

# Data Envelopment Analysis Using R

Nikita Singam

*Department of Mathematics, Chandigarh University*

\*\*\*

**Abstract:** The purpose of this paper is to estimate the technology and quantifying efficiency by the use of Data Envelopment Analysis (DEA) using the Benchmarking package in R programming language. In this paper, a number of concerns related the use of DEA are discussed. Model orientation, input and output selection/definition, mixed and raw data utilisation, and the number of inputs and outputs to utilise vs the number of decision making units (DMUs) are all problems to consider. While the DEA frontier can be seen as a production frontier, it is important to remember that DEA is ultimately a tool for evaluating performance and benchmarking against best-practice.

**Keywords:** Data Envelopment Analysis, Decision Making Units, Benchmarking, etc..

## 1. Introduction

Data Envelopment Analysis is a methodology focused on linear programming for determining the production quality of corporate units referred to as Decision-Making Units (DMUs). The data envelopment analysis (DEA) technique is used to determine a set of decision-maker units' strengths and weaknesses. Linear programming, nonlinear programming, and hybrid planning are all types of problems that are used to accomplish these objectives. In such problems, the number of constraints or factors is equivalent to the number of DMUs being evaluated, but the number of constraints multiplies as the problem layout varies from box to network. It has been effectively used to compare the efficiency of a set of companies that use a variety of similar inputs to generate a variety of identical outputs.[5] [2]

The data envelopment analysis technique is used to determine the relative efficiency of a collection of disparate decision-making units (DMUs). Using real-world observations of the community (n is the number of observable DMUs), this approach creates a collection of output possibilities, this can be used to see whether there is a better point than the DMU currently being considered.

## 2. Efficiency Measurement

The ratio of total outputs to total inputs is the fundamental performance measure in DEA. The

$$\text{Efficiency} = \frac{\text{Output}}{\text{Input}}$$

$$\text{Input Target} = \text{Actual Input} \cdot \text{Relative Efficiency}/100$$

Input Value would be less than the Actual Input Value. Input Slack is the difference between the Actual input and the input goal.

$$\text{Input Slack} = \text{Actual Input} - \text{Input Target}$$

$$\text{Input Slack Percentage} = \frac{\text{Input Slack}}{\text{Actual Input}} \times 100$$

Consequently, input slack can be calculated as a percentage.

Other entities' input targets and slacks may be calculated accordingly. Human beings can compute Output Targets and Output Slacks using a common logic.

$$\text{Output Target} = \frac{\text{Actual Output}}{\text{Relative Efficiency}/100}$$

$$\text{Output Slack} = \text{Output Target} - \text{Actual Output}$$

$$\text{Output Slack Percentage} = \frac{\text{Output Slack}}{\text{Actual Output}} \times 100$$

### 3. Mathematical Formulation

Many inputs and outputs are aggregated linearly in DEA using weights. Thus, a firm's virtual supply is the linear weighted number of all its inputs.

$$\text{Virtual Input} = \sum_{i=1}^I u_i x_i,$$

where,  $u_i$  denotes the weight applied to input  $x_i$  during aggregation and  $u_i$ .

Likewise, a firm's virtual performance is calculated as the linear weighted total of all its outputs.

$$\text{Virtual Output} = \sum_{j=1}^J v_j y_j,$$

where,  $v_j$  is the aggregation weight applied to output  $y_j$ . Also  $v_j$ .

The efficiency of the DMU in translating inputs to outputs can be described as the ratio of outputs to inputs given these virtual inputs and outputs.

$$\text{Efficiency} = \frac{\text{Virtual Output}}{\text{Virtual Input}} = \frac{\sum_{j=1}^J v_j y_j}{\sum_{i=1}^I u_i x_i}$$

### 4. Fractional DEA Programs

Imagine a collection of  $N$  DMUs whose efficiencies must be contrasted. Considering one of the DMUs, tell the  $m$ th DMU, and increase its utility using the above formula. The  $m$ th DMU is used as the reference DMU in this case.

$$\max E_m = \frac{\sum_{j=1}^J v_{jm} y_{jm}}{\sum_{i=1}^I u_{im} x_{im}}$$

subject to

$$0 \leq \frac{\sum_{j=1}^J v_{jm} y_{jn}}{\sum_{i=1}^I u_{im} x_{in}} \leq 1; \quad n = 1, 2, K, N$$

$$v_{jm}, u_{im} \geq 0; \quad i = 1, 2, K, I; \quad j = 1, 2, K, J$$

where,

$E_m$  denotes the performance of the  $m$ th DMU,  $y_{jm}$  denotes the  $j$ th output of the  $m$ th DMU,  $v_{jm}$  denotes the weight of that output,  $x_{im}$  denotes the  $i$ th input of the  $m$ th DMU,  $u_{im}$  denotes the weight of that input, and  $y_{jn}$  and  $x_{in}$  denote the  $j$ th output and  $i$ th input, respectively, of the  $n$ th

DMU,  $n = 1, 2, 3, \dots, N$ .

