Department of Computer, Control, and Management Engineering Antonio Ruberti

Sapienza Università di Roma

RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

# Methodological course on Advanced Benchmarking Models and Techniques

# Laboratory Session: Productivity and Efficiency Analysis with MATLAB

22/03/2021

| Trainers | Email |
|---|---|
| **Giammarco Quaglia** | **Email: Quaglia@diag.uniroma1.it** |
| **Simone Di Leo** | **Email: dileo@diag.uniroma1.it** |

# Index



# Additional Materials

# Reading Data in MATLAB

# Reading Data in MATLAB

In MATLAB there are several options for loading data. The main ones are:

- Loading data from file:

  A = importdata(___,delimiterIn,headerlinesIn)

  loads data from ASCII file, filename, or the clipboard, reading numeric data starting from line headerlinesIn+1.

- Loading data statically in M scripts:

  This type of data loading is used to allows faster execution of scripts.

*Important: In Matlab the decimal separator is the point, not the comma!*

# Data Analysis

# Descriptive Statistics

**Central Tendency Measures**

- Mean
- Median
- Quartile

**Dispersion Measures**

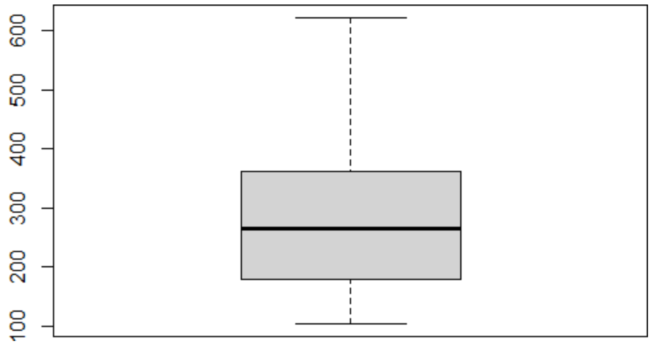
- Interquartile Range
- Variance
- Standard Deviation

**Correlation**

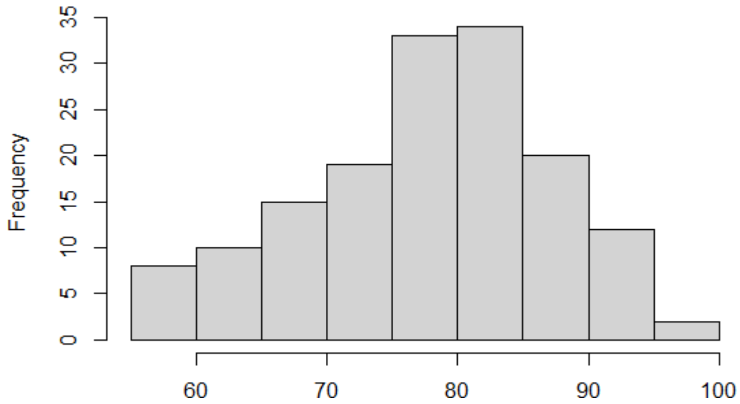
- Correlation between Inputs
- Correlation between Outputs
- Correlation Inputs/Outputs

