# ADVANCED BENCHMARKING MODELS AND TECHNIQUES

# R TUTORIAL

Thyago Nepomuceno
thyago.nepomuceno@ufpe.br

RISIS
RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

SAPIENZA
UNIVERSITÀ DI ROMA

UNIVERSIDADE FEDERAL DE PERNAMBUCO

# Outline

- INTRODUCTION TO R
- EFFICIENCY ANALYSIS IN R: SURVEY OF OPTIONS
- READING DATA
- DESCRIPTIVE STATISTICS & VISUALIZATIONS
- SEMINAL MODELS (CCR & BCC)
- TESTING RETURNS TO SCALE,CONVEXITY AND SEPARABILITY
- BOOTSTRAPPING
- DIRECTIONAL EFFICIENCY

# INTRODUCTION TO R

# Introduction

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS.

- Initially written & released as an open source software during 90s (the “S” language for data analysis)

- Always at the edge of scientific knowlegde providing over 16 thousand packages with the most advanced computations for the development of mathematical and statistical models of any type.

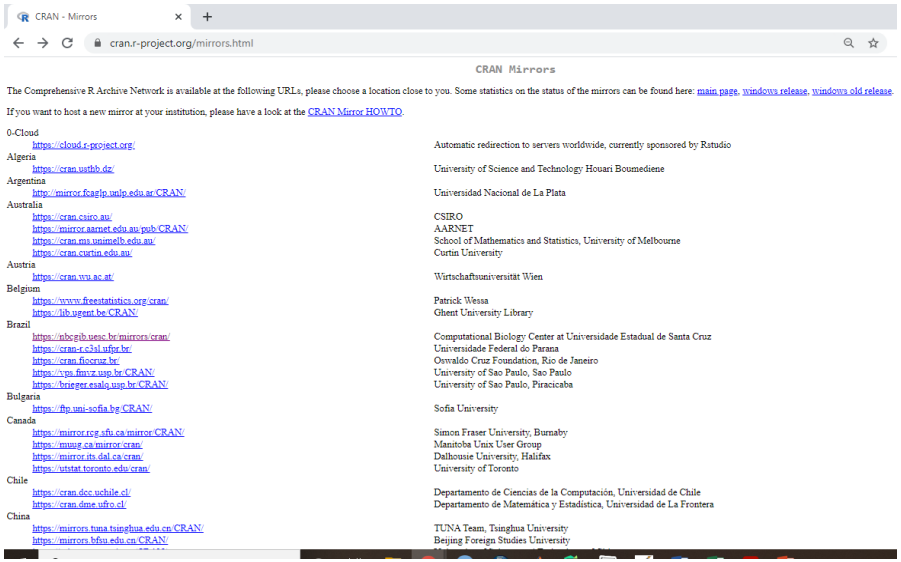
# Introduction

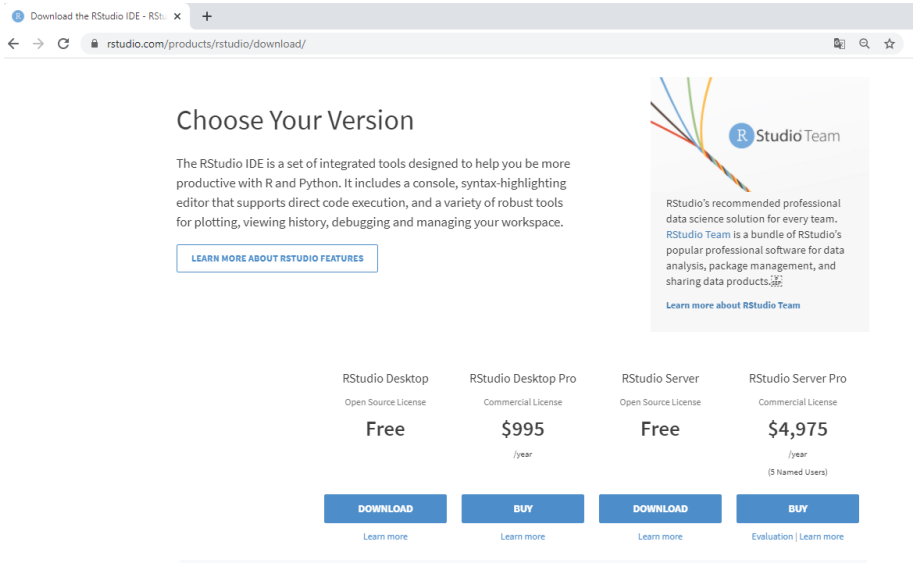
- R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible.

- Current version: R 4.0.3 (released on 10.10. 20) - https://www.r-project.org/

- **RStudio** is an open source integrated development environment (IDE) for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

