TEACHING DATA ENVELOPMENT ANALYSIS IN UNDERGRADUATE STATISTICS COURSES

José F.M. Pessanha[1], Alexandre Marinho[1,2], Luiz C. Laurencel[1] and Marcus V.P. Souza[3]
[1]Rio de Janeiro State University - Uerj, Brazil
[2]Institute for Applied Economic Research – IPEA, Brazil
[3]Federal Center for Technological Education Celso Suckow da Fonseca – UnED Valença, Brazil
professorjfmp@hotmail.com

In general the undergraduate curriculum in statistics offers one semester basic course on linear programming (LP). The conventional course on LP includes simplex algorithm, duality theory, sensitivity analysis and some applications like the feedmix and transport problems. This outline seems to be distant from the mainstream curriculum in statistics. In order to partially overcome this shortcoming, we propose to introduce a technique based on linear programming called Data Envelopment Analysis (DEA) on the LP courses for students of statistics. DEA has gained momentum as a powerful complementary method for statistical tools in the research agenda as well as in practical activities of economic efficiency evaluation.

INTRODUCTION

The linear programming LP is the main branch of the operational research with applications in several domains (Bazaraa et al, 1990). Despite its deterministic nature, the linear programming is a rather valuable technique and it must be regarded by the statisticians. Some undergraduate courses in statistics offer one semester basic course on linear programming where the students learn the simplex algorithm, the duality theory, the sensitivity analysis and applications of the LP theory to solve some classical problems like the feedmix and the transport problems. However, a course on LP to statistical students can explore other LP applications, in particular, the Data Envelopment Analysis – DEA proposed by Charnes et al (1978). DEA is a nonparametric technique based on linear programming to evaluate the efficiency of profit and non-profit organizations in a same industry, for example, schools, hospitals, banks and factories.

In the classical problems the LP is applied as an ex-ante tool in planning with aim to find the decision variables values that optimize a linear objective function subject to linear constraints. In DEA the LP is applied in order to evaluate the performance of decision making units (DMU). Therefore, different from other LP applications, DEA is an ex-post tool (Baker, 2011).

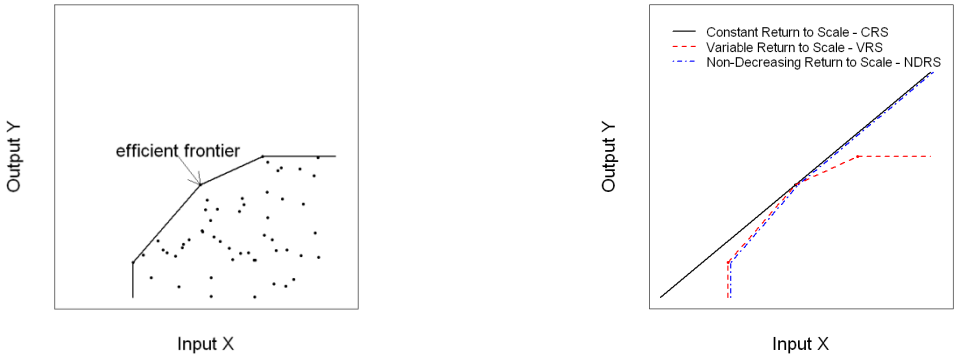
DEA is a widely used technique for evaluating the efficiency of peer entities called decision making units (DMU) which convert multiple inputs into multiple outputs. The efficiency score of a DMU depends on its distance to the efficient frontier, but the true frontier is not known. DEA is a LP model designed to identify the efficient frontier from data. In this aspect, DEA sounds like a regression linear model. In fact the statistical approaches to the same problem are based on linear regression model, for example the Corrected Ordinary Least Square (COLS) and the Stochastic Frontier Analysis (SFA). In addition, DEA presents other features found in statistical methods, for example, the choice of variables and sample units to be analyzed.