# Visualizations

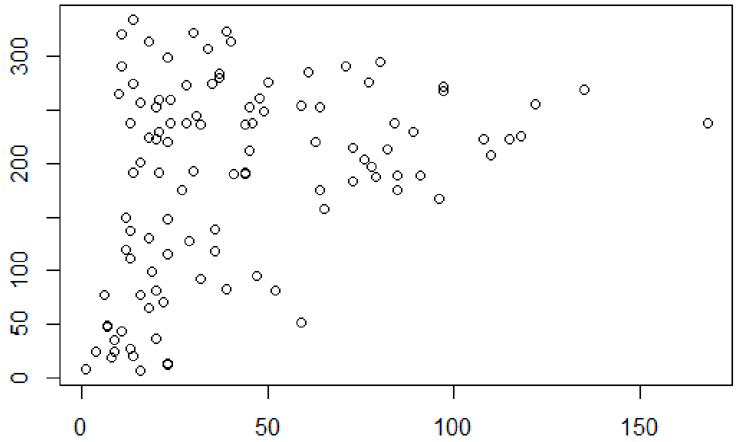
## Boxplots



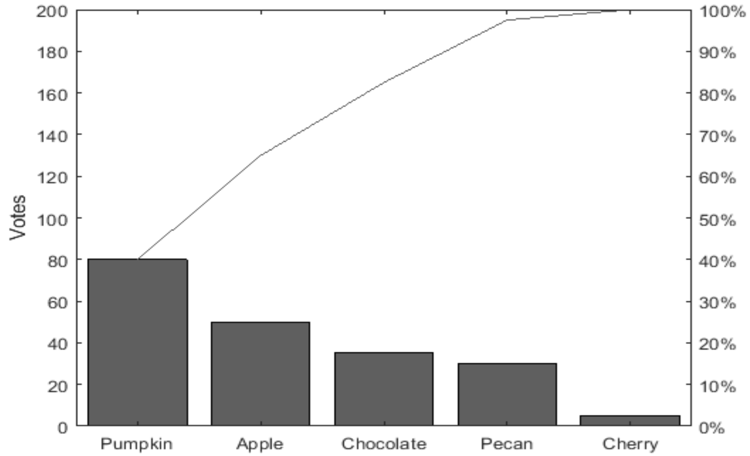

## Histograms



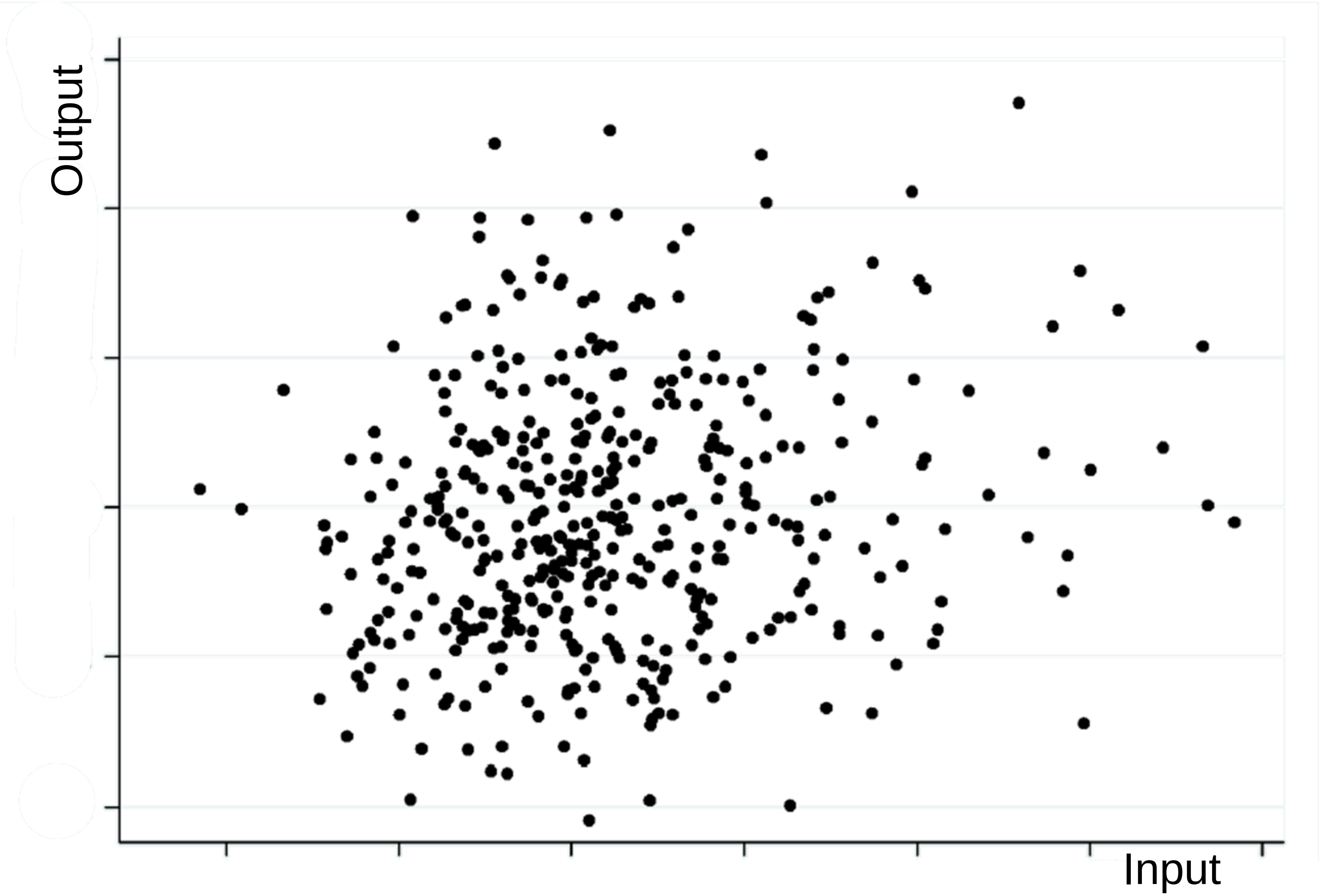

## Dispersion Graphs



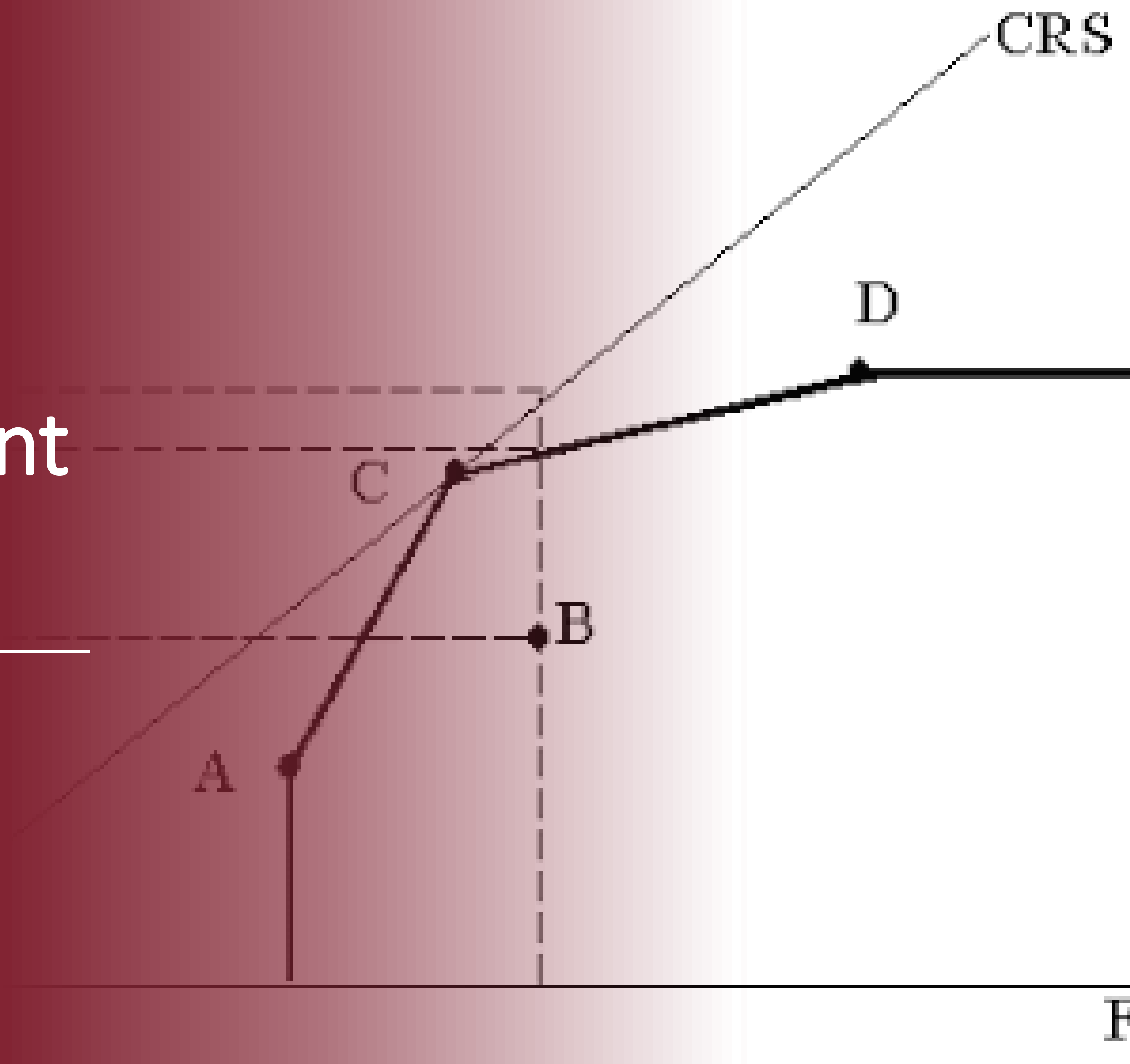

## Pareto Plot



# Example: dispersion graph



# Data Envelopment Analysis



# DEA

DEA is a technique for measuring the performance of production units. DEA allows to evaluate the efficiency of a production unit(DMU - Decision Making Unit) relative to a given set of production units chosen for comparison.

