iDEAs
Samsun, Turkey -- June 27 to 30, 2013
11th International Conference on Data Envelopment Analysis

Implementing DEA models in the R program

José Francisco Moreira Pessanha
Rio de Janeiro State University - UERJ
professorjfmp@hotmail.com

Alexandre Marinho
Institute for Applied Economic Research - IPEA
Rio de Janeiro State University - UERJ
alexandre.marinho@ipea.gov.br

Luiz da Costa Laurencel
Rio de Janeiro State University - UERJ
llaurenc.ntg@terra.com.br

Marcelo Rubens dos Santos do Amaral
Brazilian Institute of Geography and Statistics - IBGE
Rio de Janeiro State University - UERJ
mrubens@ime.uerj.br

Objective

This work aims to present a brief introduction about the implementation of Data Envelopment Analysis (DEA) classical models in the R program.

The models implemented include the DEA model with constant returns to scale (CRS) and the model with variable returns to scale (VRS), both in the multipliers forms and input oriented.

The computational implementation of DEA models is illustrated by the efficiency evaluation of the 18 biggest Brazilian electric power distribution utilities.

Data Envelopment Analysis

Introduced by Charnes, Cooper and Rhodes in 1978, the Data Envelopment Analysis is an important branch of operations research, as well as of economics as evidenced by numerous publications with practical applications and theoretical developments on little more than three decades (COOK & SEIFORD, 2009).

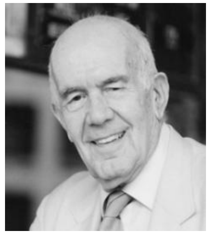
DEA can be described as a nonparametric technique based on linear programming to evaluate the efficiency of organizations working in the same industry, for example, schools, banks, factories and utilities.

Currently, the DEA has a variety of models ranging from classic models (CRS and VRS) and its variations to approaches that combine DEA models with other methods like as bootstrap (COELLI et al, 2005) and fuzzy logic (GUO and TANAKA, 2001).

Along with theoretical advances, we see the evolution of softwares dedicated to the DEA.

Data Envelopment Analysis

Abraham **Charnes**



William W. **Cooper**

Edwardo Lao **Rhodes**

CHARNES, A., COOPER, W. W. & RHODES, E. Measuring the efficiency of decision making units, *European Journal of Operational Research*, *Volume 2*, *Issue 6*, *November 1978*, pp. *429-444*

CRS model , input oriented

Multiplier form

$$efficiency = \underset{u,v}{Max} \sum_{i=1}^{m} u_i y_{i,j0}$$

$s.t.$

$$-\sum_{i=1}^{s} v_i x_{ij} + \sum_{i=1}^{m} u_i y_{ij} \le 0 \quad \forall j = 1,\dots, j0,\dots, N$$

$$\sum_{i=1}^{s} v_i x_{i,j0} = 1$$

$$u_i \ge 0 \quad \forall i = 1, m$$

$$v_i \ge 0 \quad \forall i = 1, s$$

Envelopment form

$$efficiency = \underset{\lambda,\theta}{Min} \quad \theta$$

$s.t.$

$$\theta X_{j0} \ge \sum_{j=1}^{N} \lambda_j X_j$$

$$Y_{j0} \le \sum_{j=1}^{N} \lambda_j Y_j$$

$$\lambda_j \ge 0 \forall j = 1,\dots, j0,\dots, N$$

$$\theta \ge 0$$

Data Envelopment Analysis

Rajiv D. **Banker**

BANKER, R.D., CHARNES, A. & COOPER, W. W. Some models for estimating technical scale inefficiencies in Data Envelopment Analysis, Management Scxience, v. 30, n. 9, pp. 1078-1092, 1984

VRS model , input oriented

Multiplier form

$$efficiency = \underset{u,v}{Max} \sum_{i=1}^{m} u_i y_{i,j0} + u_0$$

$$s.t.$$

$$-\sum_{i=1}^{s} v_i x_{ij} + \sum_{i=1}^{m} u_i y_{ij} + u_0 \leq 0 \quad \forall j = 1,\ldots,j0,\ldots,N$$

$$\sum_{i=1}^{s} v_i x_{i,j0} = 1$$

$$u_i \geq 0 \quad \forall i = 1, m$$

$$v_i \geq 0 \quad \forall i = 1, s$$

Envelopment form

$$efficiency = \underset{\lambda,\theta}{Min} \quad \theta$$

$$s.t.$$

$$\theta X_{j0} \geq \sum_{j=1}^{N} \lambda_j X_j$$

$$Y_{j0} \leq \sum_{j=1}^{N} \lambda_j Y_j$$

$$\sum_{j=1}^{N} \lambda_j = 1$$

$$\lambda_j \geq 0 \forall j = 1,\ldots, j0,\ldots, N$$

$$\theta \geq 0$$

R project

Language and computational environment to make statistical analyzes and data mining.

It's free and open source.

Provides a variety of functions for statistical analysis (linear and nonlinear regression, statistical tests, time series analysis temporal, multivariate statistics, design of experiments, etc.).

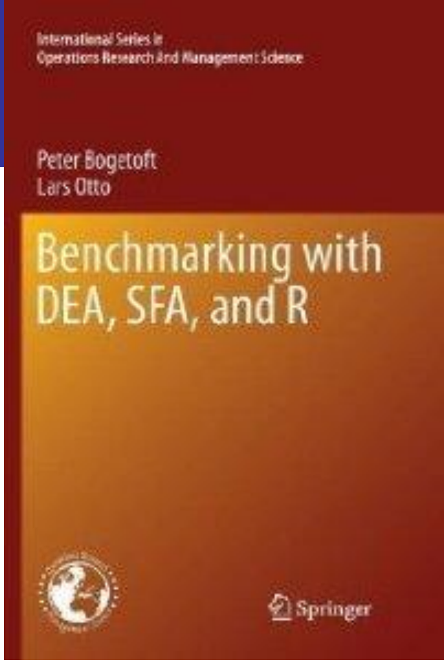
Provides functions for the development of various types of graphs, useful in exploratory data analysis and data visualization.

