = 1, 2, \ldots, r)\) and V-input multiplier of m unknowns \((v_i^c; i = 1, 2, \ldots, m)\) have to be determined by solving the linear programming (LP) model stated below.
A Hybrid Analysis of E-Learning Types and Knowledge Sharing Measurement Indicators: A Model for E-Learning Environments
Davood Qorbani, Iman Raeesi Vanani, Babak Sohrabi and Peter Forte (2016).
Business Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 395-405).
www.igi-global.com/chapter/a-hybrid-analysis-of-e-learning-types-and-knowledge-sharing-measurement-indicators/142630?camid=4v1a