## 5. Time Series Analysis Using DEA

Typically, this kind of analysis is referred to as a cross-sectional analysis. In comparison, one might consider measuring DMU efficiency over time. This is often referred to as a time series study. In practice, DMUs are observed over a span of time; the differences in DMU performance over time may aid in drawing significant conclusions. There are at least two methods for using DEA in a time series mode: Window Analysis and the Malmquist Productivity Index.

## 6. Analysis of DEA using R language

In this paper, **Benchmarking** package is used. The Benchmarking package includes tools for estimating technology and quantifying efficiency by the use of Data Envelopment Analysis (DEA) and Stochastic Frontier Functions (SFA). Data Envelopment Analysis (DEA) is provided under a variety of technological assumptions (fdh, vrs, drs, crs, irs, and add) and with a variety of performance metrics (input based, output based, additive, super, directional). Peers should be included, and the optimum expense, sales, and benefit can be measured. Comparative approaches for estimating the efficiencies of stochastic frontier functions (SFAs) and convex nonparametric least squares for convex functions (StoNED) are also used.[1]

### 6.1 Analysis

**Description** A DEA technology graph is drawn. Two goods illustrations, either isoquant (2 inputs), transformation curve (2 outputs), or a production function (1 input and 1 output) are designed. [3]

**Usage** `dea.plot(x, y, RTS="vrs", ORIENTATION="in-out", txt=NULL, add=FALSE, wx=NULL, wy=NULL, TRANSPOSE=FALSE, fex=1, GRID=FALSE, RANGE=FALSE, param=NULL, ..., xlim, ylim, xlab, ylab)`

`dea.plot.frontier(x, y, RTS="vrs", ...)`

`dea.plot.isoquant(x1, x2, RTS="vrs", ...)`

`dea.plot.transform(y1, y2, RTS="vrs", ...)`

**Method** The method `dea.plot` is a generic plotting technique; the other 3 are specialised versions for frontiers, which have 1 input and 1 output, isoquant curves, which have 2 inputs for given outputs, and transformation curves, which have 2 outputs and 2 inputs.

### Arguments

`x, y, x1, y1, x2, y2` - the good for the first and second axis

RTS - assumptions for returns to scales

`txt` - an array to label the observations

`add` - 'add=T' means on top of the existing graph the technology is drawn and 'add=F' means a default new graph is made

`wx` and `wy` - Weight to aggregate for the first axis and the second axis

TRANSDPOSE - for each axis, only relevant if there are more than one input

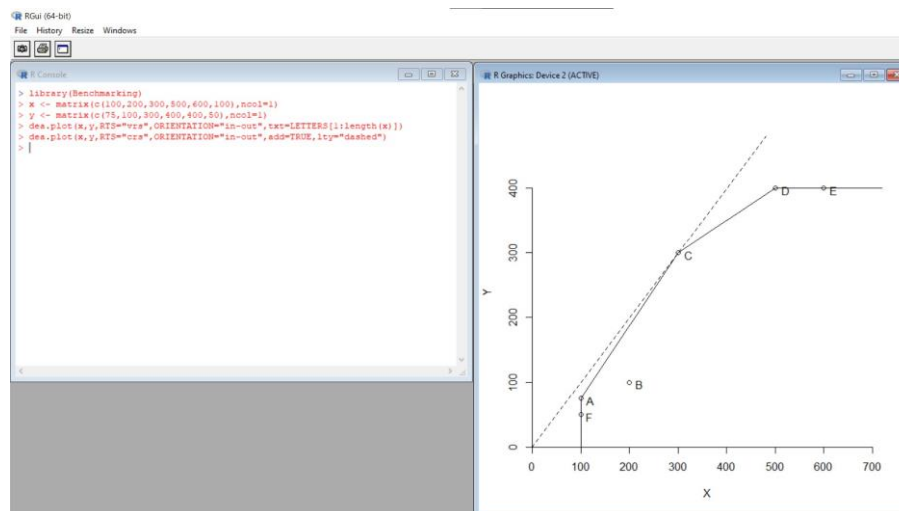
`...` - the methods usual options such `plot`, `lines`, and `abline` etc.