DATA ENVELOPMENT ANALYSIS

The production of manufactured goods or services is achieved through a process or decision making unit (DMU) that transforms an input x into a product y. Many process can make this transformation, all of them are in the production possibility set $T(X,Y)=\{(X,Y)|$it is feasible produce Y from $X\}$ in Figure 1a. The set $T(X,Y)$ is limited by the efficient frontier, a function that express the maximum output achieved from a given level of input (output orientation) or the minimum input required to achieve a determined output level (input orientation). Based on the Pareto-Koopmans definition (Coelli et al, 2005) all process sited on the efficiency frontier are efficient and all process in the interior of $T(X,Y)$ are inefficient. The efficient frontier is a benchmark to the performance of the different process that transform x into y. The distance between a DMU and the frontier express the inefficiencies in the process. Based on this distance we can build an index in the interval

In: M.A. Sorto (Ed.), *Advances in statistics education: developments, experiences and assessments. Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),* July 2015, Rio de Janeiro, Brazil. iase-web.org/Conference_Proceedings.php

[0,1] so that the unitary value indicates efficiency, otherwise the process is inefficient. However, the efficient frontier is not known. DEA can identify the efficient frontier from a sample of DMU (Figure 1a) together with assumptions about the frontier's shape (Figure 1b).



(a) sample of DMU and the efficient frontier (b) efficient frontier and returns to scale models

Figure 1. Efficient frontier

In general, a DMU consumes multiple inputs $X=(x_1,...,x_s)$ and produces multiple outputs $Y=(y_1,...,y_m)$, in this case the efficiency score is defined by the following quotient:

$$efficiency = (u_1 y_1 + \ldots + u_m y_m)/(v_1 x_1 + \ldots + v_s x_s) = (U \cdot Y)/(V \cdot X) \qquad (1)$$

where $V=(v_1,...,v_s)$ and $U=(u_1,...,u_m)$ are the weights assigned to the inputs and outputs respectively.

Charnes et al (1978) suggest that the vectors U and V must be determined by the linear programming problem - LPP (2) in Table 1, called DEA CRS (Constant Return to Scale) input oriented in the multiplier form. In (2) the objective function $\theta = u_1 y_{1,j_0} + \ldots + u_m y_{m,j_0}$ is the efficiency score of the evaluated DMU (DMU_{j0}) and the linear constraints represent the production possibility set. The DMU_{j0} is fully efficient if $\theta=1$ and all weights are positive at the optimal solution. If $\theta=1$ but some weights are equal to zero the DMU_{j0} is weakly efficient. Otherwise, if $\theta<1$ the DMU is inefficient (Cook & Zhu, 2005).

Table 1. DEA CRS input oriented

$efficiency = \underset{u,v}{Max} \sum_{i=1}^{m} u_i y_{i,j_0}$ (2)	$efficiency = \underset{\lambda,\theta}{Min} \ \theta$ (3)
s.t.	*s.t.*
$-\sum_{i=1}^{s} v_i x_{ij} + \sum_{i=1}^{m} u_i y_{ij} \le 0 \ \ \forall j = 1,\ldots,j_0,\ldots,N$	$\theta X_{j_0} \ge \sum_{j=1}^{N} \lambda_j X_j$
$\sum_{i=1}^{s} v_i x_{i,j_0} = 1$	$Y_{j_0} \le \sum_{j=1}^{N} \lambda_j Y_j$
$u_i \ge 0 \ \ \forall i = 1,m \ , \ v_i \ge 0 \ \ \forall i = 1,s$	$\lambda_j \ge 0 \forall j = 1,\ldots,j_0,\ldots,N$

Under the resources conservation approach (input orientation), the measure of technical efficiency θ ($0 \le \theta \le 1$) is defined as the maximum radial contraction of the input vector X that produces the same amount of product Y:

$$efficiency = Min \ \{\theta \mid (\theta X, Y) \in \text{ production possibilities set } T(X,Y) \ \} \qquad (4)$$

Through the duality theory we can derive an equivalent model known as DEA model in the envelopment form under input orientation whose mathematical formulation corresponds to the model (3) in Table 1. In this case, the DMU_{j0} is fully efficient if and only if θ=1 and all slack variables are equal to zero. If θ=1 but some slack variables are positive the DMU_{j0} is weakly efficient. Otherwise, the DMU is inefficient. It should be emphasized that the LPP (2) or (3) must be solved for each DMU in order to compute its efficiency score.