It is highly extensible.

Rapid diffusion (2 million users worldwide).

http://www.r-project.org/

DEA in R program



Packages dedicated to DEA models:

- **Benchmarking (BOGETOFT & OTTO, 2011)**
- **FEAR (Frontier Efficiency Analysis with R)**

http://www.clemson.edu/economics/faculty/wilson/Software/FEAR/fear.html

However, the R program is more than a library package.

R allows analysts to build their own programs or packages and distribute them.

Thus, using the R program analysts can obtain low-cost solutions.

Although commercial and freeware programs are practical and contain many templates and resources to facilitate the implementation of the DEA, the possibility of implementing DEA models in a spreadsheet or any other programming language, is interesting because it provides great flexibility to the analyst in the application of the models, research and innovations.

Case study - data

Consider the data of the <u>18 biggest Brazilian electricity distribution utilities</u> for the year 2009, where

- **OPEX denotes the annual operating expenditures (R$).**
- **NETWORK is the total length (in km) of the distribution network.**
- **MWH represents the total electricity consumption in each utility.**
- **CUSTOMERS is the number of customers supplied by the utility.**

	A	B	C	D	E
1	UTILITY	OPEX	NETWORK	MWH	CUSTOMERS
2	ELETROPAULO	1,249,143,613	45,213	39,922,710	5,987,873
3	CEMIG	1,682,334,644	460,219	37,476,802	6,832,546
4	CPFL - Paulista	497,290,782	89,879	25,267,579	3,502,793
5	COPEL	1,018,866,491	224,817	23,525,040	3,628,209
6	LIGHT	557,206,112	58,074	22,902,552	3,640,182
7	CELESC	721,455,274	144,896	18,105,811	2,237,127
8	COELBA	436,436,014	215,001	14,286,757	4,622,046
9	ELEKTRO	414,602,018	107,116	13,398,558	2,123,670
10	CPFL - Piratininga	195,789,961	22,236	13,013,378	1,367,488
11	BANDEIRANTE	286,832,273	27,496	12,536,237	1,482,518
12	CELPE	350,651,684	120,428	10,001,560	2,994,259
13	AMPLA	436,532,756	51,050	9,506,961	2,365,558
14	CELG	691,472,253	199,494	9,344,291	2,213,198
15	RGE	186,357,415	84,997	7,993,103	1,226,079
16	COELCE	316,166,876	120,300	7,929,212	2,744,830
17	ESCELSA	241,433,335	56,960	7,897,969	1,185,432
18	AES SUL	206,122,962	76,133	7,616,460	1,150,518
19	CEEE	385,990,997	71,892	7,277,929	1,438,072

Source: Aneel - Brazilian Electricity Regulatory Agency

Case study – data loading

Assuming that the data is stored in a spreadsheet called *data.xls*, located in the directory *c:\example*, the data import can be done by the following commands (commentaries after #):

```
require(xlsReadWrite)  # load xlsReadWrite package
setwd("c:/exemple") # set work directory
data<-read.xls('data.xls',colNames=TRUE,sheet=1, type="data.frame",from=1) # import data
attach(data) # add variable names to the data matriz columns
```

In the R code above

- ***colnames = TRUE*** **indicates that the first line contains the variable names**
- ***sheet = 1*** **indicates that the data is on the first sheet at *data.xls* and should be read from cell A1 (*from = 1*)**
- **the option *type* indicates the object is a data.frame (VERZANI, 2005).**

An alternative to data loading

```
require(xlsx) # load xlsx package
setwd("c:/exemple")
data <- read.xlsx("c:/exemple/data.xls", 1) # import data
```

Case study – data loading

```
data <- read.xls("data.xls",colNames =TRUE,sheet=1, type='data.frame',from=1)
```

or

```
data <- read.xlsx("c:/exemple/data.xls", 1)
```

Excel spreadsheet data.xls

R object data

	A	B	C	D	E
1	UTILITY	OPEX	NETWORK	MWH	CUSTOMERS
2	ELETROPAULO	1,249,143,613	45,213	39,922,710	5,987,873
3	CEMIG	1,682,334,644	460,219	37,476,802	6,832,546
4	CPFL - Paulista	497,290,782	89,879	25,267,579	3,502,793
5	COPEL	1,018,866,491	224,817	23,525,040	3,628,209
6	LIGHT	557,206,112	58,074	22,902,552	3,640,182
7	CELESC	721,455,274	144,896	18,105,811	2,237,127
8	COELBA	436,436,014	215,001	14,286,757	4,622,046
9	ELEKTRO	414,602,018	107,116	13,398,558	2,123,670
10	CPFL - Piratininga	195,789,961	22,236	13,013,378	1,367,488
11	BANDEIRANTE	286,832,273	27,496	12,536,237	1,482,518
12	CELPE	350,651,684	120,428	10,001,560	2,994,259
13	AMPLA	436,532,756	51,050	9,506,961	2,365,558
14	CELG	691,472,253	199,494	9,344,291	2,213,198
15	RGE	186,357,415	84,997	7,993,103	1,226,079
16	COELCE	316,166,876	120,300	7,929,212	2,744,830
17	ESCELSA	241,433,335	56,960	7,897,969	1,185,432
18	AES SUL	206,122,962	76,133	7,616,460	1,150,518
19	CEEE	385,990,997	71,892	7,277,929	1,438,072

```
> data
              UTILITY       OPEX   NETWORK      MWH CUSTOMERS
1         ELETROPAULO 1249143613  45212.99 39922710   5987873
2               CEMIG 1682334644 460219.00 37476802   6832546
3     CPFL - Paulista  497290782  89879.00 25267579   3502793
4               COPEL 1018866491 224817.29 23525040   3628209
5               LIGHT  557206112  58074.00 22902552   3640182
6              CELESC  721455274 144896.32 18105811   2237127
7              COELBA  436436014 215001.47 14286757   4622046
8             ELEKTRO  414602018 107115.75 13398558   2123670
9  CPFL - Piratininga  195789961  22235.63 13013378   1367488
10        BANDEIRANTE  286832273  27496.38 12536237   1482518
11              CELPE  350651684 120427.84 10001560   2994259
12              AMPLA  436532756  51050.29  9506961   2365558
13               CELG  691472253 199494.10  9344291   2213198
14                RGE  186357415  84996.52  7993103   1226079
15             COELCE  316166876 120299.97  7929212   2744830
16            ESCELSA  241433335  56959.90  7897969   1185432
17            AES SUL  206122962  76133.22  7616460   1150518
18               CEEE  385990997  71892.26  7277929   1438072
```