# Introduction



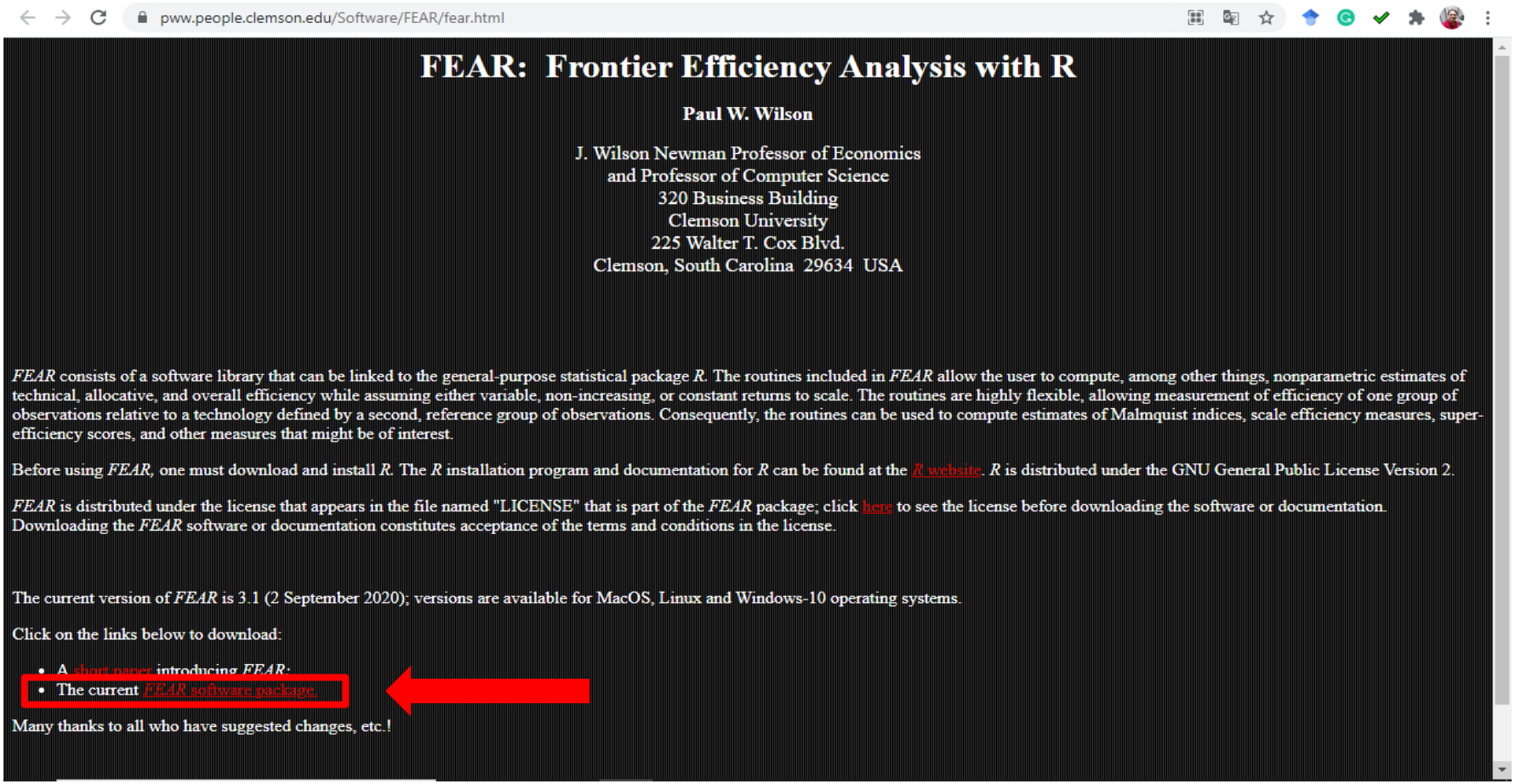




https://cran.r-project.org/mirrors.html

https://rstudio.com/products/rstudio/download/

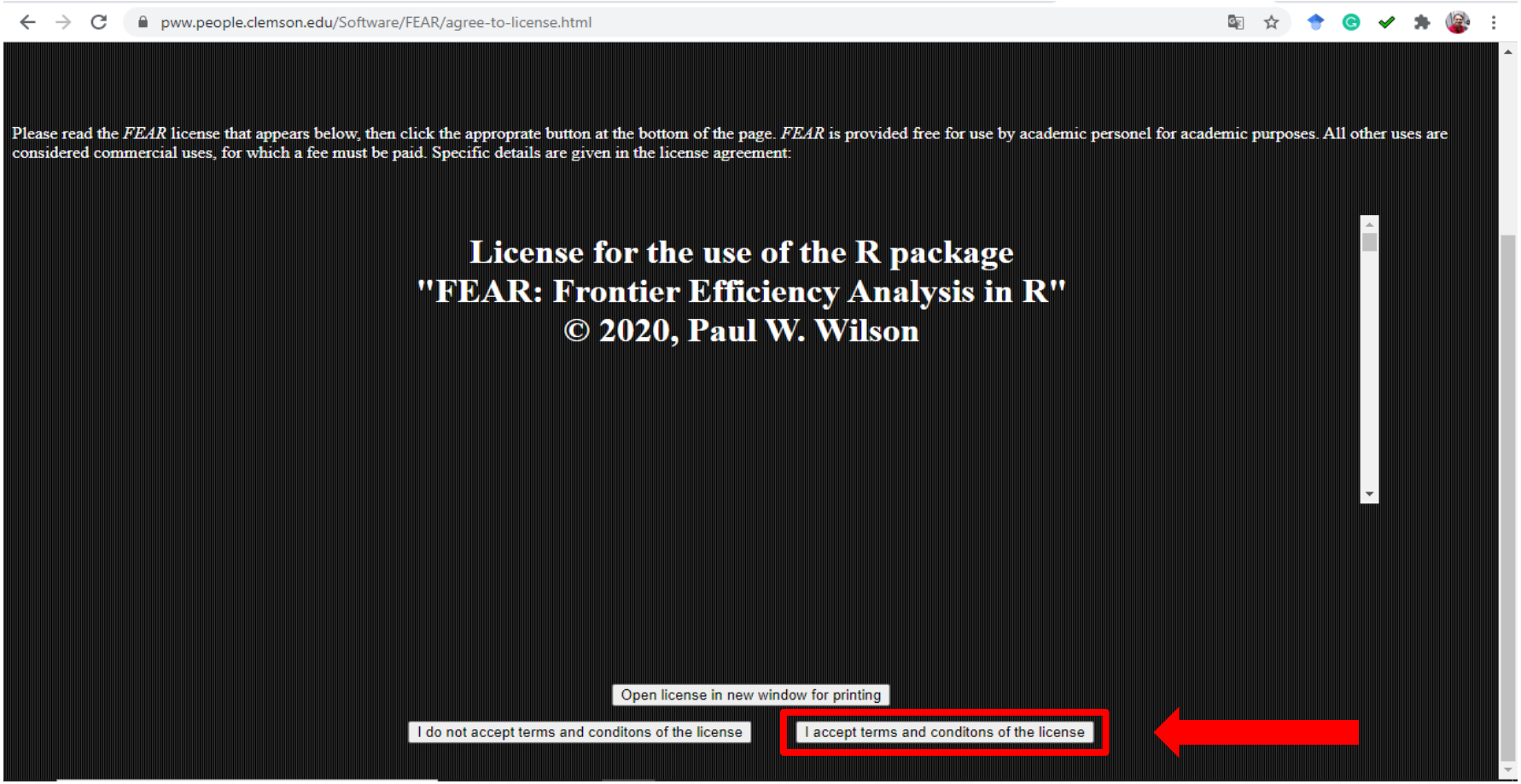
(Download First)