Later, Banker et al (1984) added the constraint $\lambda_1+\ldots+\lambda_N=1$ in the model (3). The result is the DEA model with variable return to scale (6) called DEA VRS. Model (5) in Table 2 is the VRS model in the multiplier form and input oriented, where the unconstrained variable u_0 corresponds to the constraint $\lambda_1+\ldots+\lambda_N=1$ in (6). Cook & Zhu (2005) present output oriented DEA models corresponding to the models presented in Tables 1 and 2.

Table 2. DEA VRS input oriented

$$efficiency = \max_{u,v} \sum_{i=1}^{m} u_i y_{i,j_0} + u_0 \qquad (5)$$

$$s.t.$$

$$-\sum_{i=1}^{s} v_i x_{ij} + \sum_{i=1}^{m} u_i y_{ij} + u_0 \le 0 \quad \forall j = 1,\ldots,j_0,\ldots,N$$

$$\sum_{i=1}^{s} v_i x_{i,j_0} = 1$$

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

$$efficiency = \min_{\lambda,\theta} \theta \qquad (6)$$

$$s.t.$$

$$\theta X_{j_0} \ge \sum_{j=1}^{N} \lambda_j X_j$$

$$Y_{j_0} \le \sum_{j=1}^{N} \lambda_j Y_j$$

$$\lambda_1+\ldots+\lambda_N=1$$

$$\lambda_j \ge 0 \forall j = 1,\ldots,j_0,\ldots,N$$

DEA AND THE UNDERGRADUATE CURRICULUM IN STATISTICS

Traditionally, the operational research and linear programming courses in the undergraduate programs include the simplex algorithm, duality theory, sensitivity analysis and some case studies that illustrate how apply LP to optimize the resource allocation, for example, the feedmix and transportation problems (Bazaraa et al, 1990). During the course the students learn how to use electronic worksheet with Solver function to solve linear programming problems (Baker, 2011). The modern textbooks on operational research or linear programming include a chapter or section about DEA models. Then, in a basic course it is not necessary adopt an exclusive book on DEA. Based on textbooks and classroom experience with undergraduate students, we recommend introduce the classical DEA models (CRS and VRS) only after the students learn the feedmix and the transportation problems. This sequence aligns with the linear programming development and puts in evidence the contrasts between the ex-ante and ex-post nature of the different LP applications (Baker, 2011). DEA models offer a good point to teach linear programming and duality theory, for example the equivalence between the multiplier and envelopment models show the primal and dual formulations and how to move between them.

Some microeconomic concepts like as possibility production set, efficiency, productivity, return to scale and efficient frontier should be taught before the introduction of DEA. These basic concepts are important in econometrics, other disciplines often offered in the statistic undergraduate program. Based on these concepts the students can compare the classical regression line with the efficient frontier. DEA can introduce students to productivity and efficiency analysis, it is a great opportunity to the students compare deterministic (DEA) and statistical approaches (OLS, COLS and SFA) designed to solve the same problem, but with different assumptions and weak or strong points (Thanassoulis, 1993). DEA models also offer a way to show some interfaces between LP and statistical techniques, for example, cluster analysis to select DMU, principal components to transform variable and DEA approaches with two stages, where the combination of DEA and tobit or linear regression models as well as bootstrap resampling is useful (Coelli et al, 2005). Although R (R Development Core Team, 2014) offers packages to solve DEA models, for example, the Frontier Efficiency Analysis with R - FEAR (Wilson, 2008) and Benchmarking

(Bogetoft & Otto, 2011), we recommend the resolution of exercises with R package lpSolve (Buttrey, 2005) in order to develop the students skills on R.