Case study – inputs and outputs variables

The main outputs of the distribution utilities are the amount of distributed energy (MWH) and the number of customer (CUSTOMERS).

The operating expenses are also influenced by non controllable factors, for example, the dispersion of consumers and geographical characteristics of the concession area. To address these issues the size of the distribution network (NETWORK) can also be included as an variable output.

The outputs variables are the cost drivers of a distribution utility.

For a given level of output, the utility must produce at the lowest cost.

In order to obtain an efficiency score that quantifies the potential reduction in operating costs, the Brazilian Electricity Regulatory Agency (ANEEL) proposes a DEA model orientated to input, wherein the OPEX is the unique input variable and with the same three outputs: CUSTOMERS, MWH and NETWORK.

Case study – inputs and outputs variables

```
> data
            UTILITY       OPEX   NETWORK      MWH CUSTOMERS
1       ELETROPAULO 1249143613  45212.99 39922710   5987873
2             CEMIG 1682334644 460219.00 37476802   6832546
3    CPFL - Paulista  497290782  89879.00 25267579   3502793
4             COPEL 1018866491 224817.29 23525040   3628209
5             LIGHT  557206112  58074.00 22902552   3640182
6            CELESC  721455274 144896.32 18105811   2237127
7            COELBA  436436014 215001.47 14286757   4622046
8           ELEKTRO  414602018 107115.75 13398558   2123670
9  CPFL - Piratininga  195789961  22235.63 13013378   1367488
10       BANDEIRANTE  286832273  27496.38 12536237   1482518
11             CELPE  350651684 120427.84 10001560   2994259
12             AMPLA  436532756  51050.29  9506961   2365558
13              CELG  691472253 199494.10  9344291   2213198
14               RGE  186357415  84996.52  7993103   1226079
15            COELCE  316166876 120299.97  7929212   2744830
16           ESCELSA  241433335  56959.90  7897969   1185432
17           AES SUL  206122962  76133.22  7616460   1150518
18              CEEE  385990997  71892.26  7277929   1438072
```

The input variable is the OPEX located at second column of the data matrix.

The output variables are at columns 3 (NETWORK), 4 (MWH) and 5 (CUSTOMERS) of the data matrix.

The selection of inputs and outputs variables can be done by the following commands:

```
inputs <- data.frame(data[2]) # input variable at second column of the data matrix
outputs<-data.frame(data[c(3,4,5)]) # output variables
N <- dim(data)[1] # the number of DMUs is equal to number of rows of data matrix
s <- dim(inputs)[2] # number of input variables, in this case s = 1
m <- dim(outputs)[2] # number of output variables, in this case m = 3
```

Case study – inputs and outputs variables

```
inputs <– data.frame(data[2])
outputs<–data.frame(data[c(3,4,5)])
```

> data

	UTILITY	OPEX	NETWORK	MWH	CUSTOMERS
1	ELETROPAULO	1249143613	45212.99	39922710	5987873
2	CEMIG	1682334644	460219.00	37476802	6832546
3	CPFL - Paulista	497290782	89879.00	25267579	3502793
4	COPEL	1018866491	224817.29	23525040	3628209
5	LIGHT	557206112	58074.00	22902552	3640182
6	CELESC	721455274	144896.32	18105811	2237127
7	COELBA	436436014	215001.47	14286757	4622046
8	ELEKTRO	414602018	107115.75	13398558	2123670
9	CPFL - Piratininga	195789961	22235.63	13013378	1367488
10	BANDEIRANTE	286832273	27496.38	12536237	1482518
11	CELPE	350651684	120427.84	10001560	2994259
12	AMPLA	436532756	51050.29	9506961	2365558
13	CELG	691472253	199494.10	9344291	2213198
14	RGE	186357415	84996.52	7993103	1226079
15	COELCE	316166876	120299.97	7929212	2744830
16	ESCELSA	241433335	56959.90	7897969	1185432
17	AES SUL	206122962	76133.22	7616460	1150518
18	CEEE	385990997	71892.26	7277929	1438072

> inputs

	OPEX
1	1249143613
2	1682334644
3	497290782
4	1018866491
5	557206112
6	721455274
7	436436014
8	414602018
9	195789961
10	286832273
11	350651684
12	436532756
13	691472253
14	186357415
15	316166876
16	241433335
17	206122962
18	385990997

18 x 1
(s=1)

> outputs

	NETWORK	MWH	CUSTOMERS
1	45212.99	39922710	5987873
2	460219.00	37476802	6832546
3	89879.00	25267579	3502793
4	224817.29	23525040	3628209
5	58074.00	22902552	3640182
6	144896.32	18105811	2237127
7	215001.47	14286757	4622046
8	107115.75	13398558	2123670
9	22235.63	13013378	1367488
10	27496.38	12536237	1482518
11	120427.84	10001560	2994259
12	51050.29	9506961	2365558
13	199494.10	9344291	2213198
14	84996.52	7993103	1226079
15	120299.97	7929212	2744830
16	56959.90	7897969	1185432
17	76133.22	7616460	1150518
18	71892.26	7277929	1438072