# Introduction



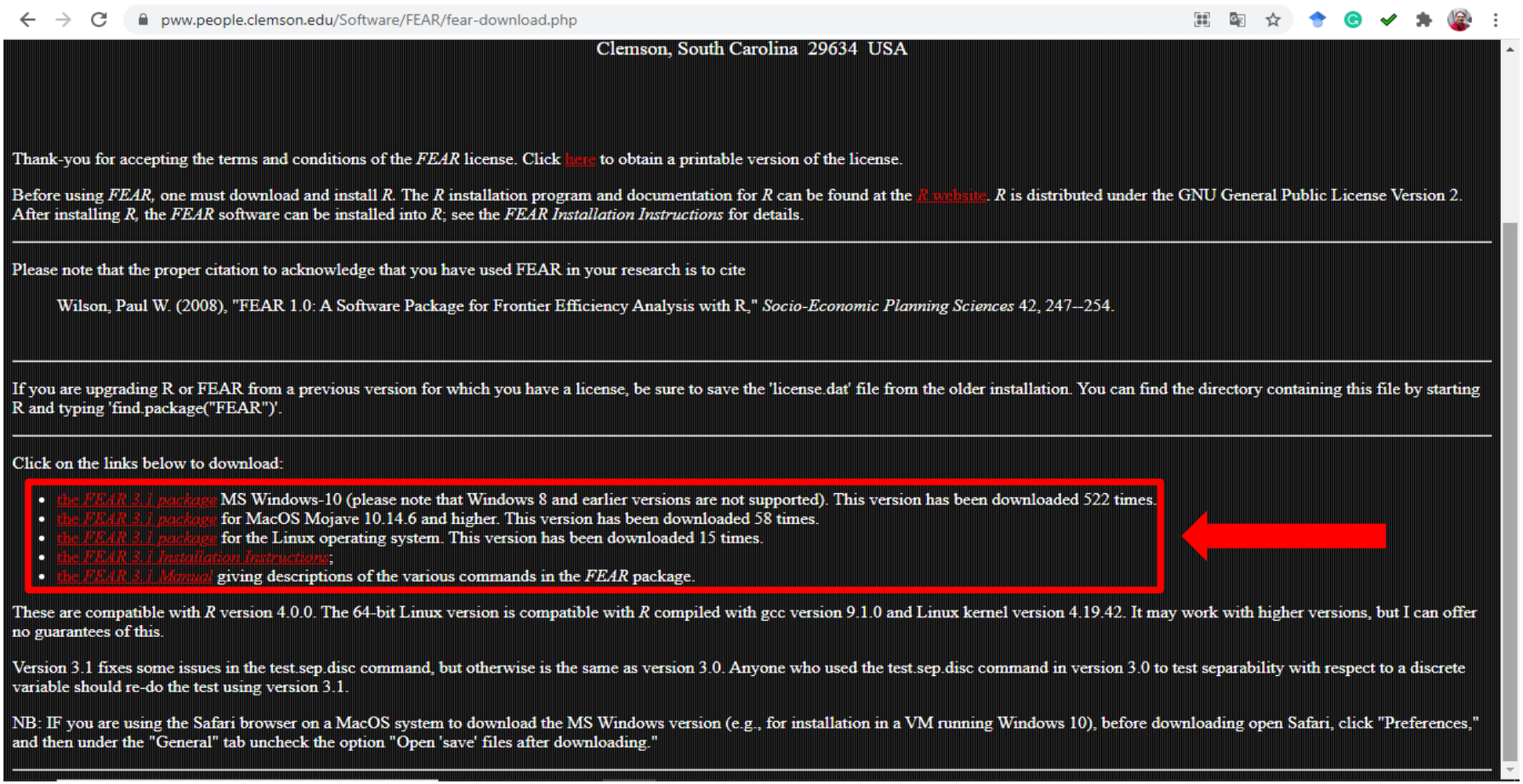

- Go to https://pww.people.clemson.edu/Software/FEAR/fear.html and select the current FEAR package for download