EXAMPLE

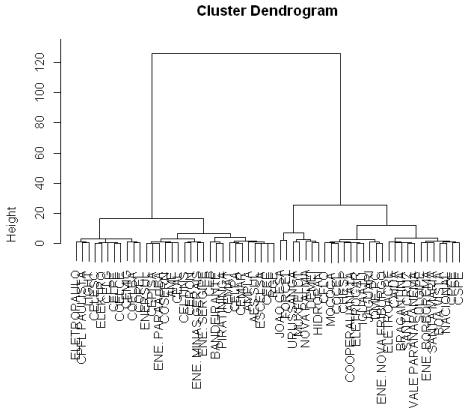
Below we present descriptions of an example designed to introduce DEA models. We consider public data from Brazilian electricity distribution utilities available on site of the Brazilian Electricity Regulatory Agency (Aneel). The dataset has been used in the discussions about the tariff review methodology. DEA has been used by regulatory agencies around the world (Jasmab & Pollitt, 2000). The following R code downloads and extracts the datafile. The datafile consists of 61 cases (distribution utilities) described by 11 variables.

```
library(downloader) ; library(xlsx) ; library(utils) # load packages
setwd("c:/IASE") # set work directory
url = "http://www.aneel.gov.br/aplicacoes/audiencia/arquivo/2014/023/documento/custos_operacionais_-_atualizada.zip"
file="OPEX.zip" ;  download(url,file,mode="wb") # download file
unzip(file,files=NULL,list=FALSE,overwrite=TRUE,junkpaths=TRUE,exdir=".",unzip="internal",setTimes=FALSE)
data=read.xlsx("Base DEA - atualizada.xlsx",header=TRUE,startRow=2, sheetIndex=1) # read spreadsheet
utility=strsplit(as.character(data[,1]),"_m") # utility's names
opex=data[,4] # average operational expenditure in the period 2011-2013
customers=data[,8] # average number of customers in the period 2011-2013
energysale=data[,9] # average energy sales in the period 2011-2013
network=data[,5]+data[,6]+data[,7] # total length distribution network (underground + overhead lines + high voltage)
datamatrix=cbind(opex,customers,energysale,network) ; rownames(datamatrix)=utility  # data matrix
```

The specification of a DEA model has three stages: selection of the DMU, choice of input and output variables and specification of the return to scale model (constant or variable). The regulatory agency aims assess the efficient level of the operational expenditure (opex) of each distribution utility. Given the market size served by a utility, it must operate at least cost defined by the efficient frontier. The deviation between the actual opex and the least cost is assigned to the inefficiency in the utility management. Based on the efficiency score the regulatory agency penalizes inefficiencies when set the tariff level for the utilities.

Then in this case we have a DEA input oriented model in which opex is the only input variable (s=1). The output variables in the DEA model are the driver of operational expenditures: numbers of customers, total energy sales and distribution network length (Appa et al, 2010), then three output variables (m=3). The efficient frontier's shape depends on the assumptions about the return to scale. For purposes of incentive regulation the frontier should not allow decreasing return to scale. In this case we should consider the DEA CRS model (Appa et al, 2010).

DEA is based on the assumption of comparability among the DMU. However the Brazilian distribution utilities are heterogeneous. In order to establish fair comparisons we can classify the 61 utilities in two groups: large and small utilities. We applied the Ward method to classify the utilities and got the cluster with the large utilities. The cluster analysis was applied to the logarithm of the output variables and the results are illustrated in Figure 2 where the dendrogram shows that the two clusters are well separated. The large utilities concentrate 97% of the total energy sales.




```
# cluster analysis
results =  hclust(dist(log(datamatrix[,c(2,3,4)])),method="ward.D")
plot(results) # dendrogram
members = cutree(results,k=2) # partition with 2 clusters
datamatrix=data.frame(cbind(datamatrix,members))
attach(datamatrix)

#selected DMU classified in cluster 1 (large utilities)
index=which(members==1)
data_dea = datamatrix[index,]
```