18 x 3
(m=3)

Case study – CRS model

CRS model, input oriented in the multiplier form

c^T = [0 *outputs*]
outputs of the
DMU evaluated j0

$$\underset{u,v}{Max} \sum_{i=1}^{m} u_i y_{i,j0}$$

multipliers
$z^T = [v \;\; u_1 \;\; u_2 \;\; u_3]$

$$\begin{aligned} &Max && c^T \cdot z && \\ &s.t. && [-inputs \quad outputs] \cdot z && \le 0 \\ & && b^T \cdot z && = 1 \\ & && z && \ge 0 \end{aligned}$$

$$-\sum_{i=1}^{s} v_i x_{ij} + \sum_{i=1}^{m} u_i y_{ij} \le 0 \quad \forall j = 1,\ldots, j0,\ldots, N$$

$$\sum_{i=1}^{s} v_i x_{i,j0} = 1$$

b^T = [*inputs* 0]
inputs of the DMU
evaluated j0

Case study – CRS model

Writing the DEA model in R program

f.obj<−c(rep(0,1,s),as.numeric(outputs[i,]))

vector with the objective function coefficients (outputs) of the i-th DMU

aux <- cbind(-1*inputs, outputs)

matrix [-inputs outputs]

$$
\begin{aligned}
Max \quad & c^T \cdot z \\
s.t. \quad & \begin{bmatrix} -inputs & outputs \end{bmatrix} \cdot z \le 0 \\
& b^T \cdot z = 1 \\
& z \ge 0
\end{aligned}
$$

f.rhs <− c(rep(0,1,N),1)

c(as.numeric(inputs[i,]),rep(0,1,m))

vector with the inputs of the i-th DMU

f.dir<−c(rep("<=",1,N),"=")

Case study – CRS model

Writing the DEA model in R program

f.obj<−c(rep(0,1,s),as.numeric(outputs[i,]))

vector with the objective function coefficients

aux <- cbind(-1*inputs, outputs)

matrix [-inputs outputs]

$$
\begin{aligned}
Max \quad & c^T \cdot z \\
s.t. \quad & \left[-inputs \quad outputs \right] \cdot z \le 0 \\
& b^T \cdot z = 1 \\
& z \ge 0
\end{aligned}
$$

f.rhs <− c(rep(0,1,N),1)

f.dir<−c(rep("<=",1,N),"=")

f.con <− rbind(aux ,c(as.numeric(inputs[i,]),rep(0,1,m)))

stacking [-inputs outputs] and $b^T z$

Case study – CRS model

Solving the DEA model in R program

```
require(lpSolve) # load lpSolve package

results<-lp("max",as.numeric(f.obj),f.con,f.dir,f.rhs,scale=0,compute.sens=TRUE) # lp solve

multipliers <-  results$solution

efficiency <- results$objval

lambdas <- results$duals[seq(1,N)]
```

lp is a function available on lpSolve package

u and v

θ

λ

Case study – DEA model

R code for the CRS/M/I DEA model

```
f.rhs <- c(rep(0,1,N),1)
f.dir <- c(rep("<=",1,N),"=")
aux <- cbind(-1*inputs,outputs)

for (i in 1:N) {
      f.obj <- c(0*rep(1,s),as.numeric(outputs[i,]))
      f.con <- rbind(aux ,c(as.numeric(inputs[i,]), rep(0,1,m)))
      results <- lp ("max",as.numeric(f.obj), f.con, f.dir, f.rhs,scale=0, compute.sens=TRUE)
      if (i==1) {
          weights <-  results$solution
          effcrs <- results$objval
          lambdas <- results$duals[seq(1,N)]
      } else {
          weights <- rbind(weights, results$solution)
          effcrs <- rbind(effcrs , results$objval)
          lambdas <- rbind(lambdas, results$duals[seq(1,N)] )
      }
}
```

loop of N DMUs

Case study – Exporting results to an Excel spreadsheet

Record multipliers and efficiency scores

	A	B	C	D	E	F
1		efficiency	OPEX	NETWORK	MWH	CUSTOMERS
2	ELETROPAULO	0.5827	8.0055E-10	0	6.0740E-09	5.6817E-08
3	CEMIG	0.5795	5.9441E-10	9.7382E-07	3.5032E-09	0
4	CPFL - Paulista	0.8854	2.0109E-09	0	1.5257E-08	1.4272E-07
5	COPEL	0.5186	9.8148E-10	7.9522E-07	1.1191E-08	2.1093E-08
6	LIGHT	0.7755	1.7947E-09	0	1.3617E-08	1.2737E-07
7	CELESC	0.5239	1.3861E-09	1.2843E-06	1.8660E-08	0
8	COELBA	1.0000	2.2913E-09	0	1.7385E-08	1.6262E-07
9	ELEKTRO	0.6879	2.4120E-09	1.9542E-06	2.7502E-08	5.1834E-08
10	CPFL - Piratininga	1.0000	5.1075E-09	0	7.6844E-08	0
11	BANDEIRANTE	0.6984	3.4864E-09	0	2.6452E-08	2.4744E-07
12	CELPE	0.8225	2.8518E-09	0	2.1638E-08	2.0240E-07
13	AMPLA	0.5498	2.2908E-09	0	1.7381E-08	1.6258E-07
14	CELG	0.5856	1.4462E-09	2.9357E-06	0	0
15	RGE	1.0000	5.3660E-09	4.3477E-06	6.1187E-08	1.1532E-07
16	COELCE	0.8198	3.1629E-09	0	0	2.9866E-07
17	ESCELSA	0.6697	4.1419E-09	3.3559E-06	4.7229E-08	8.9013E-08
18	AES SUL	0.8406	4.8515E-09	3.9308E-06	5.5319E-08	1.0426E-07
19	CEEE	0.4460	2.5907E-09	2.0991E-06	2.9541E-08	5.5677E-08

```
# merge the efficiency and multipliers
spreadsheet <- cbind(effcrs,weights)
# assign the utilities' names to the spreadsheet rows
rownames(spreadsheet) <- data[,1]
# assign the variables names to the spreadsheet columns
colnames(spreadsheet) <- c('efficiency',names(inputs),names(outputs))

# record file by xlsReadWrite package
write.xls(spreadsheet ,"resultscrs.xls",colNames = TRUE,sheet = 1,from = 1)
# or record file by xlsx package
write.xlsx(spreadsheet ,"resultscrs.xls",col.names = TRUE)
```