# Introduction



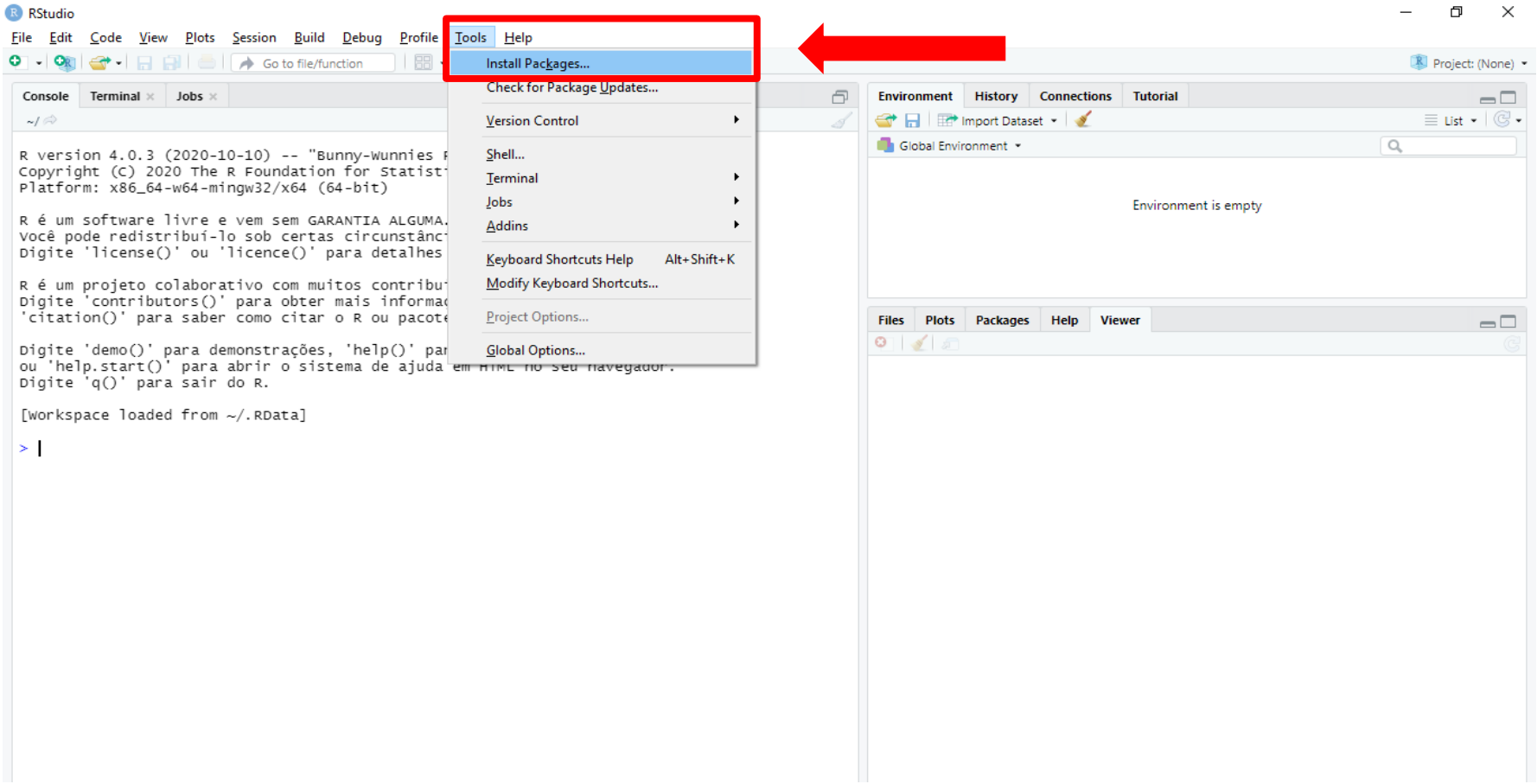

- Accept the terms and conditions of the license

# Introduction



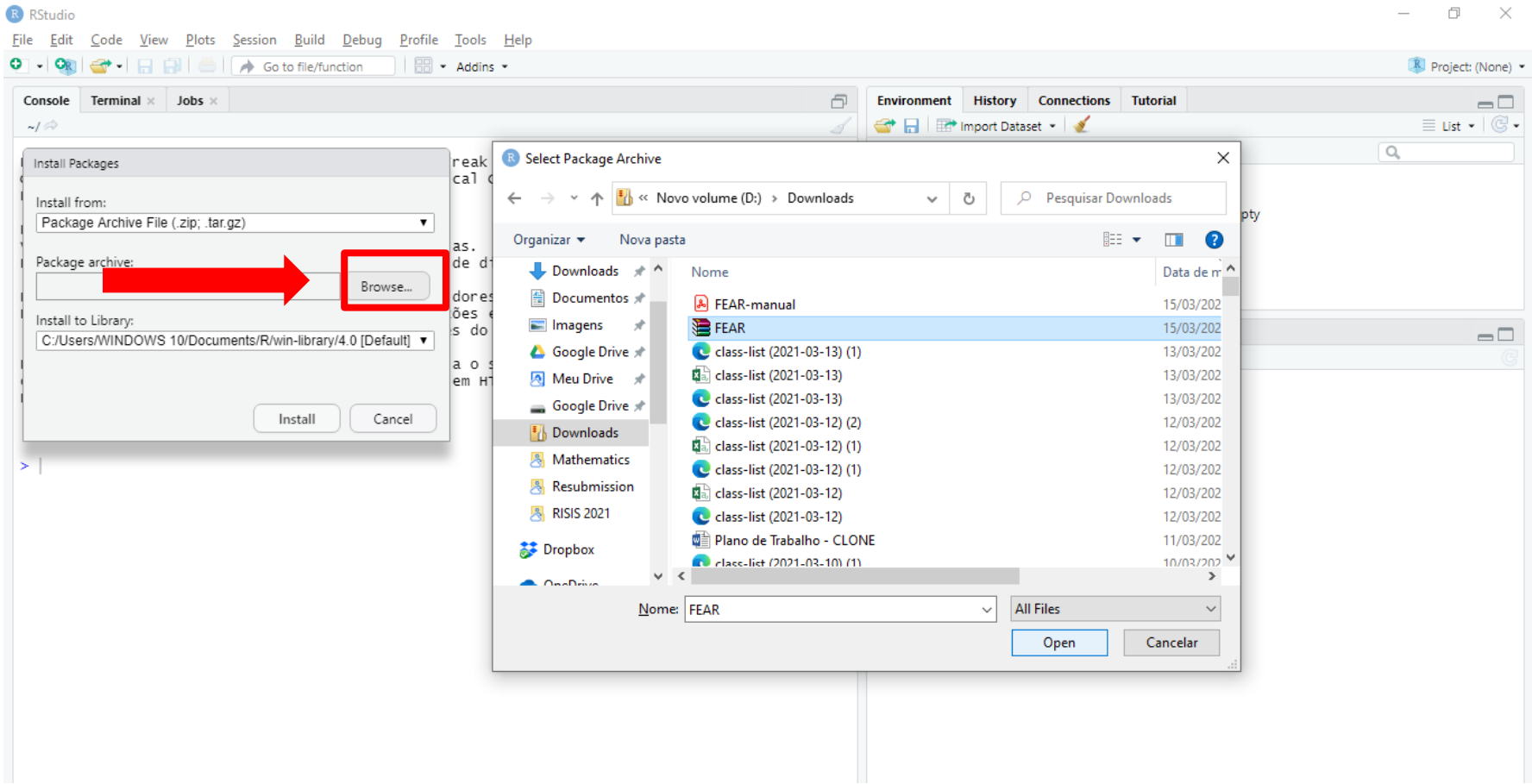

- Download the most appropriate version for your OS. Download