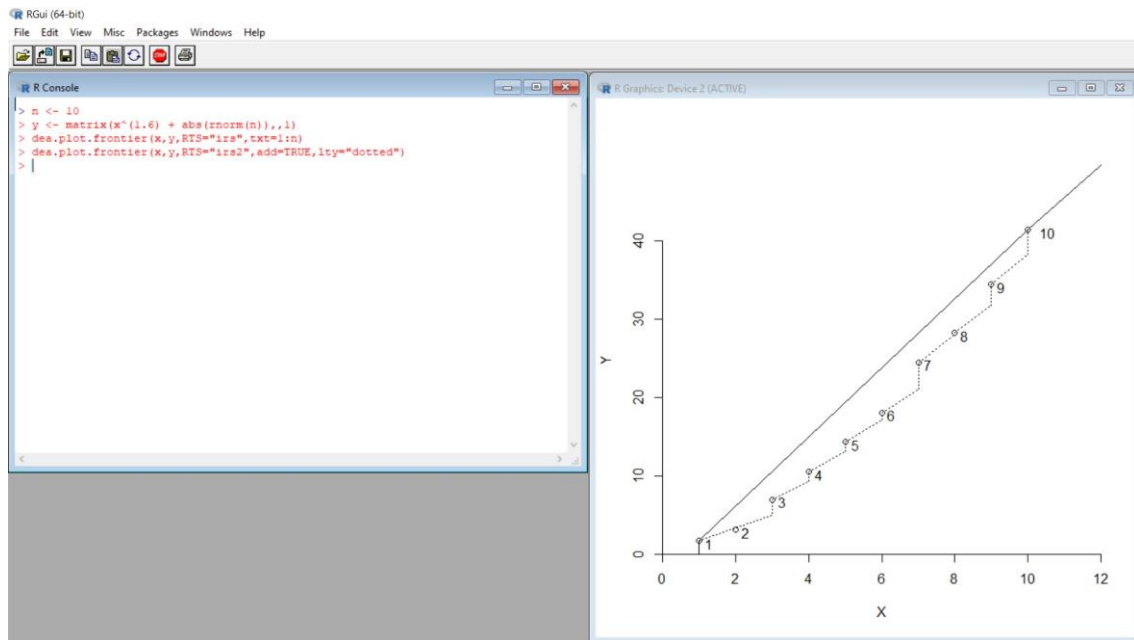
`fex` - the relative size of text on observations is the same as 'cex', except it simply affects the size

RANGE - a logical variable

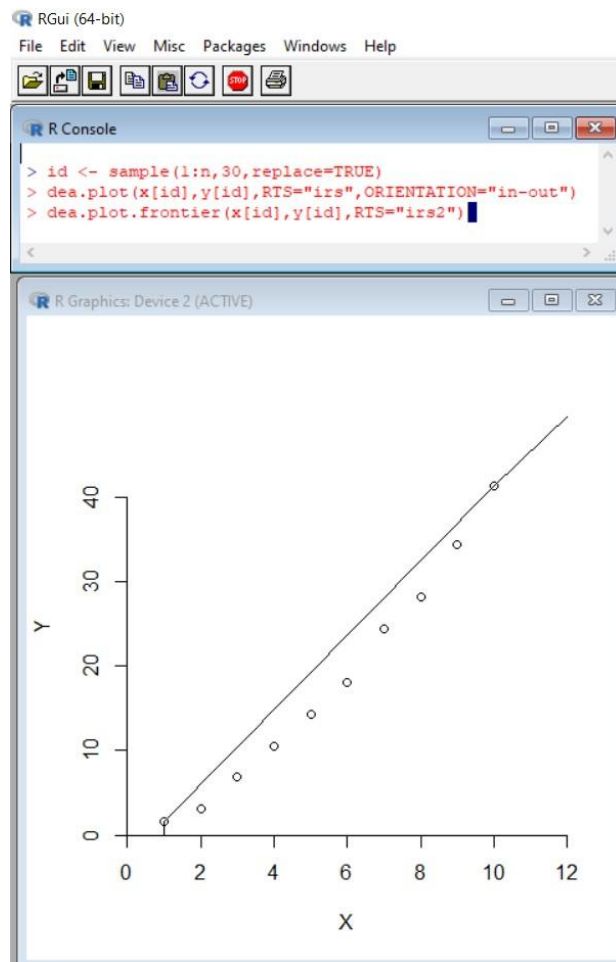
`xlim, ylim` - possible limits for the first axis and the second axis

