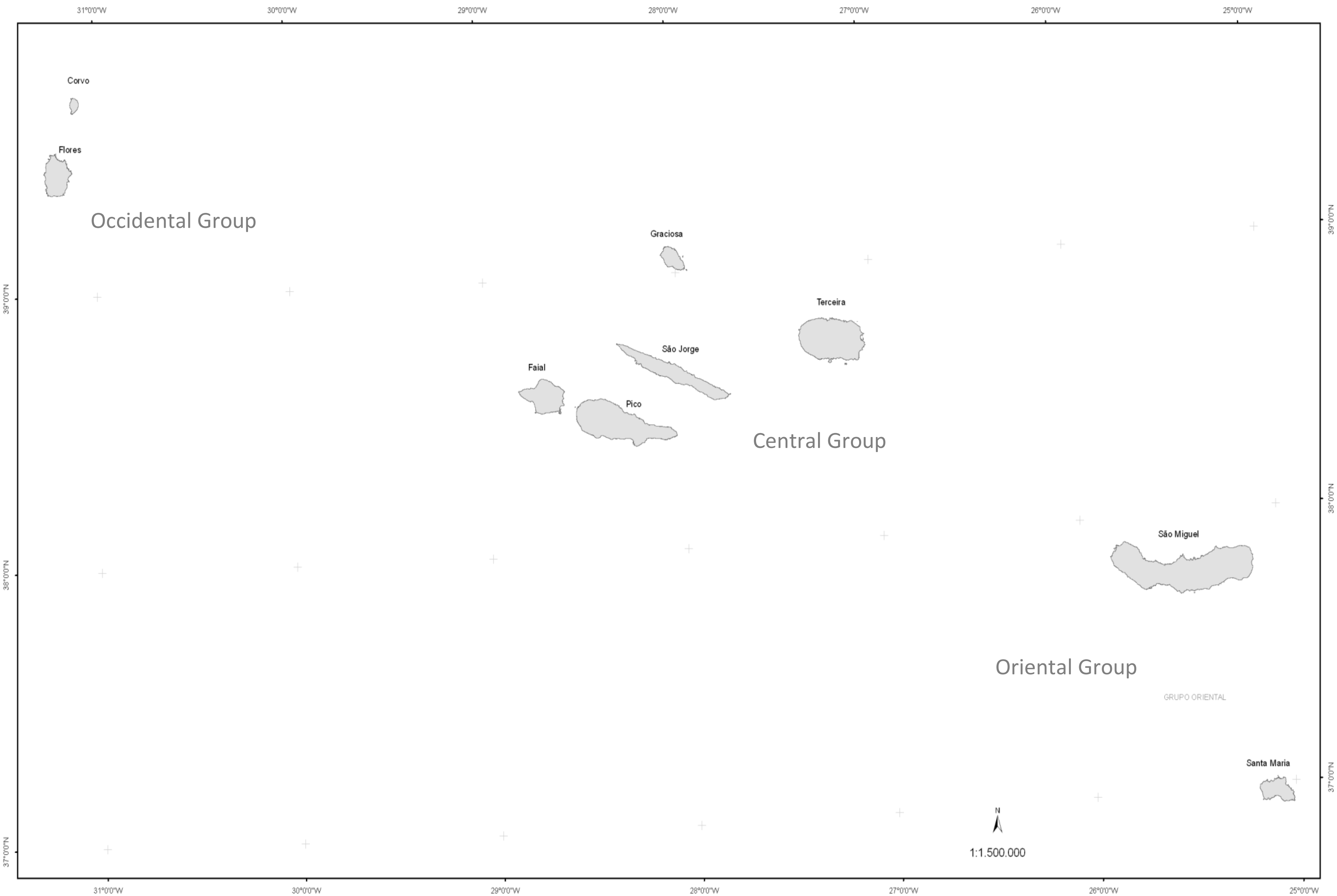
A SOFTWARE FRAMEWORK FOR MEASURING EFFICIENCY

Veska Noncheva (veska@uac.pt)
Armando Mendes (amendes@uac.pt)
Emiliana Silva (emiliana@uac.pt)

CEEAplA – Universidade dos Açores
Portugal

Introduction

- Azores is a Portuguese insular territory where the main economic activity is dairy and meat farming.
- Dairy policy depends on Common Agricultural Policy of the European Union and is still limited by quotas.
- In this context, decision makers need knowledge for deciding the best policies in promoting quality and best practices.



Corvo
Flores
Occidental Group
Graciosa
Terceira
São Jorge
Faial
Pico
Central Group
São Miguel
Oriental Group
GRUPO ORIENTAL
Santa Maria
N
1:1.500.000

Objective

- The goal of our work is to provide Azorean Government with a reliable tool for measurement of productive efficiency of the farms.
- The proposed approach is implemented in R statistical software. The output of the computer program is self explanatory.

PAR

- The “Productivity Analysis with R” (PAR) framework establishes a user-friendly DEA environment with special emphasis on variable selection and aggregation, and summarization and interpretation of the results.
- The starting point is the following R packages: CCA, DEA and FEAR .

DEA

- Data Envelopment Analysis (DEA) is a way of determining the efficiency for a group of farms called decision making units (DMUs) when measured over a set of multiple input and output variables.
- For a given set of input and output variables DEA produces a single comprehensive measure of performance called *efficiency score*.

DEA limitations

- Since DEA is an extreme point technique, noise such as measurement error can cause problems.
- When the number of inputs or outputs is increased, the number of observations must increase at an exponential rate.

One of the most important steps in the modelling using DEA is the choice of input and output variables.

Methodology

- Variable selection is crucial to the process as the omission of some of the inputs can have a large effect on the measure of efficiency. It is now recognized that improper variable selection often results in biased DEA evaluation results.
- The attention to variable selection is particularly crucial since the greater the number of input and output variable, the less discerning are the DEA results.