All units in the set must have the ratio between input and output less or equal to one or that their difference is less or equal to zero, because  each DMUs can’t produce a set of output higher than the input **(convexity of the set)**

Unlike the "statistical" approach that compares production units with the average production unit, DEA compares each DMU with the most efficient DMUs and evaluates their relative efficiency. Efficient DMUs and assesses their relative efficiency.

Thanks to this approach, DEA allows the use of multiple inputs and outputs

# DEA LP Formulation

| Orientation | Primal | Dual |
|---|---|---|
| **Input Oriented** | $\max \quad z = uy_j - (u_j)^*$<br>$\text{s.t.} \quad vx_j = 1$<br>$-vX + uY - (u_j e)^* \leq 0$<br>$v \geq 0, \quad u \geq 0, \quad (u_j \text{ free in sign})^*$ | $\min \quad \theta$<br>$s.t. \quad \theta x_j - X\lambda \geq 0$<br>$Y\lambda \geq y_j$<br>$(e\lambda = 1)^*$<br>$\lambda \geq 0$ |
| **Output Oriented** | $\min \quad z = vx_j - (v_j)^*$<br>$s.t. \quad uy_j = 1$<br>$vX - uY - (v_j e)^* \geq 0$<br>$v \geq 0, \quad u \geq 0, \quad (v_j \text{ free in sign})^*$ | $\max \quad \eta$<br>$s.t. \quad x_j - X\lambda \geq 0$<br>$\eta y_j - Y\lambda \leq 0$<br>$(e\lambda = 1)^*$<br>$\lambda \geq 0$ |

# DEA Radial Measure: input vs output

The linear programming problem can be resolved in two ways:

- Maximizing the numerator and fixing the denominator (output-oriented method)
- Keeping the numerator constant and minimizing the denominator (input-oriented method).

The distinction is important because from it descends the form of efficiency that is being evaluated.

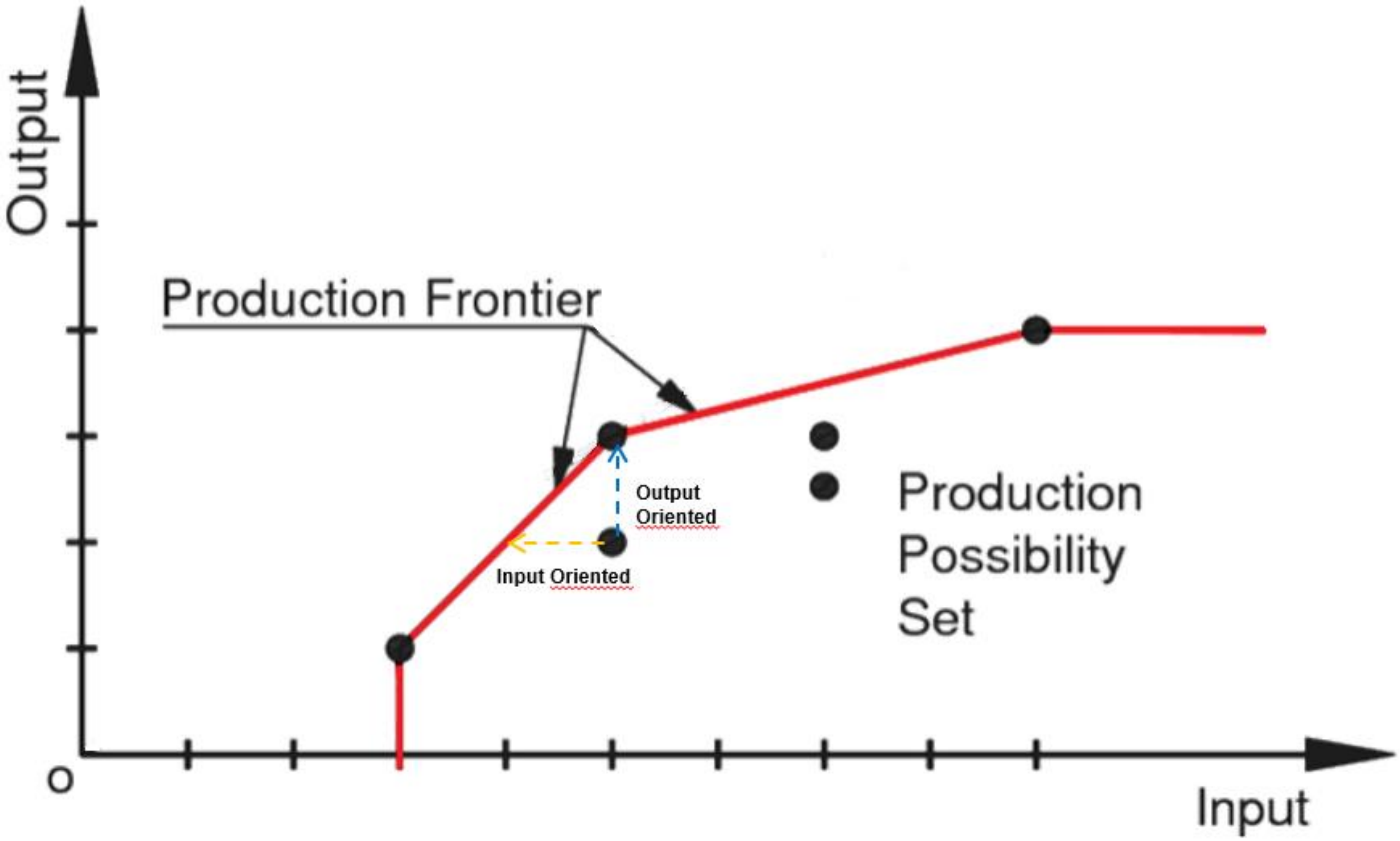
Input oriented:

$$\theta(x, y) = \inf\{\theta \mid (\theta x, y) \in \Psi\} \leq 1$$

Output oriented:

$$\lambda(x, y) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\} \geq 1$$

# DEA input vs output plot



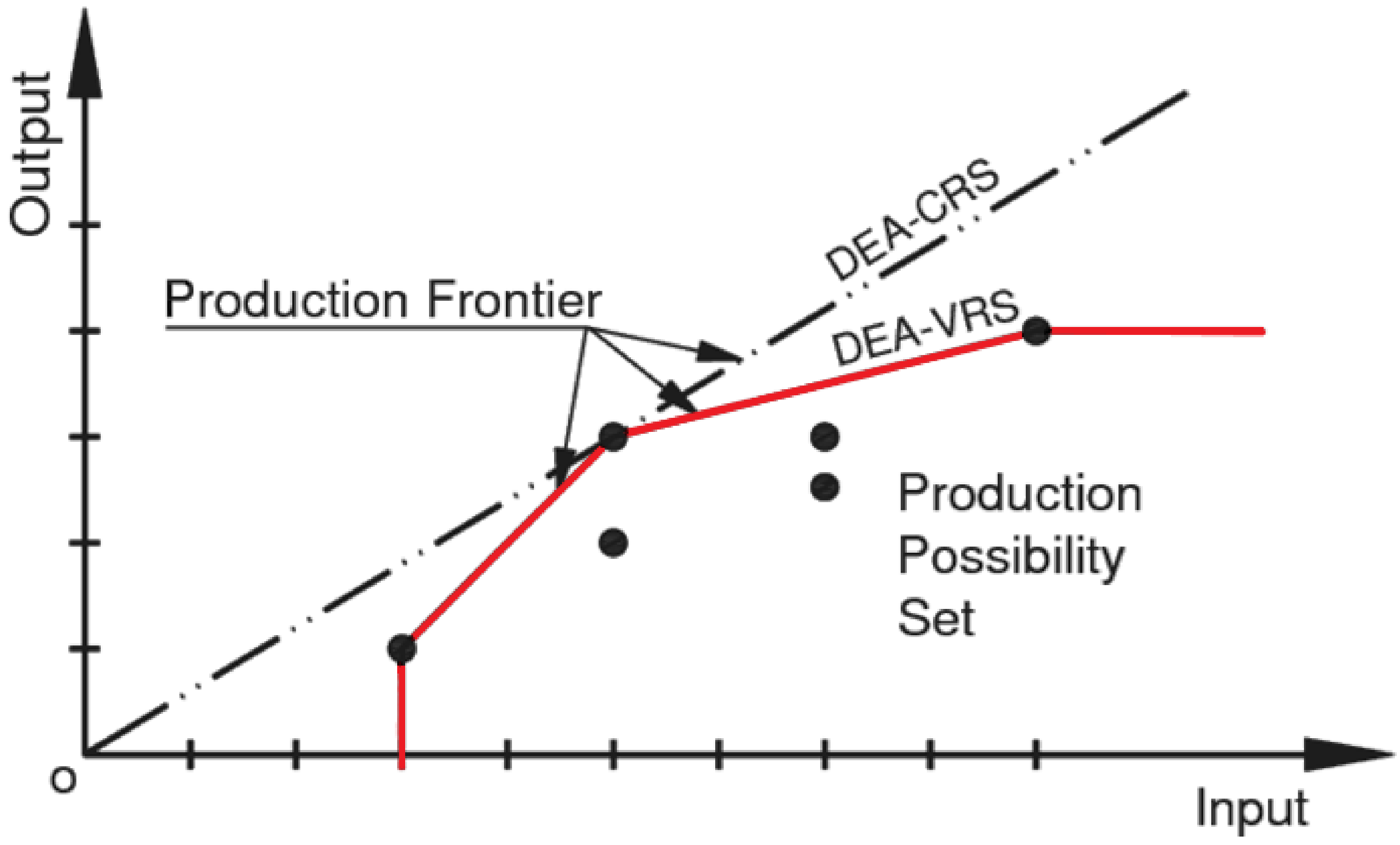
# DEA CRS vs VRS

The envelopment surface will differ depending on the scale assumptions that underpin the model.  Two scale assumptions are generally employed: Constant returns to scale (CRS) and Variable returns to scale (VRS).

CRS reflects the fact that output will change by the same proportion as inputs are changed (e.g. a doubling of all inputs will double output).

VRS reflects the fact that production technology may exhibit increasing, constant and decreasing returns to scale.

# DEA CRS vs VRS plot



# DEA Toolbox

The DEA function into the DEA Toolbox computes data envelopment analysis radial model.