Figure 2. Cluster analysis

In many practical applications the analyst should write the LP problem in a computational language and use a linear programming solver. Then, it is interesting show some aspects about the computational implementation. The DEA model processing consists in solving a LPP for each one of the *N* DMU, in this example *N*=32 (Figure 3a). The LPP corresponding to the DEA CRS model is shown in Figure 3b. The decision variable vector (v_{sx1}, u_{mx1}) has four elements, the v, u_1, u_2, u_3. The vector *inputs* in Fig 3b is displayed in the first column (opex) of the *data_dea* matrix (Figure 3a) while the *output*s is formed by the columns 2, 3 and 4 of *data_dea* matrix.

```
> data_dea
                      opex customers energysale   network
AES SUL           270415.60 1239688.0  3673950.8  65694.55
AME               374980.23  755385.7  2220492.2  37330.76
AMPLA             479317.47 2416584.0  4733589.1  55064.57
BANDEIRANTE       327364.56 1604297.0  4955248.1  27196.75
CEAL              312737.58  948724.0  1438051.7  39672.02
CEB               333767.26  912788.0  3445717.6  17496.52
CEEE              597813.96 1536012.3  3891644.5  66562.45
CELESC            842382.04 2487399.3  8300055.6 149755.43
CELG              762130.69 2509411.0  5431280.2 207674.04
CELPA             577061.08 1932670.0  3345583.6 111236.26
CELPE             549361.83 3241638.3  5493061.7 128683.31
CELTINS           158705.42  499118.7   890450.1  80383.78
CEMAR             394983.21 2033996.3  2892179.7 118231.46
CEMAT             423266.80 1163351.7  3171229.1 146233.86
CEMIG            2041586.44 7482055.3 15504078.2 496665.68
CEPISA            334005.76 1058070.0  1607821.5  76056.04
CERON             239274.79  542132.7  1007129.2  49982.03
COELBA            835616.98 5212097.3  7490080.0 252339.14
COELCE            459836.86 3077745.0  5037229.5 134134.32
COPEL            1225581.91 4045534.3 11630148.9 233505.75
COSERN            196500.78 1204178.0  2153663.9  48650.00
PIRATININGA       273902.09 1497643.7  5072889.0  22691.26
CPFL PAULISTA     720481.06 3832970.0 11480849.8 114129.04
ELEKTRO           463617.95 2311530.7  6453713.0 109656.94
ELETROPAULO      1255830.56 6476079.3 20975284.4  44956.92
ENE. MINAS GERAIS  95472.57  404450.0   703729.2  26805.73
ENERSUL           331261.32  875949.7  2063642.8  85176.16
ENE. PARAÃ□BA     249989.18 1216626.3  1846455.1  70058.60
ESCELSA           302786.96 1330784.0  3225925.4  59955.39
ENE. SERGIPE      164595.26  652585.3  1212603.8  26708.25
LIGHT             722222.02 3643105.7 10966774.5  60560.16
RGE               244263.52 1337437.0  4035004.2  81296.38
```

(a) electricity distribution utilities (DMU)

$$Max \quad \begin{bmatrix} 0_{1xs} & outputs_{j0,1xm} \end{bmatrix} \cdot \begin{pmatrix} v_{sx1} \\ u_{mx1} \end{pmatrix}$$

$$s.t. \quad \begin{bmatrix} -inputs_{Nxs} & outputs_{Nxm} \end{bmatrix} \cdot \begin{pmatrix} v_{sx1} \\ u_{mx1} \end{pmatrix} \leq 0_{Nx1}$$

$$\begin{pmatrix} inputs_{j0,1xs} & 0_{1xm} \end{pmatrix} \cdot \begin{pmatrix} v_{sx1} \\ u_{mx1} \end{pmatrix} = 1$$

$$u_{mx1} \geq 0$$

$$v_{sx1} \geq 0$$

where

```
N<- dim(data_dea)[1] # number of DMU
s<- 1 ; m<- 3 # number of inputs and outputs respectively
inputs <- data_dea[,1]
outputs<-data_dea[,c(2,3,4)]
```