the manual containing the commands to perform efficiency analysis.

# Introduction



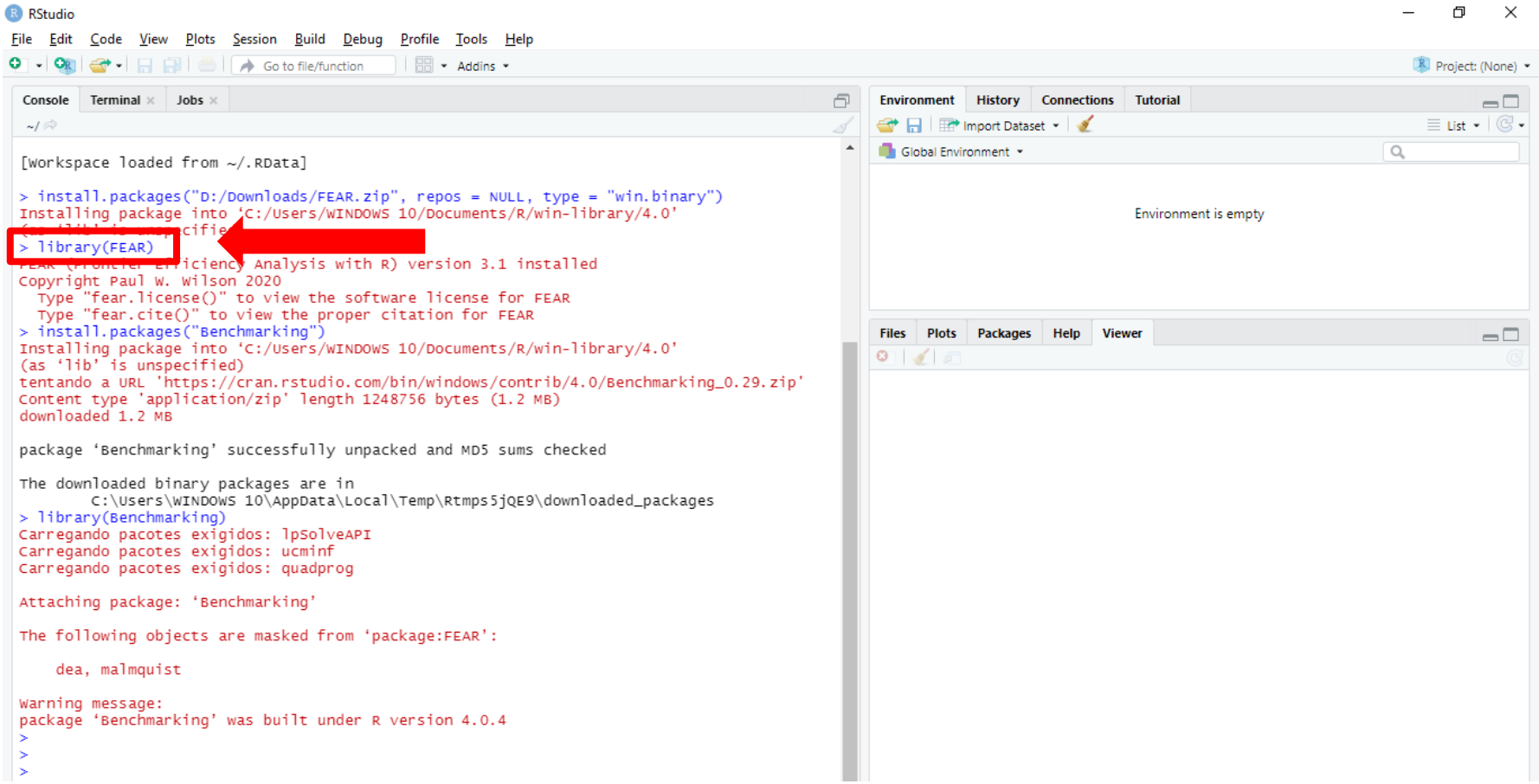

- Open your R Studio version and proceed to the tab “Tools” and select “Install Packages…”