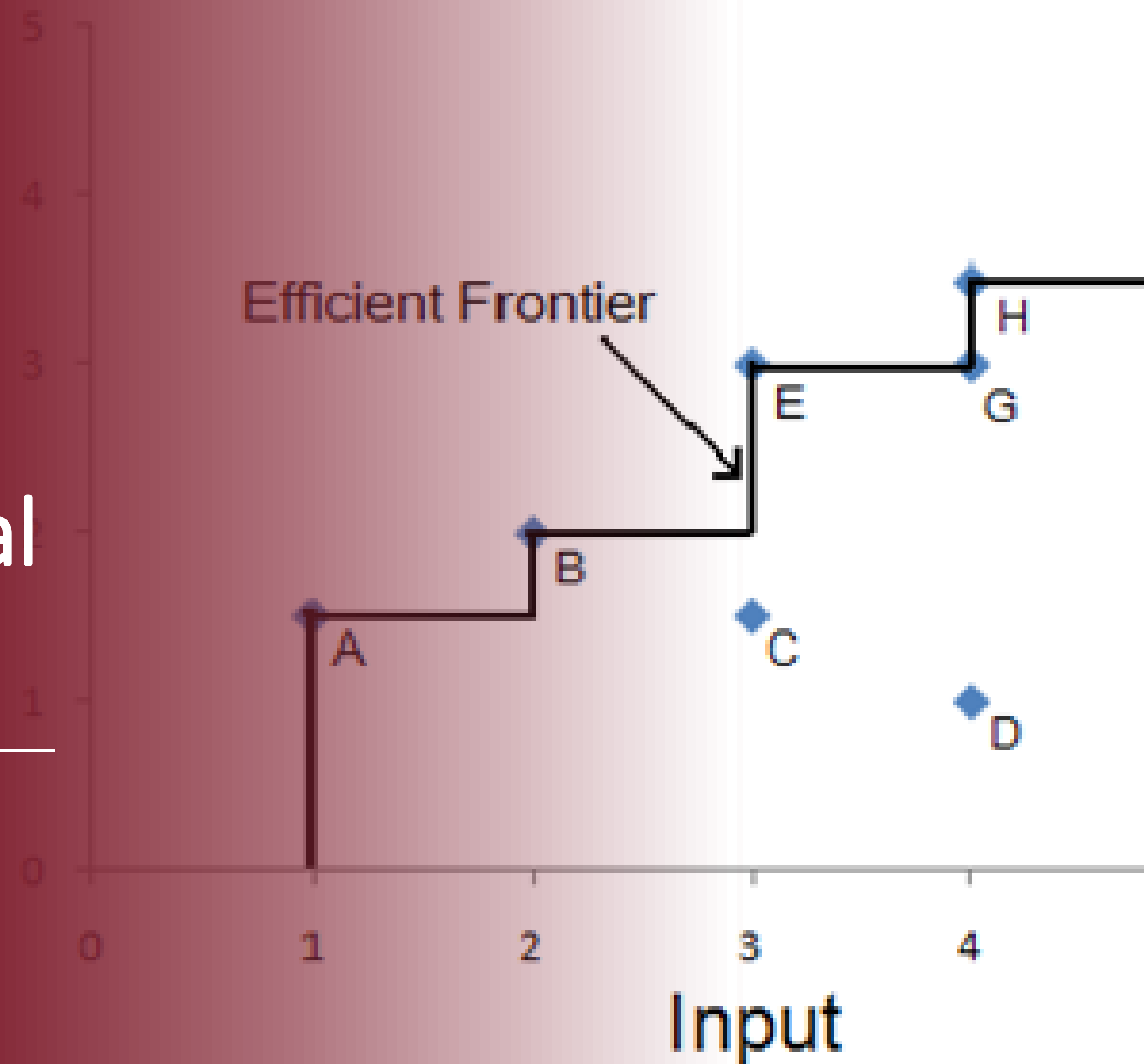
out = DEA(X, Y, Name, Value) computes data envelopment analysis model with inputs X and outputs Y. Model properties are specified using one or more Name ,Value pair arguments.

Properties:

- - 'orient': orientation. Input oriented 'io', output oriented 'oo'
- - 'rts': returns to scale. Constant returns to scale 'crs', variable returns to scale 'vrs'.
- - 'names': DMU names.
- - 'secondstep': 1 to compute input and output slacks. Default is 1.

Example: oo_vrs = dea(X, Y, 'orient', 'oo', 'rts', 'vrs');

# Free Disposal Hull



# FDH

Free Disposal Hull (FDH) model is a non-parametric method to measure the efficiency of production units.

FDH model relaxes the convexity assumption of basic DEA models. The computational technique to solve FDH program considers the mixed integer programming problem compared to the DEA model with a linear programming problem.

All the observations that have more efficient observations are considered inefficient , i.e. when higher quantities of output are produced with the use of lower inputs.

# FDH Assumption

FDH model can aims to minimize inputs while satisfying at least the given output levels. This is called the input-oriented model.

The other one is called the output-oriented model that attempts to maximize outputs without requiring more inputs.

Under input-oriented condition, the efficiency scores of FDH input-oriented model are always greater than the input-oriented variable returns to scale(VRS) model. Also, the efficiency scores of input-oriented VRS model are always greater than those of input-oriented CRS model.

In other words, the production possibility set of FDH model is subset of VRS model as well as CRS model.

# Run FDH Script

The function for run FDH analysis is called **FDH_Analysis**. The results of the function are both input and output oriented scores, stored into 2 columns:

- The first column is the input-oriented analysis;
- The second column is equal to the output-oriented analysis

---

The arguments for the function are:

X : Matrix of input(s) (n x p) (Loaded from the File)

Y : Matrix of output(s) (n x k) (Loaded from the File)

Example: [Fdh_,Fdh_O]=FDH_ANALYSIS(X,Y);