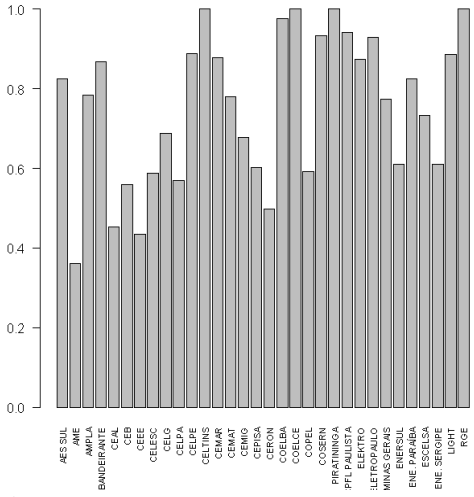
(b) DEA CRS model

Figure 3. Data and DEA CRS model

The efficiency scores can be evaluated by the R code below, where *i* is the index of the evaluated DMU and the vectors $input_{j0}$ and $output_{j0}$ are modified automatically for each DMU:

```
library(lpSolve) # load lpSolve package previously installed
f.rhs<- c(rep(0,N),1) # RHS constraints of the LPP 2 at Table 1
f.dir<- c(rep("<=",N),"=") # directions of the constraints of the LPP 2
aux<- cbind(-1*inputs,outputs) # matrix of constraint coefficients of the LPP 2
for (i in 1:N) {
  f.obj<-c(rep(0,s),t(data_dea[i,c(2,3,4)])) # objective function coefficients
  f.con<- rbind(aux ,c(data_dea[i,1], rep(0,m))) # complete matrix of constraint coefficients of the LPP 2
  results <- lp("max",f.obj,f.con,f.dir,f.rhs,scale=1,compute.sens=TRUE) # solve LPP
  multipliers<- results$solution # input and output weights
  efficiency<- results$objval # efficiency score
  duals<- results$duals # shadow prices λ , LPP 3 at Table 1
  if (i==1) {
    weights<- c(multipliers[seq(1,s+m)]) ; effcrs<- efficiency ; lambdas<- duals [seq(1,N)]
  } else {
    weights<- rbind(weights,c(multipliers[seq(1,s+m)])) ; effcrs<- rbind(effcrs , efficiency)
    lambdas<- rbind(lambdas,duals[seq(1,N)])
  }
}
matrix_results <- cbind(effcrs,weights,lambdas) ; rownames(matrix_results) <- rownames(data_dea)
colnames(matrix_results)<-c("efficiency",colnames(data_dea)[1:(s+m)], rownames(data_dea))
```

The barplot in Figure 4a presents the efficiency scores. The results show four utilities with efficiency scores equal to 1. These four utilities are the peer set for the inefficient utilities. However, Figure 4b shows some output weights equal to zero for these four utilities, then they are weakly efficient. In order to avoid null weights the DEA model can be improved with the addition of constraints to the weights (Cook & Zhu, 2005).