Methodology

- Several methods have been proposed that involve the analysis of correlation among the variables, with the goal of choosing a set of variables that are not highly correlated with one another.
- Unfortunately, studies have shown that these approaches yield results which are often inconsistent in the sense that removing variables that are highly correlated with others can still have a large effect on the DEA results.

- Several methods for variable selection have been proposed.
- However, there is no consensus on how best to limit the number of variables.

Variable Selection in PAR

- In our work, we propose Canonical Correlation Analysis (CCA) to be used in order the most appropriate variables to be selected. In our approach we apply CCA to select both input and output variables and to get final input and output sets, respectively.

Canonical Correlation Analysis

- CCA is a multidimensional exploratory statistical method. A canonical correlation is the correlation of two latent variables, one representing a set of independent variables, the other a set of dependent variables. The canonical correlation is optimized such that the linear correlation between the two latent variables (called canonical variates) is maximized.

CCA and Variable Selection

- We interpret the relations of the original variables to the canonical variates in terms of the correlations of the original variables with the canonical variates – that is by the structure coefficients.
- The absolute values of the structure coefficients are closely related to the strength of the relation between input and output sets of variables in a production process.
- We chose both input and output variables with the biggest absolute values of their structure coefficients to be included in the DEA model.

The Mathematical Intuition

- DEA

  to maximize the ratio of a weighted sum of outputs to a weighted sum of inputs

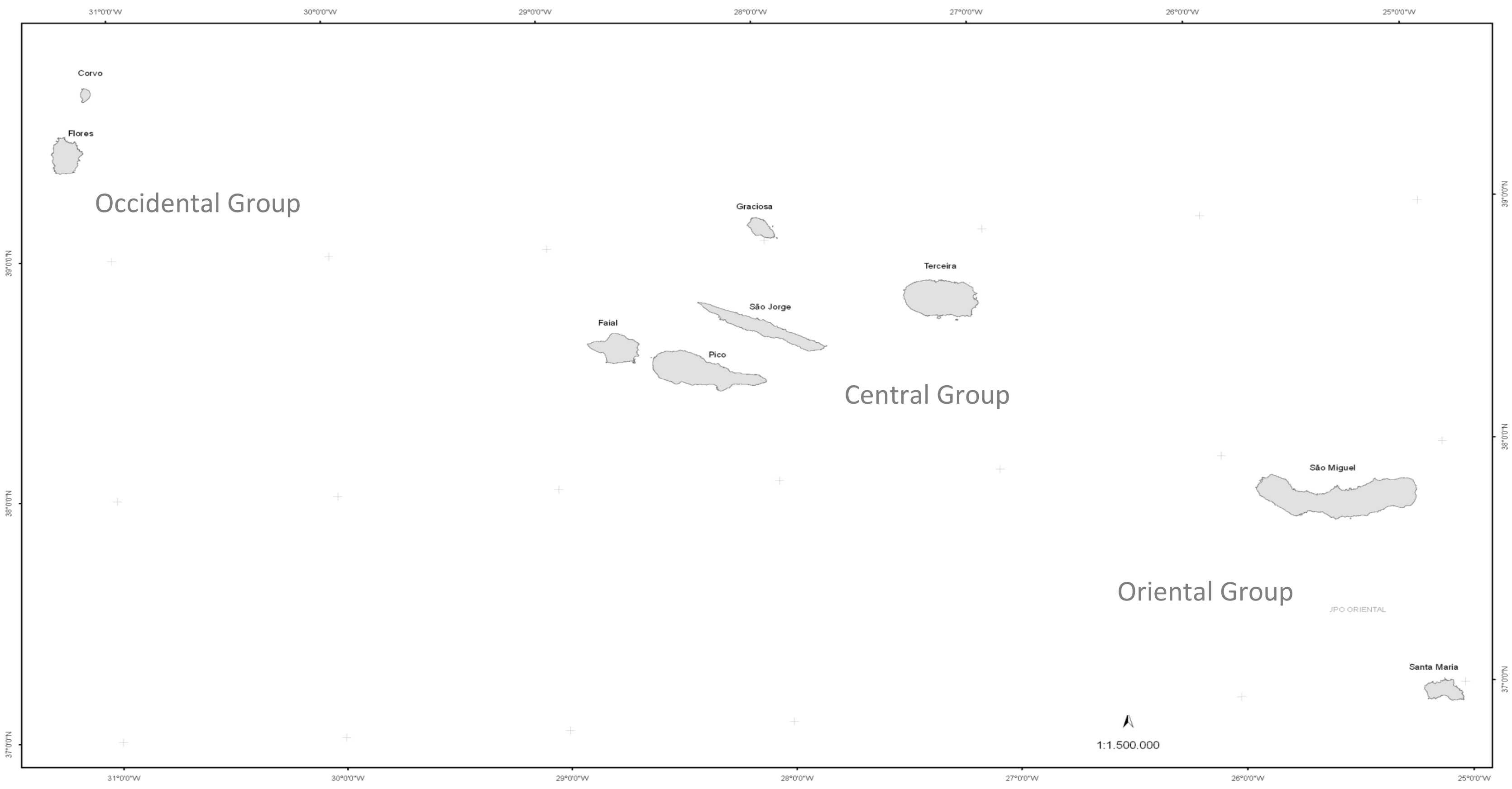
$$\frac{\sum_j b_{jd} y_{id}}{\sum_i a_{id} x_{id}}$$