*References FDH Analysis script in MATLAB: L. SIMAR (april, 2002)*

# DEA Vs FDH Theory Graphs

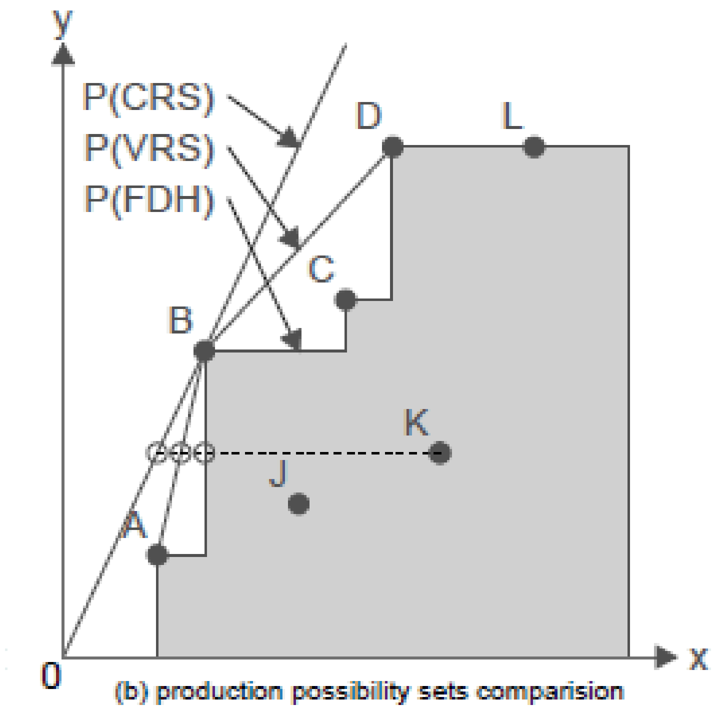



Figure 1: production possibility sets in CRS, VRS, and FDH

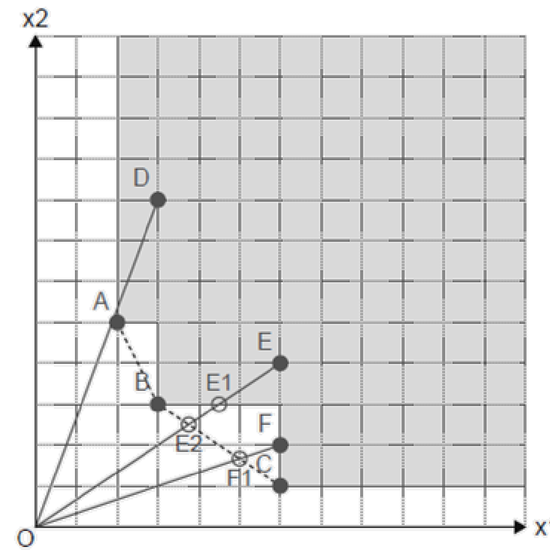
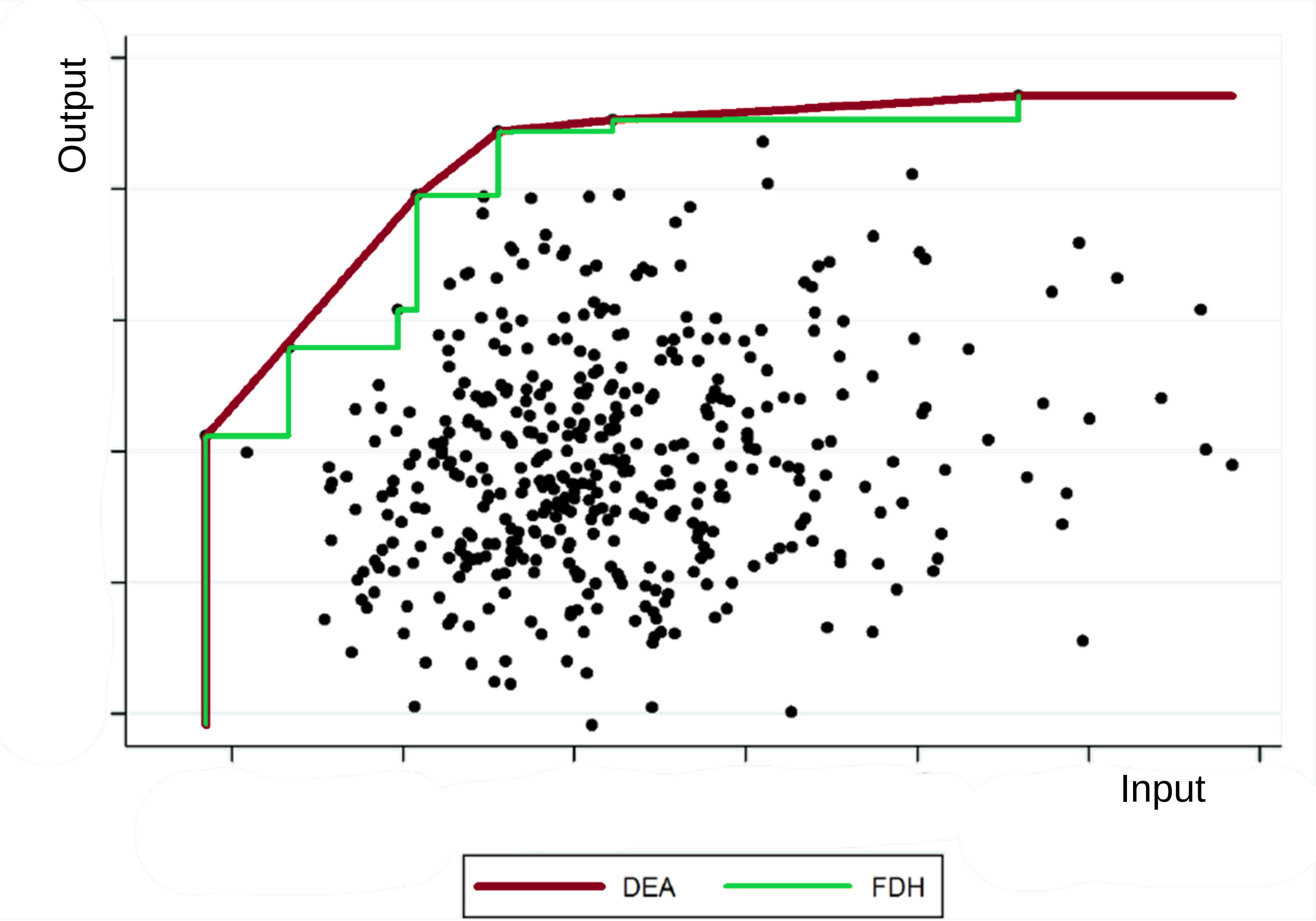



Figure 2: Input Oriented FDH Efficiency Measures

$$\theta_{FDH,input} \geq \theta_{VRS,input} \geq \theta_{CRS,input}$$

*References: Lim, B., Lee, K., & Lee, C. (2016). Free Disposal Hull (FDH) analysis for efficiency measurement: An update to DEA. The Stata Journal, 10(2), 1-8.*

