

Randomizing DEA Efficiency Scores with Beta Distribution

R

P. Sunil Dharmapala

College of Economics & Political Science, Sultan Qaboos University, Sultanate of Oman

INTRODUCTION

Since the original publication on Data Envelopment Analysis (DEA) by Charnes et al. (1978), a considerable amount of research publications have appeared in decision science literature, a significant portion of which focusing on efficiency and productivity in the banking sector. A comprehensive survey of bank efficiency studies could be found in Fethi and Pasiouras (2010). They have examined bank branch efficiencies in more than 30 studies over the period 1998-2009.

A shortcoming of DEA modeling that has come under criticism is that input/output data in DEA are treated as deterministic, but in reality, these data are contaminated with statistical noise, and therefore, efficiency score of each decision making unit (DMU) does not reflect reality. In this chapter, we propose a method to randomize efficiency scores by treating each score as an order statistic of an underlying Beta distribution. Then, we apply this method to a small set of banks from Sri Lanka and show how to do the randomization.

To our knowledge, to date, very few authors have discussed similar randomization in an efficiency study.

LITERATURE REVIEW

A comprehensive listing and analysis of DEA research that covers the first 30 years of its history could be found in Emrouznejad et al. (2008). In this chapter, we narrow down our literature survey to stochastic DEA modeling, as it suits the theme of discussion of the chapter.

DEA measures relative efficiency of decision making units (DMUs), assuming that input/output data are deterministic. While highlighting this as a drawback in DEA methodology, some authors in the past incorporated statistical noise into DEA modelling and berthed Stochastic DEA (SDEA). Some of them 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.

Cooper et al. (1998, 2002a, 2004), Despotis and Smirlis (2002), Gstach (1998), Huang and Li (1996), Kao (2006), Olesen and Petersen (1995), and Sengupta (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, with “bootstrap percentiles” reflecting the degree of confidence through high probability.

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

DOI: 10.4018/978-1-4666-5202-6.ch178

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} ($c = 1, 2, \dots, n$) is characterized by an input vector X_c ($x_{1c}, x_{2c}, \dots, x_{mc}$) and an output vector Y_c ($y_{1c}, y_{2c}, \dots, y_{rc}$). U -output multiplier of r unknowns (u_k ; $k = 1, 2, \dots, r$) and V -input multiplier of m unknowns (v_i ; $i = 1, 2, \dots, m$) have to be determined by solving the respective linear programming (LP) models stated below.

Here we consider the following four inputs ($m = 4$) and two outputs ($r = 2$) for the banks:

- X₁:** Total deposits include demand deposits, time and savings deposits, CDs, and purchased funds
- X₂:** Fixed assets in terms of bank premises, furniture, and equipment
- X₃:** Total non-interest expenses include employee salaries, benefits, and expenses on fixed assets
- X₄:** Loan loss provisions. Accounting allocations to cover possible loan defaults
- Y₁:** Total loans include commercial, industrial, real-estate, and installment loans
- Y₂:** Non-interest income includes commissions, securities, and service charges

The two outputs capture the total earnings (interest and non-interest income) for the banks, while the four inputs capture the total borrowings and operating expenses.

CCR/AR and BCC/AR Efficiency Models

We recall the input-oriented CCR/AR (Charnes-Cooper-Rhodes with Assurance Regions; Constant returns) and BCC/AR (Banker-Charnes-Cooper with Assurance Regions; Variable returns) ef-

iciency models as formulated by Thompson et al. (1996).

CCR/AR Model

$$\text{Max } Z = \sum_{k=1}^r u_k y_{kc} \quad (\text{for DMU}_c)$$

$$\sum_{i=1}^m v_i x_{ic} = 1 \quad (1)$$

$$\sum_{k=1}^r u_k y_{kj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j = 1, 2, \dots, n \quad (2)$$

$$\alpha_i v_j < v_i < \beta_i v_j \quad j < i; j = 1, \dots, m-1; \\ i = 2, \dots, m \quad (\text{input cone}) \quad (3)$$

$$\gamma_k u_j < u_k < \delta_k u_j \quad j < k; j = 1, \dots, r-1; \\ k = 2, \dots, r \quad (\text{output cone}) \quad (4)$$

The non-negative scalars $\alpha_i, \beta_i, \gamma_k, \delta_k$, in constraints (3) and (4) are specified or estimated using socio-economic/market data and/or expert opinion. They are also called "price/cost" data. As illustrated in Thompson et al. (1996), the input cone has $2^{(mC_2)}$ constraints and the output cone has $2^{(rC_2)}$ constraints. In our study, with $m = 4$ and $r = 2$, input cone has 12, and output cone has 2 constraints.

In computing $\alpha_i, \beta_i, \gamma_k, \delta_k$, first we use the bounds on v_i and u_k :

$$\text{Input multipliers: } LV_i \leq v_i \leq UV_i \\ i = 1, 2, \dots, m \quad (5)$$

$$\text{Output multipliers: } LU_k \leq u_k \leq UU_k \\ k = 1, 2, \dots, r \quad (6)$$

Notice that LV is lower and UV is upper bound for v_i , and LU is lower and UU is upper bound for u_k . Next, we compute the non-negative scalars using (5) and (6).

$$\alpha_i = LV_i / UV_j, j < i; j = 1, \dots, m-1; \\ i = 2, \dots, m \quad (\text{for input cone}) \quad (7)$$

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/randomizing-dea-efficiency-scores-with-beta-distribution/107386

Related Content

Cyber Physical Systems: A Review

Siddhartha Khaitanand James D. McCalley (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 574-579).

www.irma-international.org/chapter/cyber-physical-systems/107260

Social Media Tools for Quality Business Information

Sérgio Maravilhas (2014). *Information Quality and Governance for Business Intelligence* (pp. 167-193).

www.irma-international.org/chapter/social-media-tools-for-quality-business-information/96150

The Construction of Economic Intelligence as an Academic Discipline: A Theoretical Perspective

Florian Fougy (2014). *International Journal of Business Intelligence Research* (pp. 1-16).

www.irma-international.org/article/the-construction-of-economic-intelligence-as-an-academic-discipline/120049

Knowledge Generation Using Sentiment Classification Involving Machine Learning on E-Commerce

Swarup Kr Ghosh, Sowvik Deyand Anupam Ghosh (2019). *International Journal of Business Analytics* (pp. 74-90).

www.irma-international.org/article/knowledge-generation-using-sentiment-classification-involving-machine-learning-on-e-commerce/226973

Comprehensive Study and Analysis of Partitional Data Clustering Techniques

Aparna K.and Mydhili K. Nair (2015). *International Journal of Business Analytics* (pp. 23-38).

www.irma-international.org/article/comprehensive-study-and-analysis-of-partitional-data-clustering-techniques/124180