- CCA

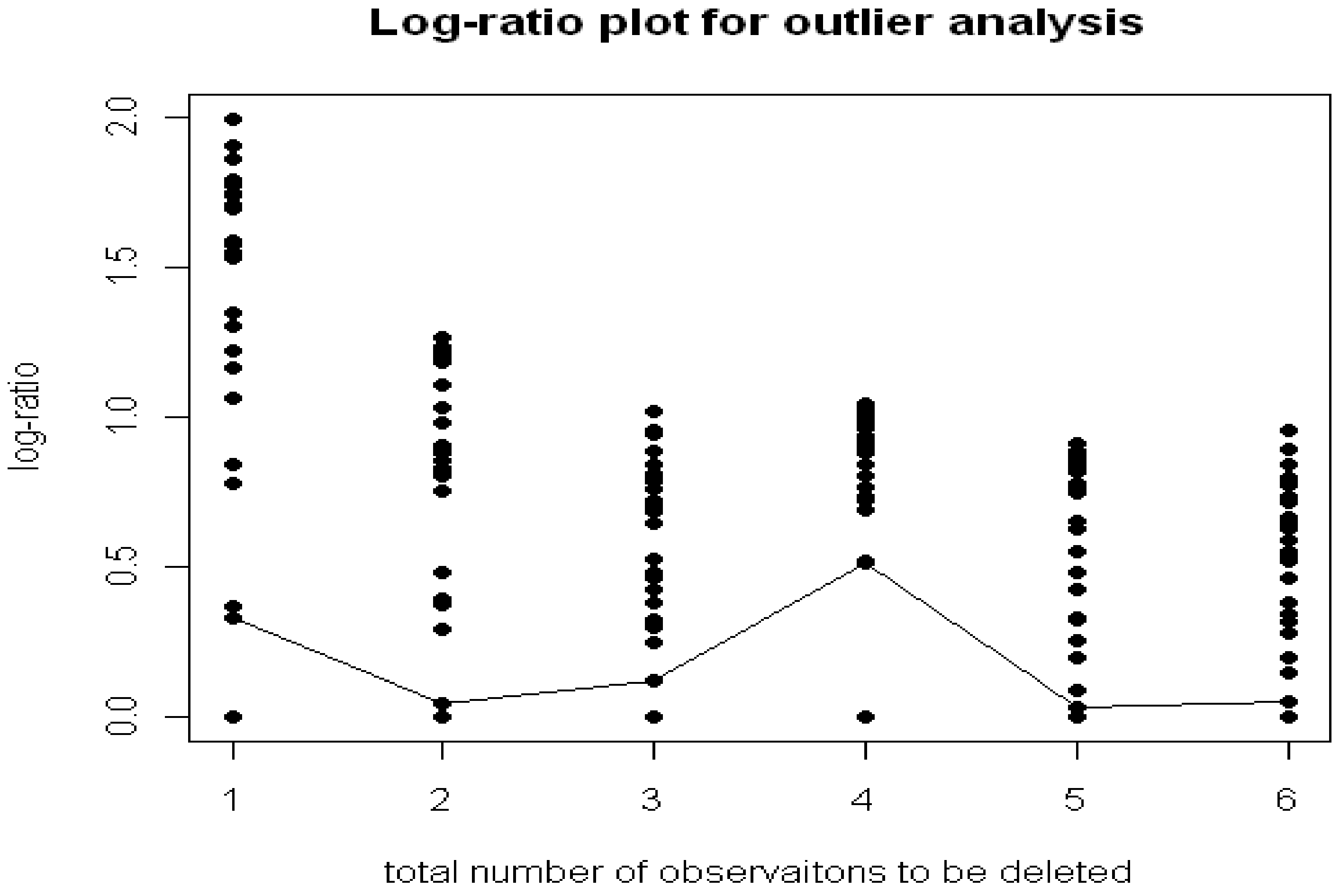
  to maximize the correlation

$$cor(\sum_i u_{id} x_{id}, \sum_j b_{jd} y_{id})$$

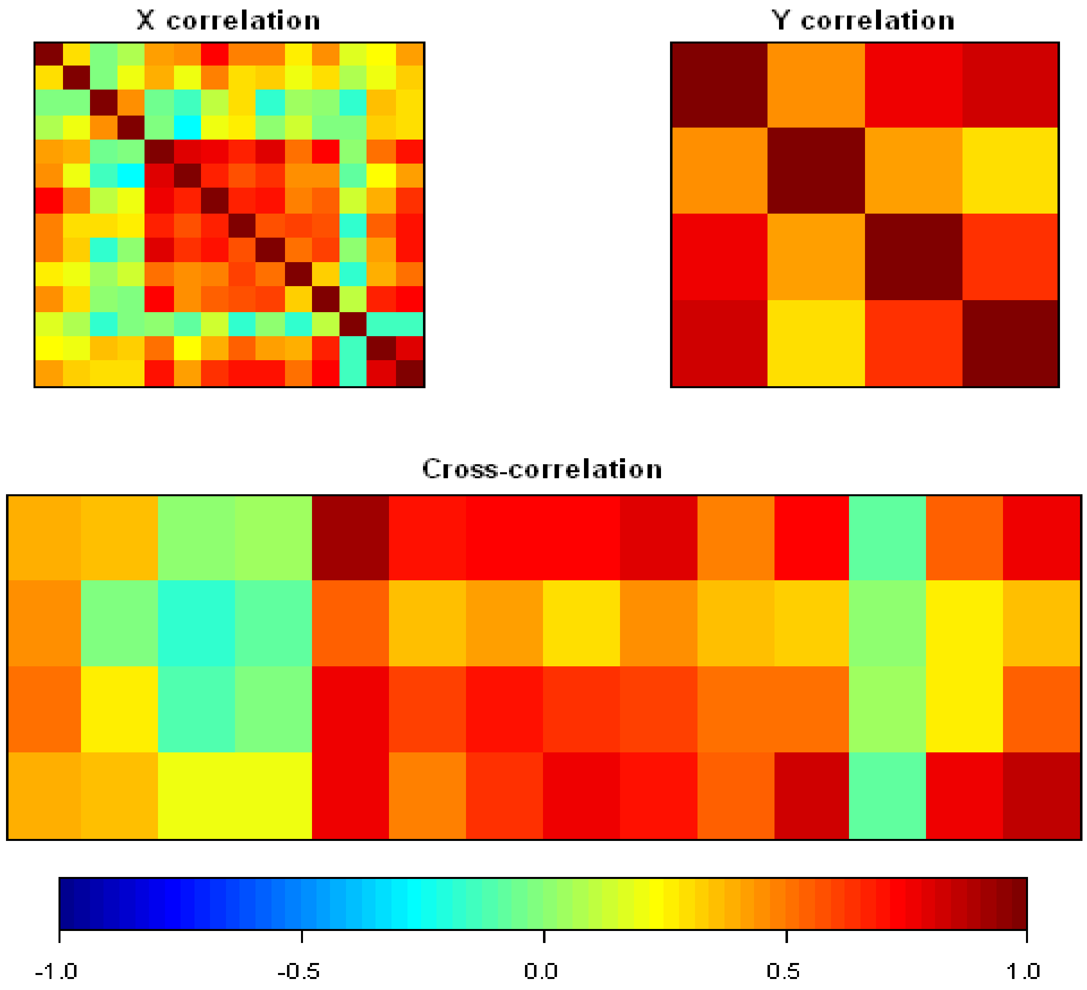
Terceira Problem



Terceira Problem – Outlier detection



Terceira Problem – Correlations Matrices



Terceira Problem -CCA

	Input variables (X)	structure weights	structure weights
1	EquipmentRepair	-0.463883	-0.447400
2	Oil	-0.367265	-0.354215
3	Lubricant	-0.084217	-0.081225
4	EquipmentAmortization	-0.122065	-0.117728
5	**AnimalConcentrate**	**-0.923175**	**-0.890372**
6	VeterinaryAndMedicine	-0.659249	-0.635824
7	OtherAnimalCosts	-0.756439	-0.729560
8	PlantasSeeds	-0.807528	-0.778834
9	Fertilizers	-0.819276	-0.790165
10	Herbicides	-0.575547	-0.555096
11	LandRent	-0.831971	-0.802408
12	Insurance	0.071831	0.069279
13	AreaDimension	-0.706559	-0.681453
14	**DairyCows**	**-0.884397**	**-0.852972**

	Output variables (Y)	structure weights	structure weights
1	**Milk**	**-0.923248**	**-0.957263**
2	Cattle	-0.486988	-0.504929
3	ProductionSubsidy	-0.728532	-0.755373
4	**FactorsSubsidy**	**-0.908093**	**-0.941549**

The subsidies are important for the dairy farms, and in 2004 they were about 61.6% of all profit. Some of these subsidies are compensations for low selling prices received by farmers. There are also subventions to improve ecological production.

Terceira Problem

- The chosen input variables are *AnimalConcentrate* and *DairyCows*.
- The chosen output data for DMUs are *Milk* and *FactorsSubsidy*.