(a) barplot(matrix_results[,1],cex.names=0.7,las=2)

```
> matrix_results[,2:5]
                          opex    customers   energysale      network
AES SUL           3.698012e-06 3.920314e-07 8.107944e-08 6.373741e-07
AME               2.666807e-06 2.827120e-07 5.847013e-08 4.596400e-07
AMPLA             2.086300e-06 2.464010e-07 3.990250e-08 0.000000e+00
BANDEIRANTE       3.054698e-06 3.607731e-07 5.842406e-08 0.000000e+00
CEAL              3.197569e-06 4.777394e-07 0.000000e+00 0.000000e+00
CEB               2.996100e-06 0.000000e+00 1.561753e-07 1.250606e-06
CEEE              1.672761e-06 1.773317e-07 3.667553e-08 2.883103e-07
CELESC            1.187110e-06 0.000000e+00 6.187954e-08 4.955130e-07
CELG              1.312111e-06 1.009306e-07 8.249475e-09 1.872480e-06
CELPA             1.732919e-06 1.333001e-07 1.089517e-08 2.473004e-06
CELPE             1.820294e-06 2.149846e-07 3.481488e-08 0.000000e+00
CELTINS           6.300982e-06 0.000000e+00 1.683702e-07 1.057520e-05
CEMAR             2.531753e-06 2.199296e-07 0.000000e+00 3.632972e-06
CEMAT             2.362576e-06 0.000000e+00 6.313100e-08 3.965211e-06
CEMIG             4.898152e-07 3.767771e-08 3.079555e-09 6.990028e-07
CEPISA            2.993960e-06 2.600809e-07 0.000000e+00 4.296222e-06
CERON             4.179295e-06 3.214810e-07 2.627597e-08 5.964166e-06
COELBA            1.196721e-06 1.039573e-07 0.000000e+00 1.717249e-06
COELCE            2.174684e-06 2.568396e-07 4.159294e-08 0.000000e+00
COPEL             8.159389e-07 8.649882e-08 1.788958e-08 1.406319e-07
COSERN            5.089038e-06 6.010374e-07 9.733277e-08 0.000000e+00
PIRATININGA       3.650940e-06 0.000000e+00 1.971263e-07 0.000000e+00
CPFL PAULISTA     1.387962e-06 1.471397e-07 3.043126e-08 2.392234e-07
ELEKTRO           2.156948e-06 2.286611e-07 4.729141e-08 3.717628e-07
ELETROPAULO       7.962858e-07 9.404479e-08 1.522973e-08 0.000000e+00
ENE. MINAS GERAIS 1.047421e-05 8.057006e-07 6.585324e-08 1.494748e-05
ENERSUL           3.018765e-06 2.322103e-07 1.897951e-08 4.308002e-06
ENE. PARAÍBA      4.000173e-06 3.474891e-07 0.000000e+00 5.740099e-06
ESCELSA           3.302652e-06 3.501188e-07 7.241112e-08 5.692316e-07
ENE. SERGIPE      6.075509e-06 6.440732e-07 1.332064e-07 1.047150e-06
LIGHT             1.384616e-06 1.635291e-07 2.648211e-08 0.000000e+00
RGE               4.093939e-06 0.000000e+00 2.134016e-07 1.708856e-06
```

(b) matrix_results[,2:6]

Figure 4. Efficiency scores (a) and weights (b)

CONCLUSION

We proposed including DEA in the basic linear programming (LP) courses offered to the undergraduate statistical students. DEA is based on LP therefore the introduction of DEA in the undergraduate courses is straightforward. However different from other LP applications DEA sounds like statistical techniques. DEA opens the opportunity to align LP to the mainstream of the curriculum in statistics and contribute to introduce students to the efficiency and productivity analysis, a branch where deterministic and statistical approaches compete.

REFERENCES

Appa, G., Bana e Costa, C.A., Chagas, M.P., Ferreira, F.C., & Soares, J.O. (2010). DEA in X-Factor evaluation for the Brazilian Electricity Distribution Industry, *Working paper LSEOR 10-121*, London School of Economics and Political Science.

Baker, K. R. (2011). *Optimization Modeling with Spreadsheets*, 2nd ed. Wiley.

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

Bazaraa, M.S., Jarvis, J.J., & Sherali, H.D. (1990). *Linear programming and network flows*, 2nd ed., John Wiley & Sons.

Bogetoft, P., & Otto, L. (2011). *Benchmarking with DEA, SFA and R*, Springer Science.

Buttrey, S.E. (2005). Calling the lp_solve linear program software from R, S-Plues and Excel, *Journal of Statistical Software*, *14*(4).

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

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., & Battese, G.E. (2005). *An introduction to efficiency and productivity analysis*, 2nd ed., Springer.

Cook, D.W., & Zhu, J. (2005). *Modeling Performance Measurement: applications and implmentations issues in DEA*, Springer.

Jasmab, T., Pollitt, M. (2000). Benchmarking and regulation: International electricity experience. *Utilities Policy*, *9,* 107-130.

Thanassoulis, E. (1993). A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments. *The Journal of the Operational Research Society*, *44* (11), 1129-1144.

Wilson, P.W. (2008). FEAR: A software package for frontier efficiency analysis with R, *Socio-Economic Planning Sciences*, *42*, 247-254.