`xlab, ylab` - possible label for the x-axis and y-axis

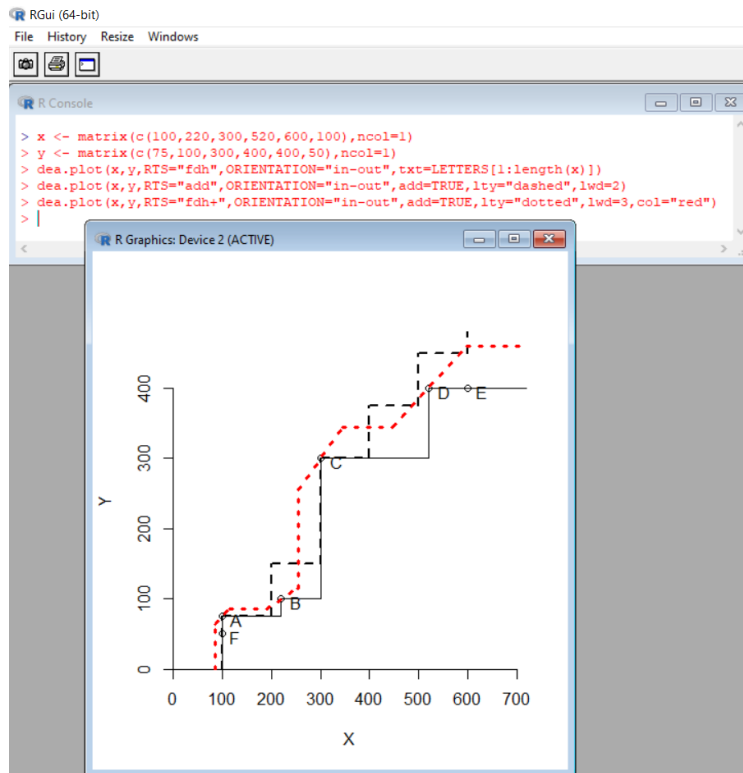




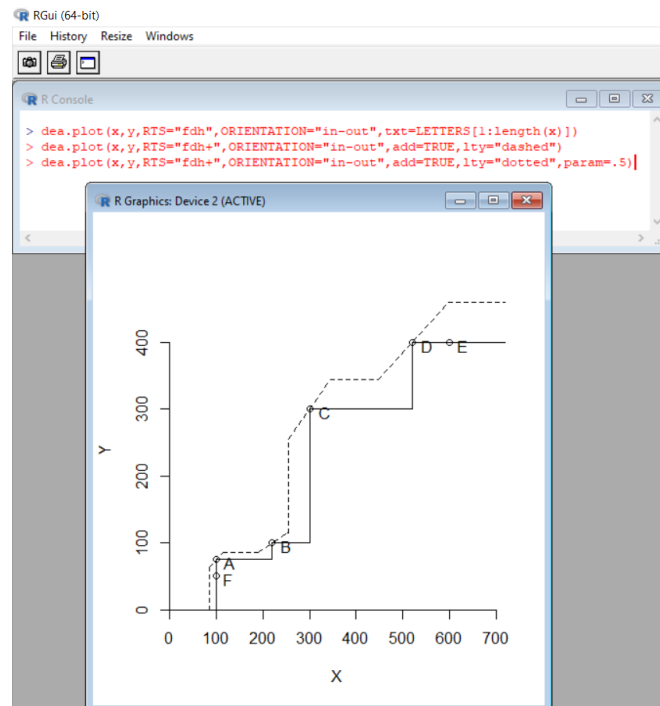
The below example tells the two different forms of irs: irs and irs2, and two different ways to make a frontier. The terms irs means increasing returns to scale. [6]



The below picture tells difference between the FDH technology and the additive FRH technology. FDH stands for free disposal hull. FDH is formulated as a series of mixed integer primal programs and a series of linear dual program. [4]



Using the parameter in FDH+



## 7. Conclusion

In this paper, R programming is used to analyse the Data Envelopment Analysis for plotting two different technologies, and designed a production function using parameters. And using software programming is more efficient to calculate DEA.

## Acknowledgement

I would like to show my special thanks of gratitude to my Professor Dr. Sandeep Mogha as well as our Department of Mathematics, Chandigarh University who gave me the golden opportunity to do this paper, this also encouraged me in conducting extensive research and learning about a variety of new topics.

## References

- [1] Peter Bogetoft, Lars Otto, and Maintainer Lars Otto. Package 'benchmarking'. Retrieved January, 4:2016, 2015.
- [2] Wade D Cook, Kaoru Tone, and Joe Zhu. Data envelopment analysis: Prior to choosing a model. *Omega*, 44:1-4, 2014.
- [3] DataCamp.  
<https://www.rdocumentation.org/packages/Benchmarking/versions/0.10/topics/dea.plot>  
Rdocumentation.
- [4] CA Knox Lovell and Philippe Vanden Eeckaut. Frontier tales: Dea and fdh. In *Mathematical Modelling in Economics*, pages 446-457. Springer, 1993.
- [5] Ramu Ramanathan. *An introduction to data envelopment analysis: a tool for performance measurement*. Sage, 2003.
- [6] Lawrence M Seiford and Joe Zhu. An investigation of returns to scale in data envelopment analysis. *Omega*, 27(1):1-11, 1999.