# Introduction



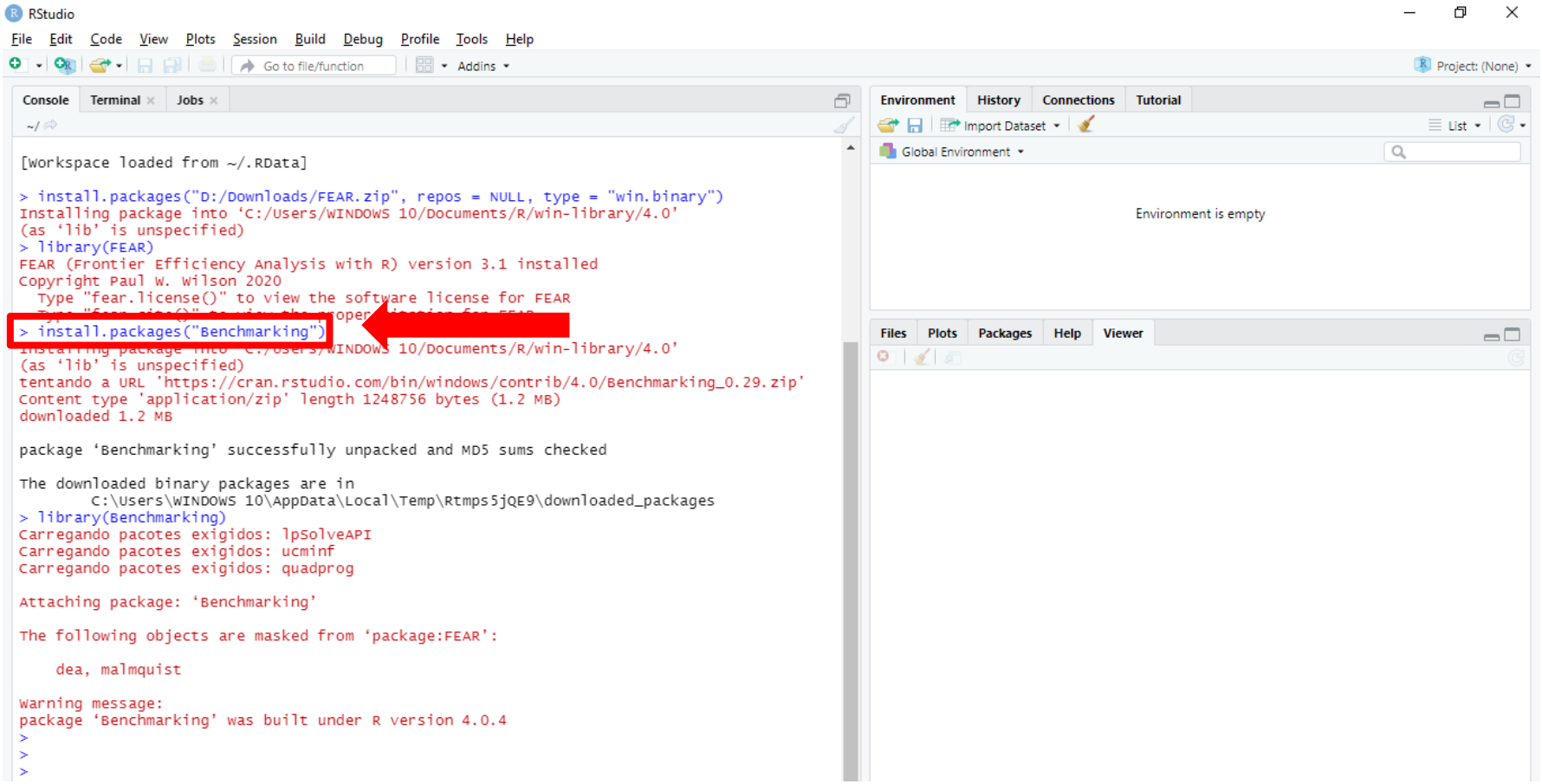

- Select the downloaded package and click “Open”, then “Install”.