Terceira Problem -Results

- > summary.BCC.IO (inputs=input.2, dmu.numb, inputs.numb, BCC.io.version, eps=0.0000001)
- $fully.efficient
- [1] 5 6 12 13 14 20
- $radial.efficient.only
- [1] NA
- $inefficient.zero.slack
- [1] 2 4 9 16 21 24
- $inefficient.nonzero.slack
- [1] 1 3 7 8 10 11 15 17 18 19 22 23 25 26 27 28 29 30

Terceira Problem –Results

- > report.BCC.IO (inputs=input.2, dmu.numb, inputs.numb, BCC.io.version, eps=0.0000001)
- [1] " The optimal solution for DMU24 is:"
- [1] "theta* = 0.887 Hence DMU24 is technically inefficient. (Zerro slacks)"
- [1] "The input values needed to bring DMU24 into efficient status are the following:"
- [1] " projection X1 = 8380.99(input x1=9452.19);
  projection X2 = 18.62 (input x2=21);"
- [1] "Reference set = {DMU12;DMU14;DMU20;DMU13;}"

Terceira Problem –Results

- In the absence of environmental differences (i.e. differences in soil quality, animal genetics, climate) and errors in the measurement of inputs and outputs, **pure technical inefficiency would reflect departures from best-practice farm management.** The way to eliminate this latter source of inefficiency is to form a benchmarking partnership with relevant best-practice farms with a view to identifying and then emulating their farm management practices.

The result includes measures of each farm's

- scale efficiency (SE),
- pure technical efficiency,
- overall technical efficiency and
- identification of its best-practice benchmark