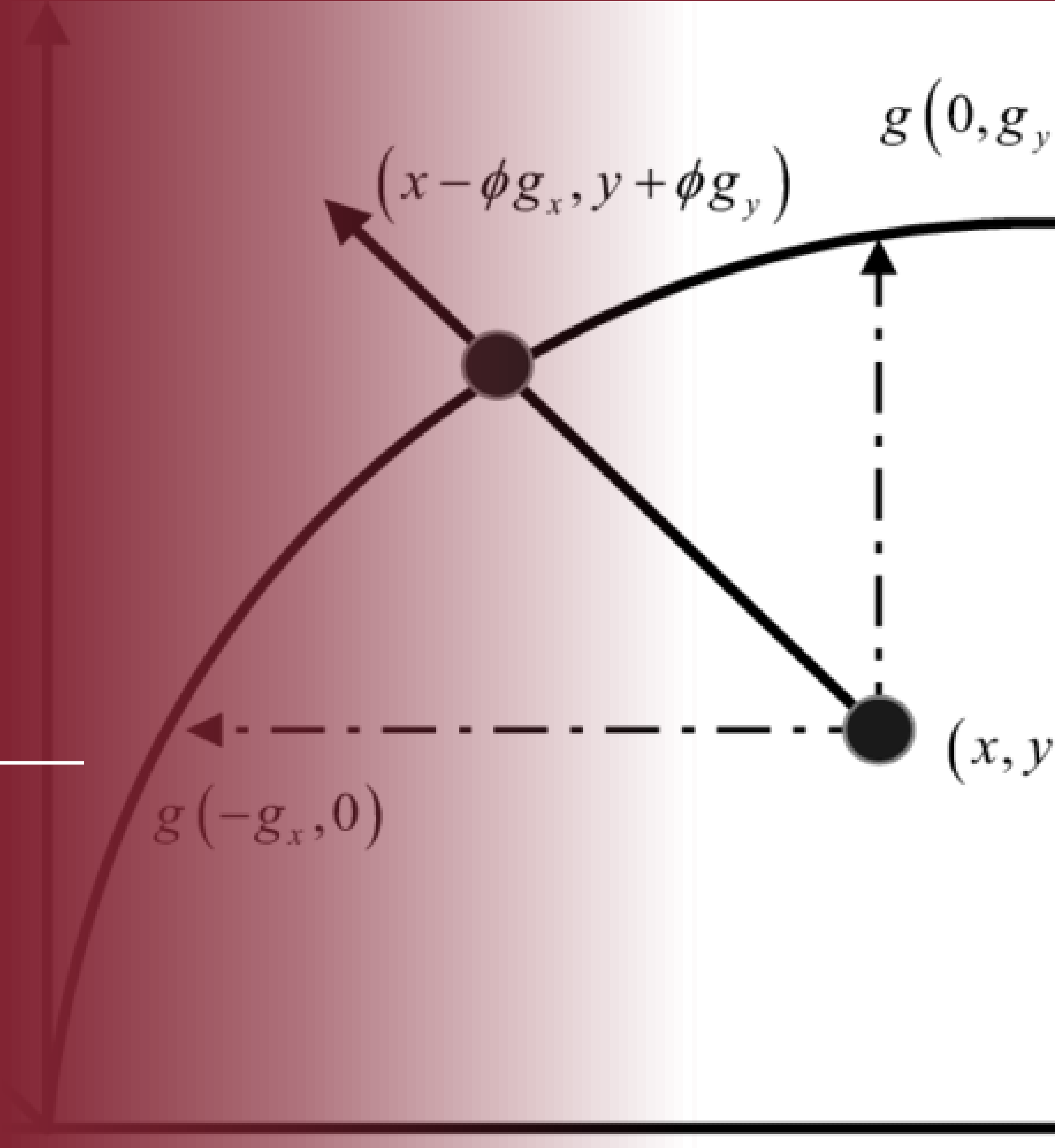
# DEA Vs FDH Example



**Thank you for your attention!**

# Additional material

# Directional Efficiency



# Directional Distances

Directional distances provide useful, flexible measures of technical efficiency of production units relative to the efficient frontier of the attainable set in input-output space.

In addition, the additive nature of directional distances permits negative input or outputs quantities.

The choice of the direction allows analysis of different strategies for the units attempting to reach the efficient frontier.

DDF accounts for both the reduction in the inputs or the expansion in the output simultaneously in the direction $g(x_i, y_r)$

# Directional Distances

Directional distance is the projection of (x, y) onto the technology frontier in a direction $d = \left(-d_x, d_y\right)$.

$$\delta\left( x, y \mid d_x, d_y, \Psi \right) = \sup\left\{ \delta \mid \left(x - \delta d_x, y + \delta d_y\right) \in \Psi \right\}$$

The distance function translates the vector (x,y) into a direction $g_{(x_i, y_r)}$ to the boundary of the production frontier.

Instead of a proportional contraction/expansion of the input/output vector, the direction distances functional approach allows for other options, such as non-proportional reductions

# Direction Choice

The directional model can be computed specifying the Gx and Gy parameters as a matrix or as a scalar (usually, 0 or 1)

Other directions are possible, including elaborated transformations driven by the data as proposed by Daraio and Simar (2016)

Other option is the projecting observations to economic optima introduced by Zofío, Pastor,and Aparicio (2013)


*Daraio C, Simar L (2016). "Efficiency and Benchmarking with Directional Distances: A Data-Driven Approach." Journal of the Operational Research Society, **67**(7), 928–944. doi:10.1057/jors.2015.111.*

*Zofío JL, Pastor JT, Aparicio J (2013). "The Directional Profit Efficiency Measure: On Why Profit Pnefficiency Is Either Technical or Allocative." Journal of Productivity Analysis, **40**(3), 257–266. doi:10.1007/s11123-012-0292-0.*

# DEA Toolbox Directional Distance

The DEA toolbox function for directional DEA analysis is the same as previously used, but with changes in the function arguments.

Example = dea(X, Y, 'ddf', 'Gx', X, 'Gy', Y);

Properties:

'orient': orientation. Input oriented 'io', output oriented 'oo', **directional distance function 'ddf'**

- 'rts': returns to scale. Constant returns to scale 'crs', variable, returns to scale 'vrs'.
- **'Gx': input directions for 'ddf' orientation. Default is Xeval.**
- **'Gy': output directions for 'ddf' orientation. Default is Yeval.**
- 'names': DMU names.
- 'secondstep': 1 to compute input and output slacks. Default is 1.

Advanced parameters:

- 'Xeval: inputs to evaluate if different from X.
- 'Yeval': outputs to evaluate if different from Y.