Case study – Exporting results to an Excel spreadsheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		ELETROPAULO	CEMIG	CPFL - Paulista	COPEL	LIGHT	CELESC	COELBA	ELEKTRO	CPFL - Piratininga	BANDEIRA	CELPE	AMPLA	CELG	RGE	COELCE	ESCELSA	AES SUL	CEEE
2	ELETROPAULO	0	0	0	0	0	0	0.5744	0	2.4372	0	0	0	0	0	0	0	0	0
3	CEMIG	0	0	0	0	0	0	0.9781	0	0	0	0	0	0	2.9403	0	0	0	0
4	CPFL - Paulista	0	0	0	0	0	0	0.2716	0	1.6435	0	0	0	0	0	0	0	0	0
5	COPEL	0	0	0	0	0	0	0.0751	0	0.2590	0	0	0	0	2.3874	0	0	0	0
6	LIGHT	0	0	0	0	0	0	0.3953	0	1.3260	0	0	0	0	0	0	0	0	0
7	CELESC	0	0	0	0	0	0	0	0	0.4101	0	0	0	0	1.5974	0	0	0	0
8	COELBA	0	0	0	0	0	0	1.0000	0	0	0	0	0	0	0	0	0	0	0
9	ELEKTRO						0	0.1244	0	0.3720	0	0	0	0	0.8483	0	0	0	0
10	CPFL - Piratininga						0	0	0	1.0000	0	0	0	0	0	0	0	0	0
11	BANDEIRANTE						0	0.0529	0	0.9052	0	0	0	0	0	0	0	0	0
12	CELPE	0	0	0	0	0	0	0.6227	0	0.0849	0	0	0	0	0	0	0	0	0
13	AMPLA	0	0	0	0	0	0	0.4379	0	0.2498	0	0	0	0	0	0	0	0	0
14	CELG	0	0	0	0	0	0	0.9279	0	0	0	0	0	0	0	0	0	0	0
15	RGE	0	0	0	0	0	0	0	0	0	0	0	0	0	1.0000	0	0	0	0
16	COELCE	0	0	0	0	0	0	0.5939	0	0	0	0	0	0	0	0	0	0	0
17	ESCELSA	0	0	0	0	0	0	0.0575	0	0.2639	0	0	0	0	0.4555	0	0	0	0
18	AES SUL	0	0	0	0	0	0	0.0041	0	0.0440	0	0	0	0	0.8739	0	0	0	0
19	CEEE	0	0	0	0	0	0	0.1682	0	0.1387	0	0	0	0	0.3840	0	0	0	0

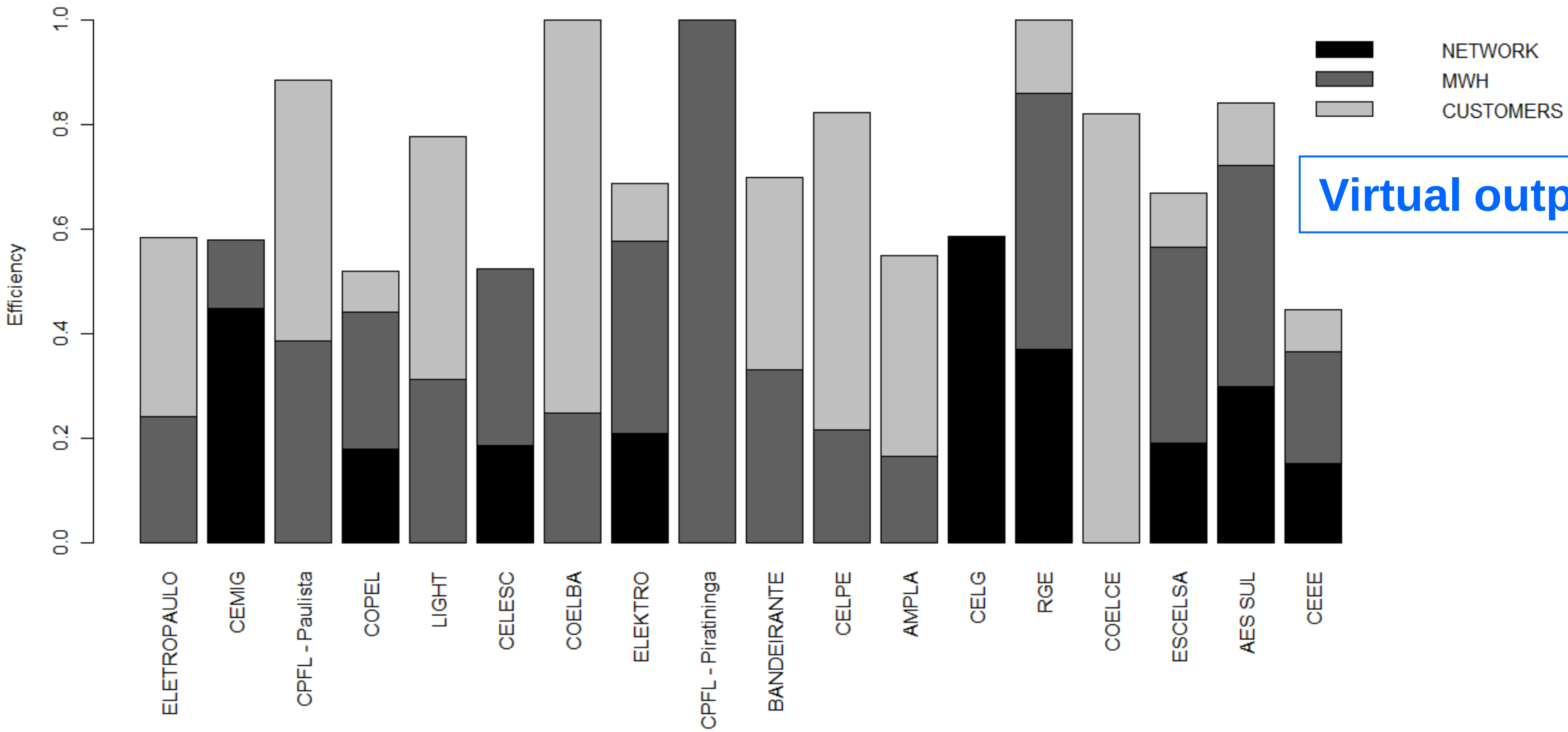
Duals variables λ

```
# duals variables
spreadsheet<-lambdas
# assign the utilities' names to the spreadsheet rows and columns
rownames(spreadsheet)<-data[,1]
colnames(spreadsheet)<- data[,1]

# record file by xlsReadWrite package
write.xls(spreadsheet ,"dualscrs.xls",colNames = TRUE,sheet = 1,from = 1)
# or record file by xlsx package
write.xlsx(spreadsheet ,"dualscrs.xls",col.names = TRUE)
```