Terceira Problem -Results

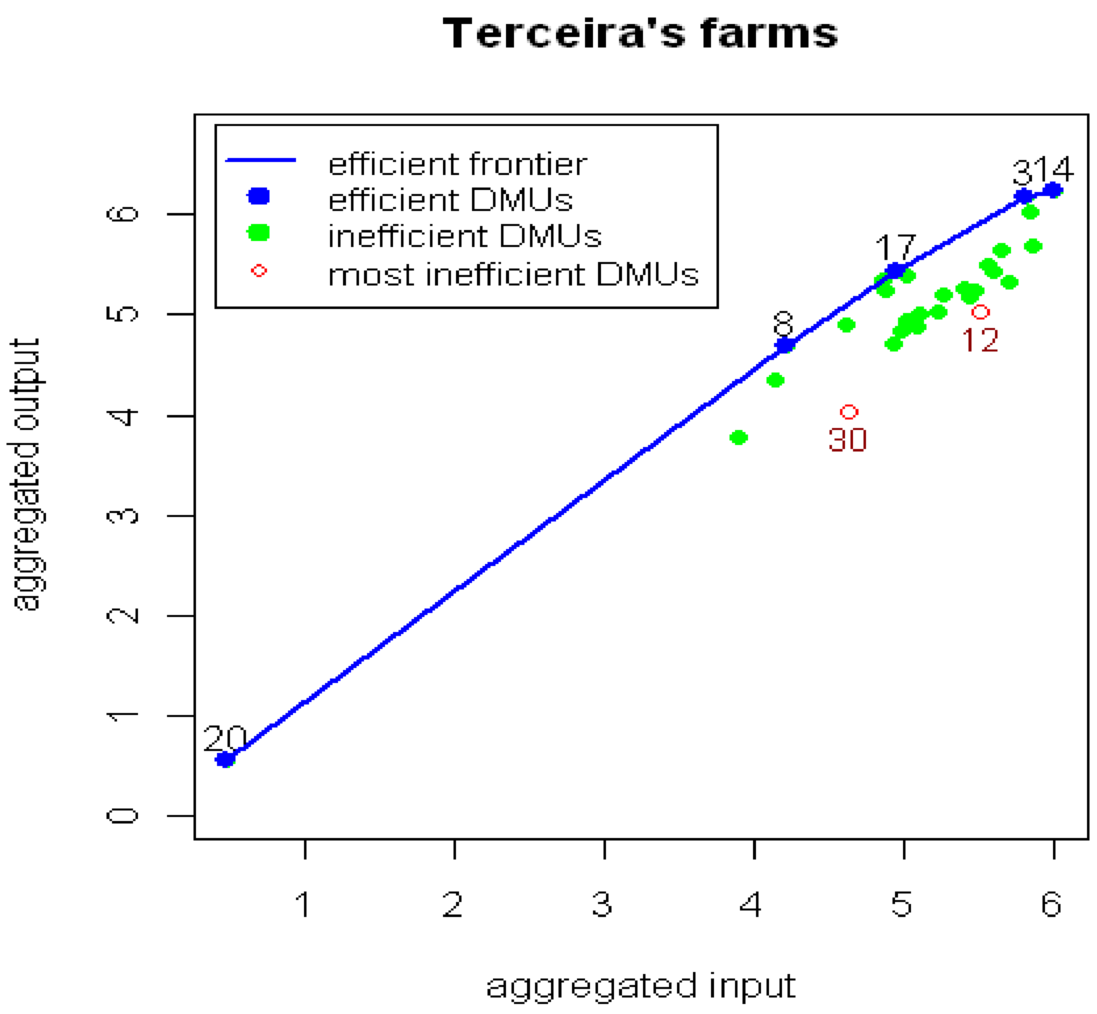
- The farms 12, 13, 14, and 20 are scale-efficient. This means that the farms are operating at its optimum size and hence that the productivity of inputs cannot be improved by increasing or decreasing this kind of production factors.
- The farms 1, 3, 7, 10, 11, 15, 18, 22, 23, 25, 27, 28 and 29 can improve the productivity of inputs and thereby reduce unit costs.
- The others 13 farms are too big and so, the farmer can improve the productivity of inputs and hence reduce unit costs by reducing the size of the farm (the number of cows, the pasture, etc.). The reference set for each inefficient farm identifies potential benchmark partners.

- On the basis of this study, senior management can only make some preliminary conclusions. The extent to which any of these results can be interpreted in a context which is relevant to managing the farms, is not clear at this point. Extensive and detailed subsequent analysis of pointed farms is required before any sound decision can be made.