# Introduction



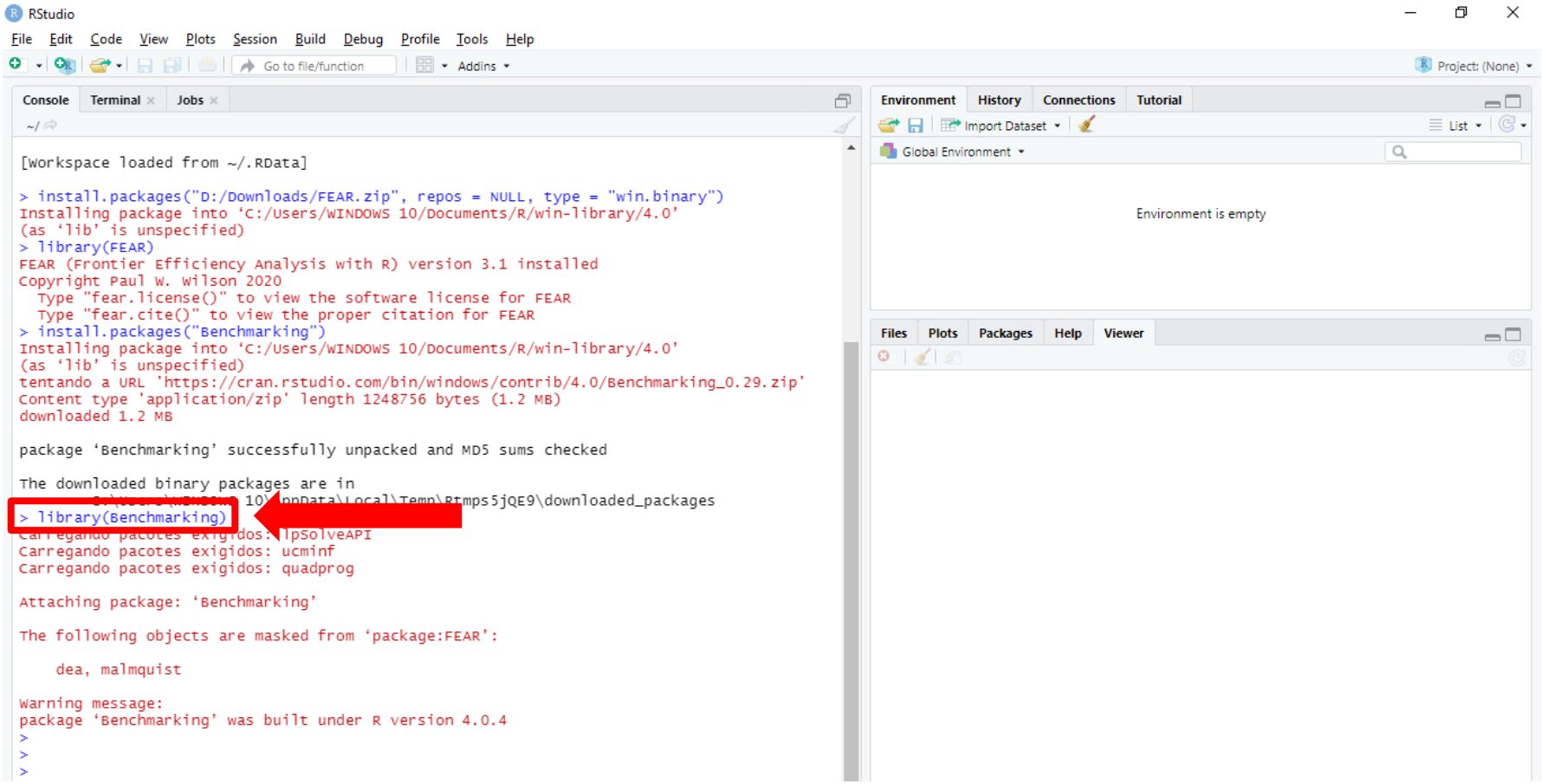

- To load the package type the command “library(FEAR)”

# Introduction




- To install the Benchmaking package you need to type the command “install.packages(“Benchmarking”)

# Introduction

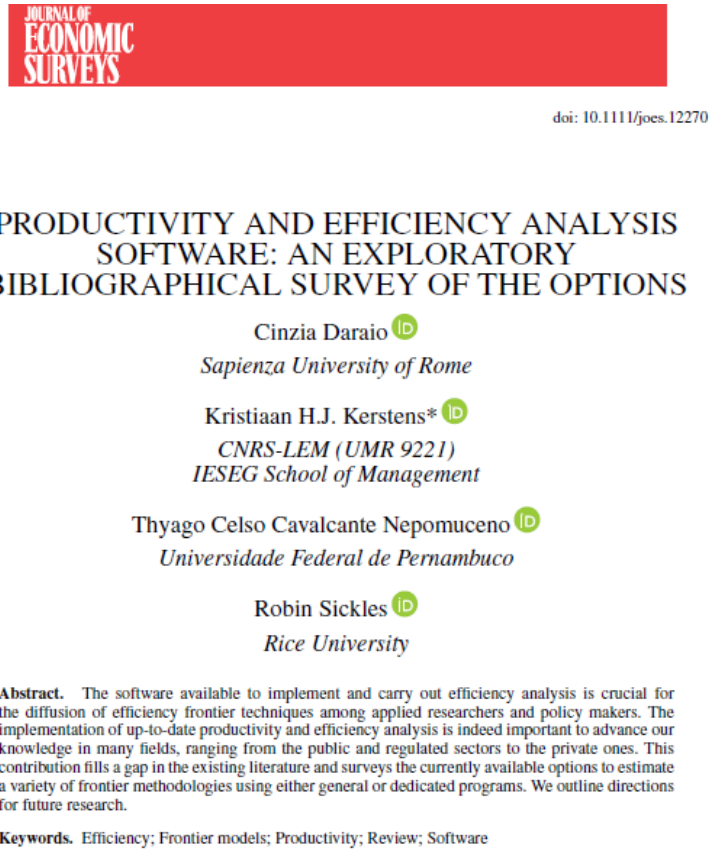



- To load the package type the command "library(Benchmarking)"

# EFFICIENCY ANALYSIS IN R

## A SURVEY OF OPTIONS

# PEA Software Options Survey



JOURNAL OF ECONOMIC SURVEYS

doi: 10.1111/joes.12270

PRODUCTIVITY AND EFFICIENCY ANALYSIS SOFTWARE: AN EXPLORATORY BIBLIOGRAPHICAL SURVEY OF THE OPTIONS

Cinzia Daraio
*Sapienza University of Rome*

Kristiaan H.J. Kerstens*
*CNRS-LEM (UMR 9221)*
*IESEG School of Management*

Thyago Celso Cavalcante Nepomuceno
*Universidade Federal de Pernambuco*

Robin Sickles
*Rice University*

**Abstract.** The software available to implement and carry out efficiency analysis is crucial for the diffusion of efficiency frontier techniques among applied researchers and policy makers. The implementation of up-to-date productivity and efficiency analysis is indeed important to advance our knowledge in many fields, ranging from the public and regulated sectors to the private ones. This contribution fills a gap in the existing literature and surveys the currently available options to estimate a variety of frontier methodologies using either general or dedicated programs. We outline directions for future research.