Case study – Visualizing the results

El-Mahgary, S.; Lahdelma, R. Data Envelopment Analysis: Visualizing the results, European Journal of Operational Research, 85, pp. 700-710, 1995.

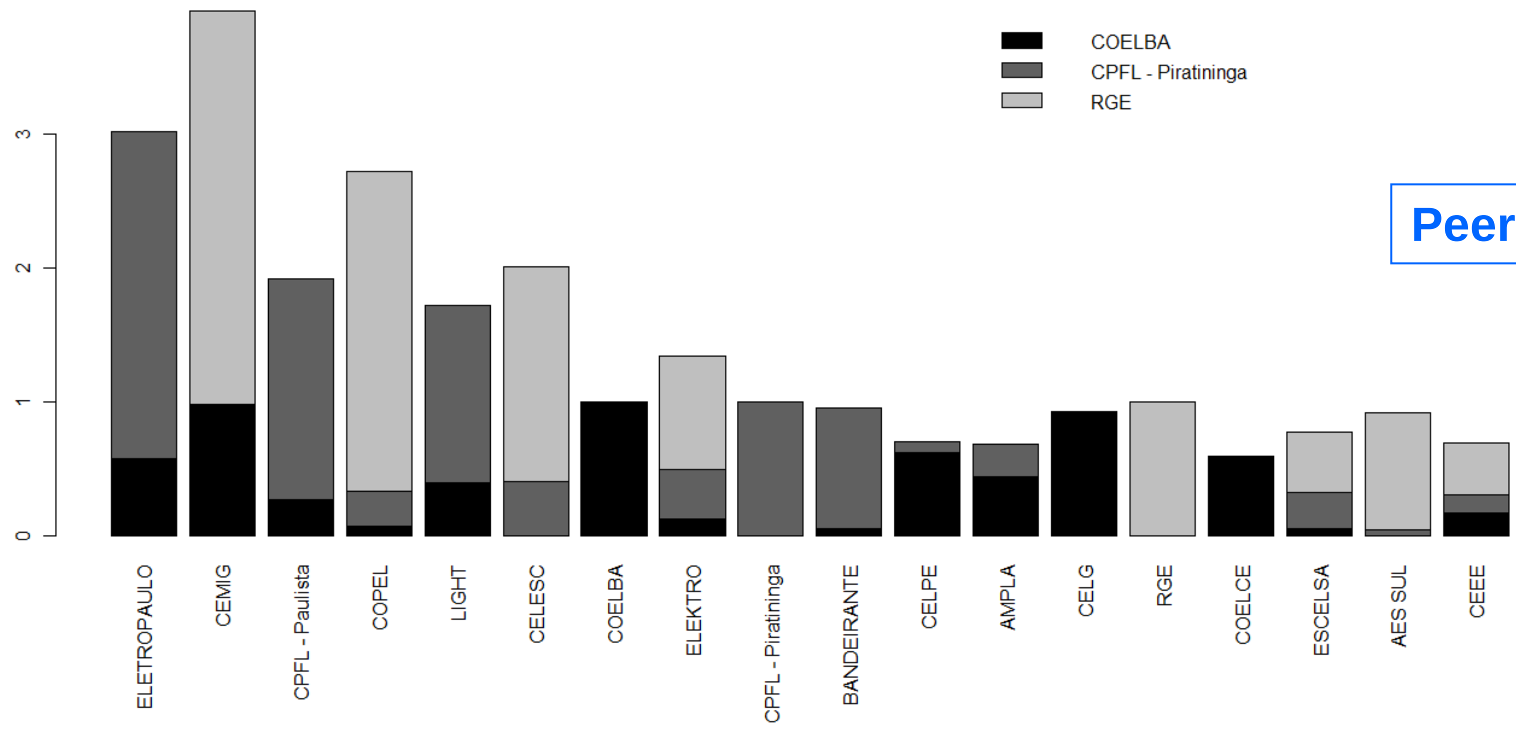


Virtual outputs

```
par(mar=c(10,5,1,10),xpd=TRUE) # set plot margin
palette(gray(0:8 / 8)) # set color palette
virtual<-weights[,(s+1):(s+m)]*outputs  # virtual outputs
rownames(virtual)<-data[,1] # assign utilities' names to the rows of the object virtual
barplot(t(virtual),col=palette()[c(1,4,7)],ylab="Efficiency",cex.axis=1,cex.lab=1,cex.names=1,las=3)
legend("topright",inset=c(-0.45,0),colnames(virtual),fill=palette()[c(1,4,7)],bty="n") # add legend
```

Case study – Visualizing the results

El-Mahgary, S.; Lahdelma, R. Data Envelopment Analysis: Visualizing the results, European Journal of Operational Research, 85, pp. 700-710, 1995.