# Full Frontier Directional Distance MATLAB Script

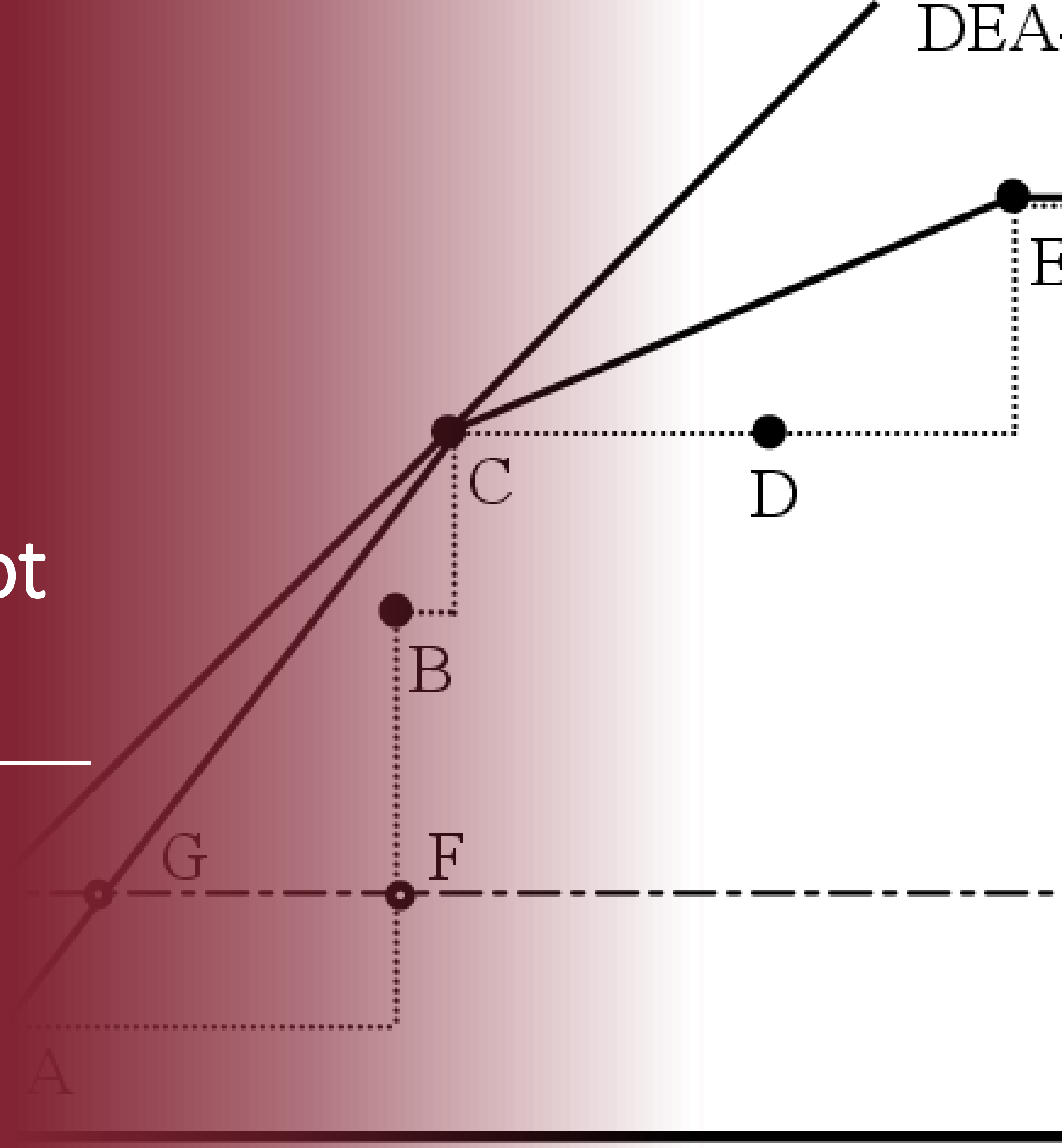
The functions FDH_Dir_Dist, has as arguments (X, Y) which are the matrices of inputs and outputs, with dimensions $(n \times p)$ and $(n \times q)$ respectively and the vectors (dx,dy) which are the direction vectors.

Some elements of the direction vectors gx and gy may be zeros, but at least one element of the full vector (gx, gy) of dimension $((p + q) \times 1)$ must be strictly positive.

Example : FDH_DD=FDH_Dir_Dist(dx,dy,X,Y);

For more detail see Daraio, C., Simar, L., & Wilson, P. W. (2020). Fast and efficient computation of directional distance estimators. *Annals of Operations Research*, *288*(2), 805-835.

# Production Frontier Plot in MATLAB



# Multiplot Script

MultiPlot is a function for represent graphically on a Cartesian plane the production set and the various efficiency Scores Frontier. This is a work in progress script, so any feedback is welcome

The concept behind this script is simple :

- Initially creates a scatter plot with the various DMUs.
- Then it takes the results of the efficiency analysis and creates lines connecting the points of the DMUs with Efficiency =1.

This script will work **ONLY** with 1 input and 1 output models (and for now only with the DEA crs, vrs and FDH model, not Directional distance Function)

# Multiplot Script

MULTIPLOT ARGUMENTS:

- X : Matrix of input(s) (n x p) (Loaded from the File)
- Y : Matrix of output(s) (n x k) (Loaded from the File)
- DEAM, vector with all the efficiency score used for plot, dimension of the vector n*number_analysis
- Rtsm, Vector with all the rts used,dimension n*number_analysis
- Max_Scale, indicates the max x used for represent the frontier scalar

---

RTS AVAILABLE:

- 'CRS'=COSTANT RETURN TO SCALE
- 'VRS'=VARIABLE RETURN TO SCALE
- 'NDRS'=NON DECREASING RTS (NOT USED IN THE SEMINAR)
- 'NIRS'=NON INCREASING RTS (NOT USED IN THE SEMINAR)
- 'FDH'=FREE DISPOSAL HULL OPTION

# References

MATLAB SCRIPTS:

Dea Toolbox:

*Álvarez, I., Barbero, J., & Zofío, J. L. (2016). A data envelopment analysis toolbox for MATLAB*

Fdh Analysis Script:

*FDH Analysis script in MATLAB: L. SIMAR (april, 2002)*

---

**OTHER MATERIALS REFERENCES:**

*Daraio, C., Simar, L., & Wilson, P. W. (2020). Fast and efficient computation of directional distance estimators*

*Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach*

*Bădin, L., Daraio, C., & Simar, L. (2019). A bootstrap approach for bandwidth selection in estimating conditional efficiency measures*