More Results

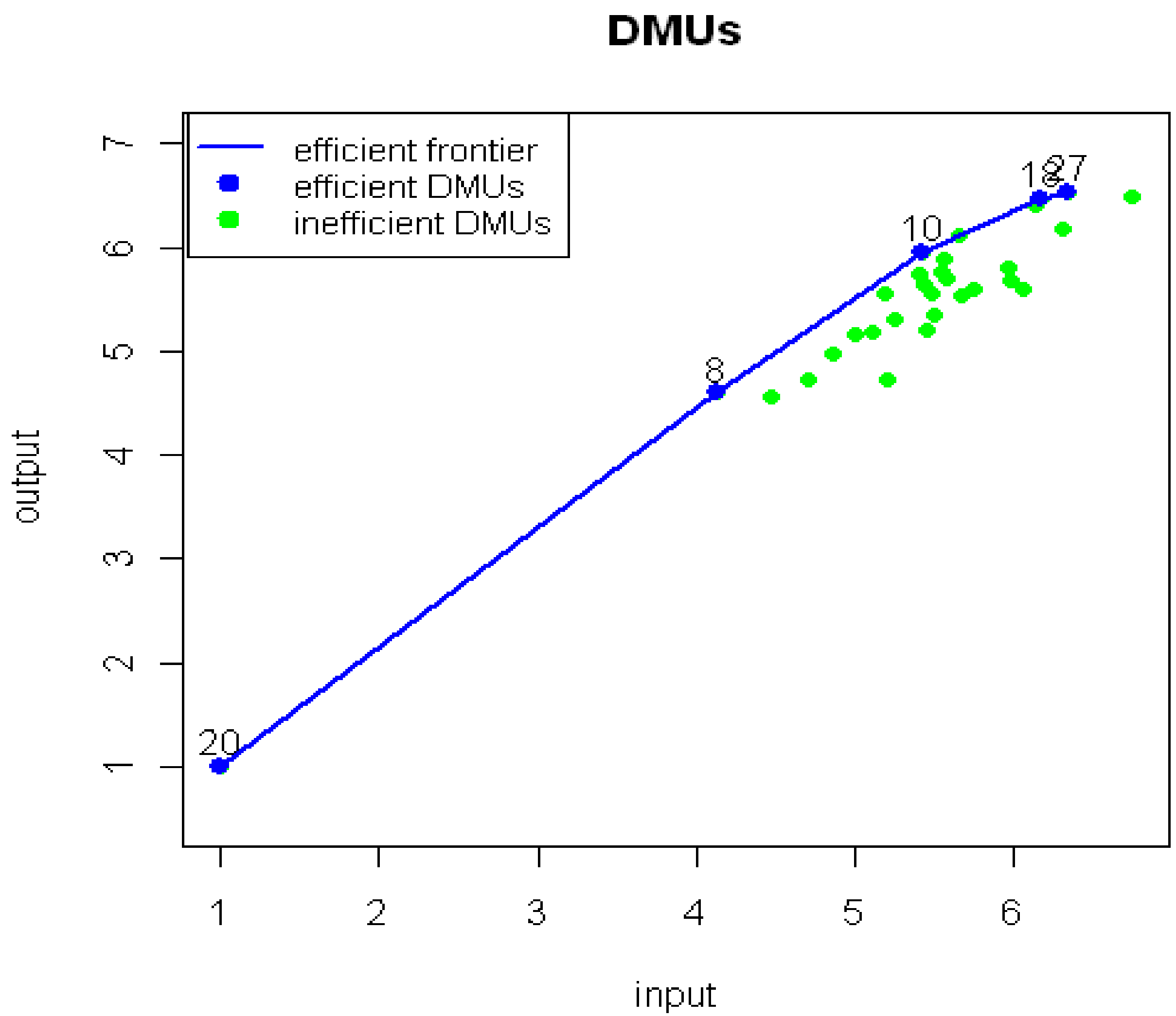
- ***Variable aggregation***

Results – Azores Data

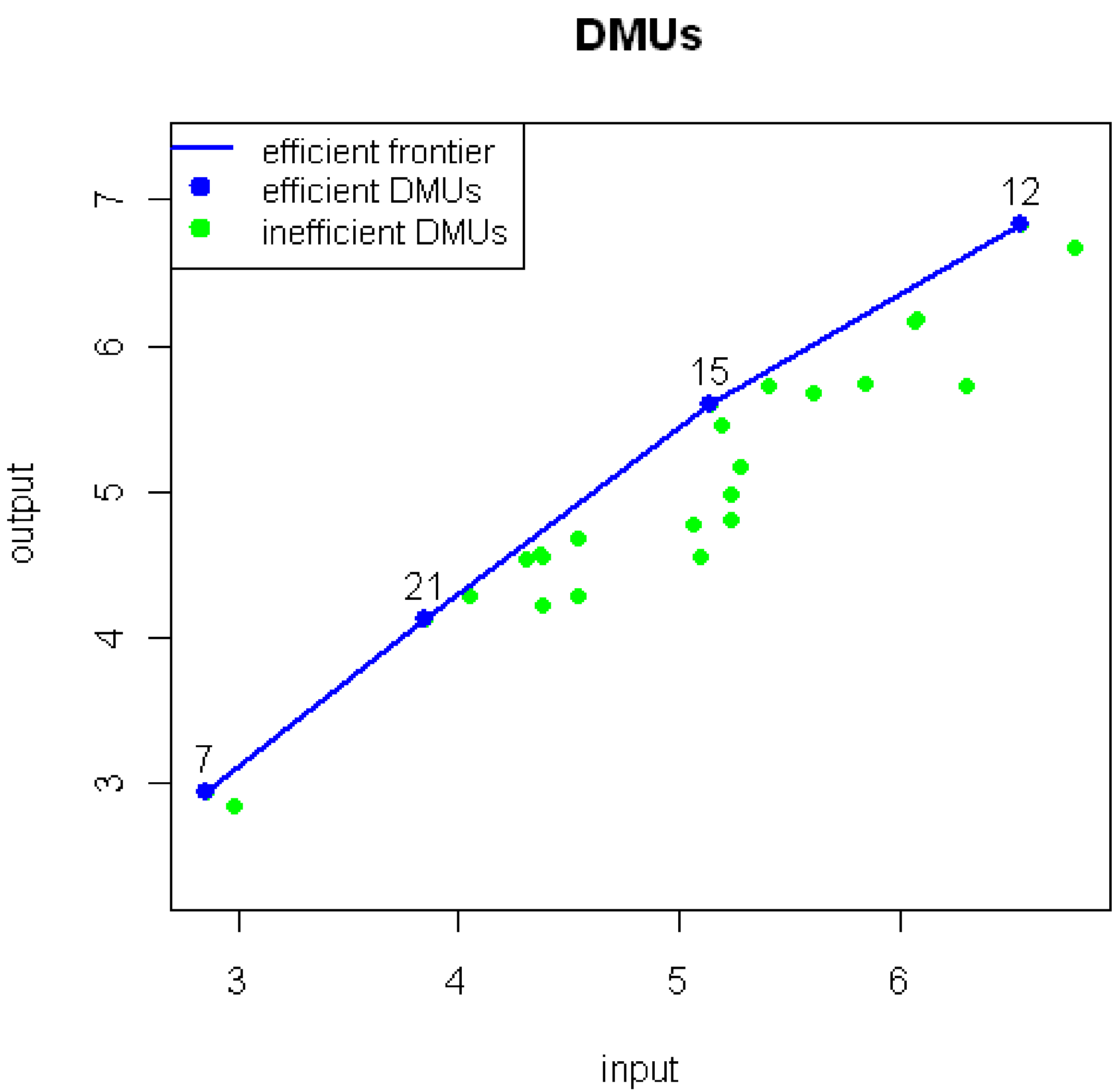


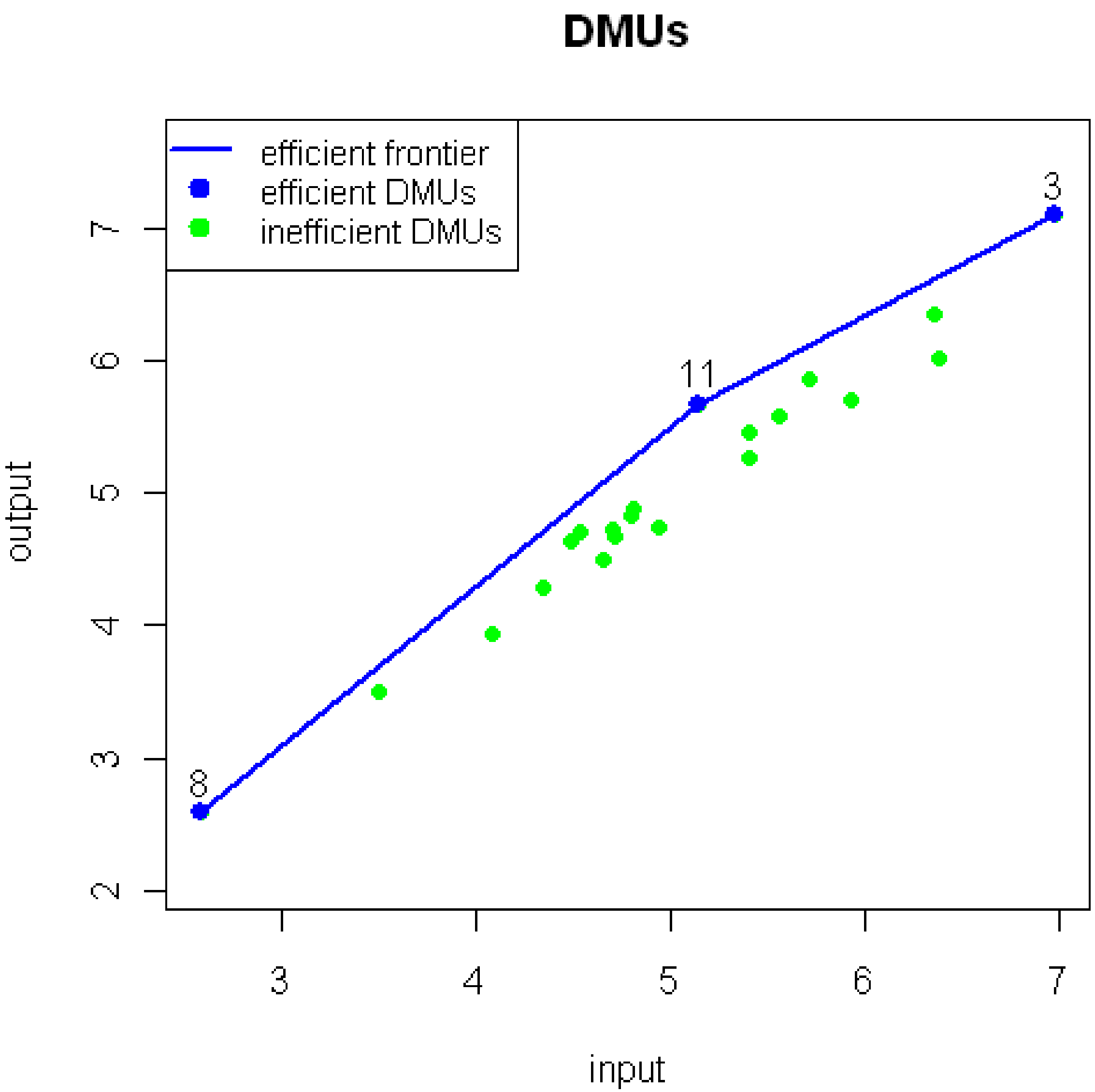
Summary Function

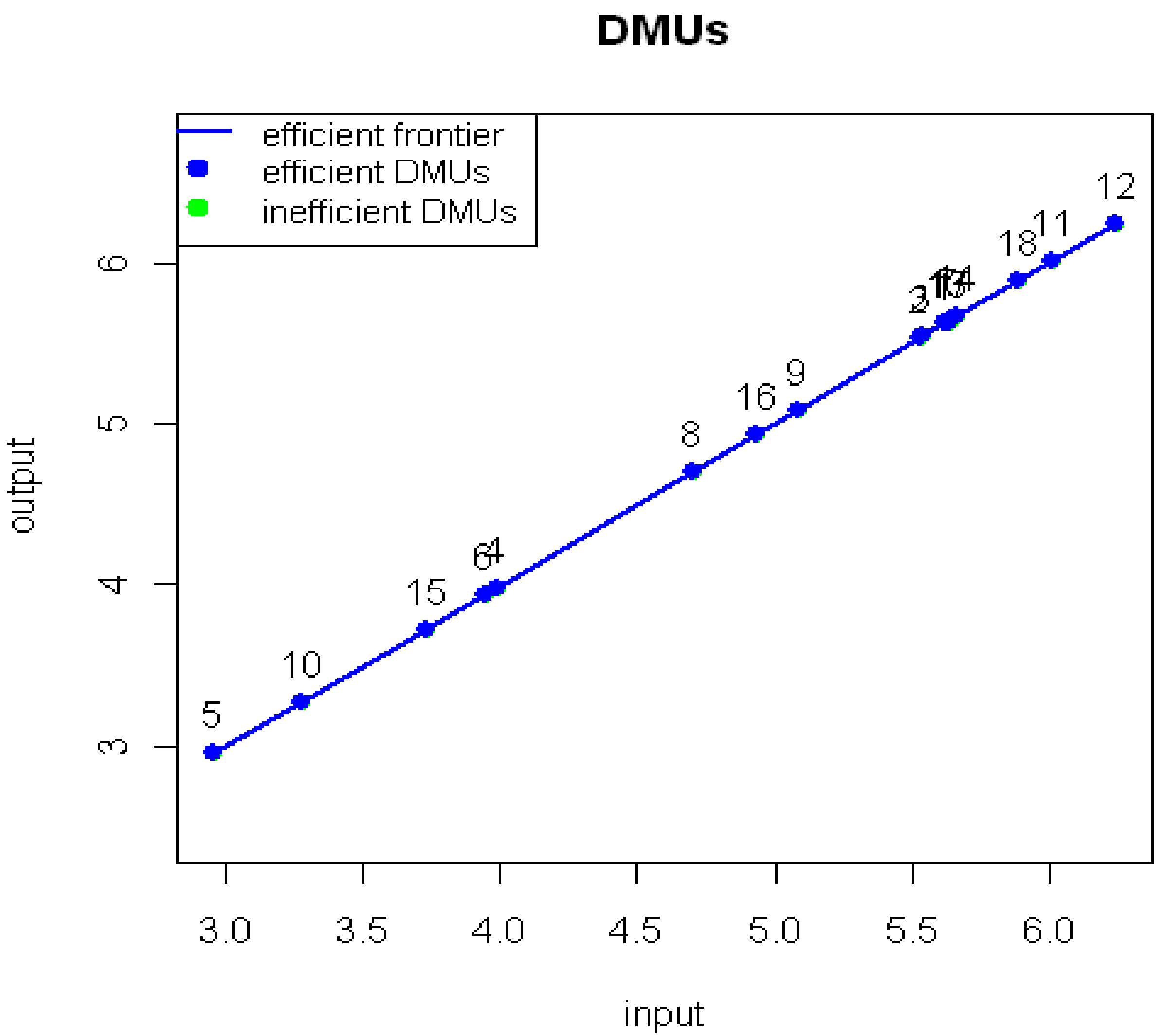
```
> summary.BCC.IO (inputs, dmu.numb, inputs.numb,
   BCC.io.version,  eps=0.0000001)
$fully.efficient
[1]  8 10 18 20 27
$radial.efficient.only
[1] NA
$inefficient.zero.slack
[1]  1  2  3  4  5  6  7  9 11 12 13 14 15 16 17 19 21
   22 23 24 25 26 28 29 30
$inefficient.nonzero.slack
[1] NA
```