Peer set

```
par(mar=c(10,5,1,1)) # set plot margin
palette(gray(0:8 / 8)) # set color palette
rownames(lambdas)<-data[,1] # assigns utilities' names to the rows of duals variables matrix
colnames(lambdas)<-data[,1] # assigns utilities' names to the colimns of duals variables matrix
peer<-which(apply(lambdas,2,sum)>0) # identifies reference units (efficiency units)
barplot(t(lambdas[,peer]),col=palette()[c(1,4,7)],cex.axis=1,cex.lab=1,cex.names=1,las=3)
legend("topright",colnames(lambdas)[peer],fill=palette()[c(1,4,7)],bty="n") # add legend
```

Case study – VRS model

The difference between CRS and VRS models resides in the unconstrained variable u_0.

This variable can be modeled by the difference of two non negative variables ($u_0 = u^+ - u^-$, where $u^+ \geq 0$ and $u^- \geq 0$),

```
f.rhs <− c(rep(0,1,N),1)
f.dir<−c(rep("<=",1,N), "=")
aux <− cbind(-1*inputs,outputs,1,-1)
```

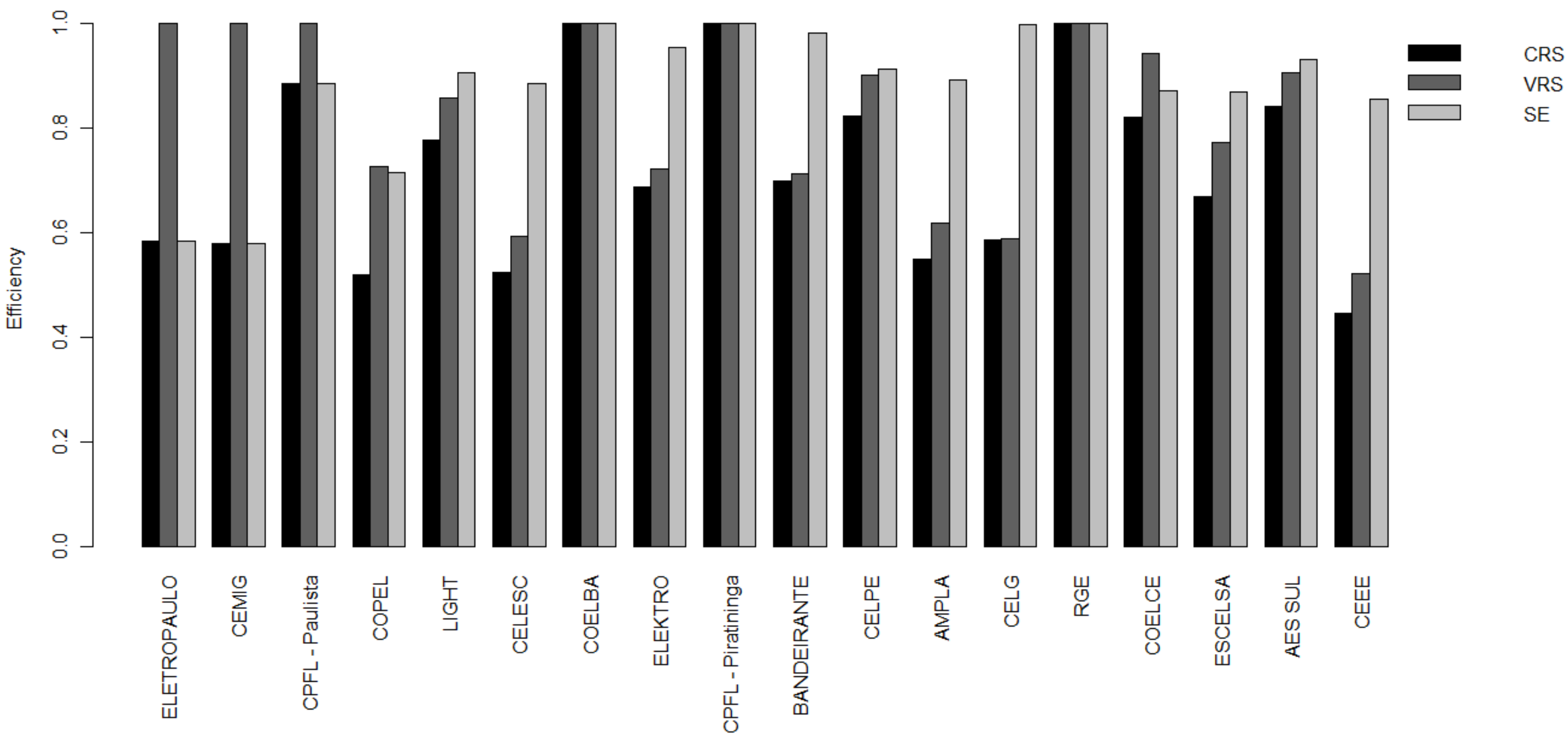
```
for (i in 1:N) {
     f.obj<−c(rep(0,1,s),as.numeric(outputs[i,]),1,-1) # 1 and -1 represents u+ - u-
     f.con <− rbind(aux,c(as.numeric(inputs[i,]),rep(0,1,m),0,0))
     results <− lp ("max",as.numeric(f.obj), f.con, f.dir, f.rhs,scale=1, compute.sens=TRUE)
     multipliers <− results$solution
     u0 <− multipliers[s+m+1]-multipliers[s+m+2]
     if (i==1) {
               weights <− c(multipliers[seq(1,s+m)],u0)
               effvrs <− results$objval
               lambdas <− results$duals[seq(1,N)]
     } else {
               weights<−rbind(weights,c(multipliers[seq(1,s+m)],u0))
               effvrs <− rbind(effvrs , results$objval)
               lambdas <− rbind(lambdas,results$duals[seq(1,N)])
     }
}
```