DMUs
efficient frontier
efficient DMUs
inefficient DMUs
20
8
10
18
27
output
input
1
2
3
4
5
6
7


DMUs
efficient frontier
efficient DMUs
inefficient DMUs
12
15
21
7
output
input
3
4
5
6
7


DMUs
efficient frontier
efficient DMUs
inefficient DMUs
3
11
8
output
2
3
4
5
6
7
3
4
5
6
7
input


DMUs
efficient frontier
efficient DMUs
inefficient DMUs
output
input
3.0
3.5
4.0
4.5
5.0
5.5
6.0
3
4
5
6
5
10
15
8
16
9
18
11
12

Terceira Problem –SDM

Level 1		Level 2		Level 3	
DMUs	Efficiency	DMUs	Efficiency	DMUs	Efficiency
1	1.000.00				
2	0.919509	1	1.000.000		
3	1.000.00				
4	0.751622	2	1.000.000		
5	0.706989	3	1.000.000		
6	0.794443	4	1.000.000		
7	0.837575	5	1.000.000		
8	0.508521	6	0.635923	1	1.000.000
9	0.862141	7	1.000.000		
10	1.000.00				

DMUs	Efficiency	DMUs	Efficiency	DMUs	Efficiency
11	0.90467	8	1.000.000		
12	1.000.00				
13	1.000.00				
14	1.000.00				
15	1.000.00				
16	0.74788	9	0.909254	2	1.000.000
17	0.54428	10	1.000.000		
18	1.000.00				
19	0.44322	11	0.572976	3	0.876868
20	1.000.00				

DMUs	Efficiency	DMUs	Efficiency	DMUs	Efficiency
21	0.784165	12	0.870596	4	1.000.000
22	0.945037	13	1.000.000		
23	1.000.00				
24	0.896165	14	1.000.000		
25	0.912602	15	1.000.000		
26	0.740936	16	0.863479	5	1.000.000
27	1.000.00				
28	1.000.00				
29	0.882734	17	1.000.000		
30	1.000.00				

11	projection X1 = 18138.47 (input x1=21922.77); projection X2 = 4408.57 (input x2=4873.07); projection X3 = 1596.95 (input x3=3056.15); projection X4 = 30.76 (input x4=34)
16	projection X1 = 6822.06 (input x1=7502.92); projection X2 = 4052.59 (input x2=5389.48); projection X3 = 829.45 (input x3=912.23); projection X4 = 25.46 (input x4=28)
17	projection X1 = 6391.56 (input x1=14535.37); projection X2 = 1091.40 (input x2=2005.19); projection X3 = 419.63 (input x3=770.98); projection X4 = 16.33 (input x4=30)
19	projection X1 = 7832.20 (input x1=8932.01); projection X2 = 5251.49 (input x2=5988.91); projection X3 = 953.21 (input x3=2197.39); projection X4 = 28.05 (input x4=36)
21	projection X1 = 10184.44 (input x1=14551.7); projection X2 = 3557.92 (input x2=4086.76); projection X3 = 919.37 (input x3=1056.02); projection X4 = 23.51 (input x4=27)
22	projection X1 = 9368.36 (input x1=9913.21); projection X2 = 3395.44 (input x2=3592.92); projection X3 = 787.32 (input x3=833.11); projection X4 = 21.74 (input x4=23)

Conclusion - PAR Methodology

- PAR is very flexible, extensible software based on CCA and DEA models, implemented as CCA and FEAR packages in R.
- The cost of this flexibility is that the user must type commands at a command-line prompt.
- The CCA provides an aggregation of both input and output units and then DEA provides efficient units.

Future Research

- The effects of the input aggregation on efficiency indicators have not been investigated.
- Some critics argue that the linear aggregation of inputs introduces a bias in the efficiency measurement.
- Estimating the aggregation bias is a question of our future theoretical research.

Final Comments

- In PAR methodology CCA provides an aggregation of both input and output units and then DEA provides efficient units.
- The effects of the variable selection and aggregation on efficiency indicators have not been investigated.