Case study – VRS model

	A	B	C	D	E	F	G
1		efficiency	OPEX	NETWORK	MWH	CUSTOMERS	u0
2	ELETROPAULO	1.0000	8.0055E-10	0	4.1071E-08	0	-6.3965E-01
3	CEMIG	1.0000	5.9441E-10	3.0201E-06	0	0	-3.8990E-01
4	CPFL - Paulista	1.0000	2.0109E-09	2.2662E-06	3.6966E-08	0	-1.3774E-01
5	COPEL	0.7263	9.8148E-10	2.1527E-06	2.9968E-08	0	-4.6262E-01
6	LIGHT	0.8561	1.7947E-09	0	2.2572E-08	1.23869E-07	-1.1174E-01
7	CELESC	0.5927	1.3861E-09	1.5621E-06	2.5481E-08	0	-9.4940E-02
8	COELBA	1.0000	2.2913E-09	4.6511E-06	0	0	0
9	ELEKTRO	0.7210	2.4120E-09	2.7537E-06	3.8957E-08	0	-9.5958E-02
10	CPFL - Piratininga	1.0000	5.1075E-09	0	7.6844E-08	0	0
11	BANDEIRANTE	0.7122	3.4864E-09	0	0	2.5779E-07	3.3008E-01
12	CELPE	0.9015	2.8518E-09	4.1028E-08	0	2.0844E-07	2.7241E-01
13	AMPLA	0.6176	2.2908E-09	0	0	1.6938E-07	2.1688E-01
14	CELG	0.5880	1.4462E-09	2.7819E-06	0	0	3.3056E-02
15	RGE	1.0000	5.3660E-09	4.9719E-06	7.2238E-08	0	0
16	COELCE	0.9421	3.1629E-09	4.5503E-08	0	2.3117E-07	3.0212E-01
17	ESCELSA	0.7719	4.1419E-09	0	0	0	7.7188E-01
18	AES SUL	0.9041	4.8515E-09	0	0	0	9.0411E-01
19	CEEE	0.5225	2.5907E-09	3.7272E-08	0	1.8935E-07	2.4747E-01

```
spreadsheet<-cbind(effvrs,weights) # merge de results
rownames(spreadsheet)<-data[,1] # assign the utilities' names to spreadsheet rows
colnames(spreadsheet)<-c('efficiency',names(inputs), names(outputs),'u0')
# record file by xlsReadWrite package
write.xls(spreadsheet,"resultsvrs.xls",colNames = TRUE,sheet = 1,from = 1)  # record file
# or record file by xlsx package
write.xlsx(spreadsheet,"resultsvrs.xls",col.names = TRUE)  # record file
```

Case study – CRS and VRS efficiencies




```
par(mar=c(10,5,1, 8),xpd=TRUE) # set plot margin
scale <- effcrs/effvrs
spreadsheet <- cbind(effcrs,effvrs,scale)
rownames(spreadsheet) <- data[,1]
colnames(spreadsheet) <- c("CRS","VRS","SE")
barplot(t(spreadsheet),
col=palette()[c(1,4,7)],ylab="Efficiency",beside=TRUE,cex.axis=1,cex.lab=1,cex.names=1,las=3)
legend("topright",inset=c(-0.2,0),colnames(spreadsheet),fill=palette()[c(1,4,7)],bty="n") # add legend
```

Conclusions

The R codes presented in this paper are examples of how to implement DEA models in R and they can be easily adapted to more sophisticated DEA models, for example, models with constraints to the weights, cross-evaluation, DEA two-stage and resource allocation.

The R is free, open source, highly extensible, widely available in the academic community and has features that might enable friendly graphical interfaces and integration with MS Excel spreadsheets and other quantitative techniques, an important aspect when considering a broader view of the decision making process.

References

Angulo-Meza, L.; Estellita Lins, M.P. Review of methods for increasing discrimination in data envelopment analysis, Annals of Operational Research, 116, pp. 225-242, 2002.

Banker, R.D.; Charnes, A. & Cooper, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30, 1078-1092, 1984.

Bogetoft, P., Otto, L. Benchmarking with DEA, SFA and R, Springer Science, 2011.

Charnes, A.; Cooper, W.W. & Rhodes, E. Measuring the Efficiency of Decision Making Units. European Journal of Operational Research, 2, 1978.

Coelli, T.J.; Rao, D.S.P.; O'Donnell, C.J.; Battese, G.E. An introduction to efficiency and productivity analysis, Springer, 2005.

Cook, W.D.; Seiford, L.M. Data Envelopment Analysis (DEA) – Thirty years on, European Journal of Operational Research, 192, pp. 1-17, 2009.

Cooper, W.W.; Seiford, L.M.; Tone, K. Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver Software, Kluwer Academic Publishers, 2002.

El-Mahgary, S.; Lahdelma, R. Data Envelopment Analysis: Visualizing the results, European Journal of Operational Research, 85, pp. 700-710, 1995.

Guo, P.; Tanaka, H. Fuzzy DEA: a perceptual evaluation method, Fuzzy Sets and Systems, 119, pp. 149-160, 2001.

Ragsdale, C.T. Spreadsheet modeling & decision analysis: a practical introduction to management science. 4.ed. Thomson South-Western, 2004.

R Development Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/, 2011.

Verzani, J. Using R for introductory statistics, Chapman & Hall/ CRC Press, 2005.

José Francisco Moreira Pessanha
professorjfmp@hotmail.com
Alexandre Marinho
alexandre.marinho@ipea.gov.br
Luiz da Costa Laurencel
llaurenc.ntg@terra.com.br
Marcelo Rubens dos Santos do Amaral
mrubens@ime.uerj.br

Thank you