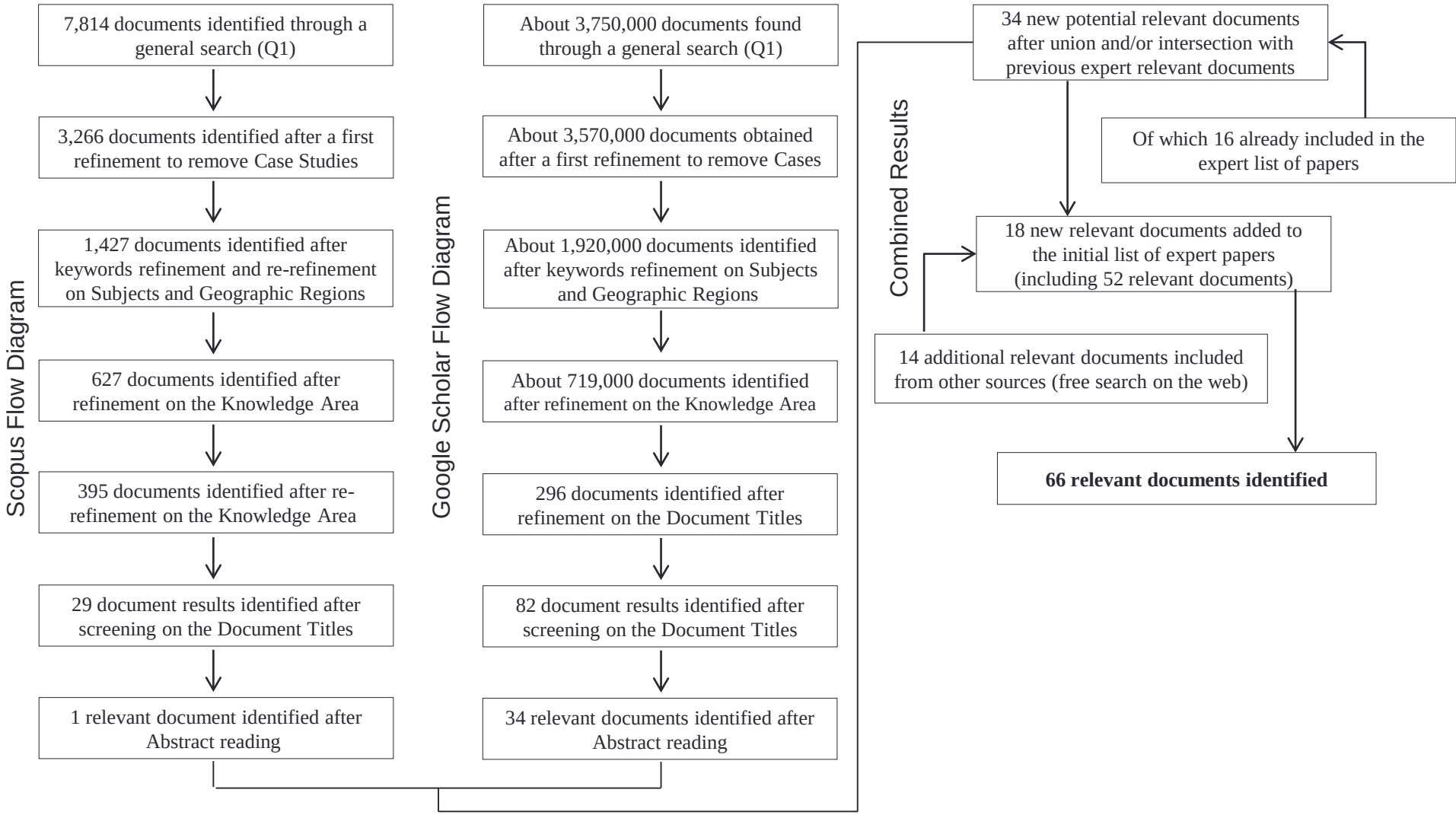
**Keywords.** Efficiency; Frontier models; Productivity; Review; Software

The survey investigates the current available options to estimate a variety of frontier methodologies using either general or dedicated programs, outlining directions for future research.

https://www.researchgate.net/publication/325660186_PRODUCTIVITY_AND_EFFICIENCY_ANALYSIS_SOFTWARE_AN_EXPLORATORY_BIBLIOGRAPHICAL_SURVEY_OF_THE_OPTIONS

# Methodology

## Document Results



# Methodology

## Document Results

Table 2: 18 Relevant Documents added after the systematic search on Scopus and Google Scholar.

| Order | Reference | Document Type | Versions |
|---|---|---|---|
| [35] | Argyrioy and Sifaleras (2013) | Proceedings | 6 |
| [36] | Barr and Durchholz (1992) | Proceedings | 1 |
| [37] | Ceyhan and Benneyan (2010) | Proceedings | 1 |
| [38] | Charnes et al. (1994) | Book | 1 |
| [39] | Chatzigeorgiou and Stiakakis (2011) | Article | 20 |
| [40] | Coelli et al. (2005) | Book | 10 |
| [41] | Coelli and Henningsen (2015) | Manual | 210 |
| [42] | Daouia and Laurent (2015) | Manual | 329 |
| [43] | Diaz-Martinez et al. (2008) | Manual | 3 |
| [44] | Emrouznejad and Thanassoulis (2009) | Proceedings | 2 |
| [45] | Iliyasu et al. (2015) | Article | 3 |
| [46] | Jablonsky (2014) | Article | 4 |
| [47] | Li et al. (2016) | Article | 3 |
| [48] | Meza et al. (2004) | Proceedings | 4 |
| [49] | Morgunov (2005) | Report | 3 |
| [50] | O'Donnell (2010) | Report | 5 |
| [51] | Straub (2015) | Manual | 117 |
| [52] | Surco and Wilhelm (2006) | Article | 8 |

Table 3: 12 Relevant Documents Added from Additional Sources

| Order | Reference | Document Type | Versions |
|---|---|---|---|
| [53] | Badunenko et al. (2017) | Manual | - |
| [54] | Belotti and Ilardi (2013) | Article | 10 |
| [55] | Bogetoft and Otto (2015)* | Manual | 197 |
| [56] | Dakpo et al. (2016) | Manual | - |
| [57] | Ferrara and Vidoli (2015) | Manual | 60 |
| [58] | Fusco and Vidoli (2015) | Manual | 44 |
| [59] | Kalvelagen (2002) | Report | 6 |
| [60] | Lim and Anderson (2012) | Proceedings | 4 |
| [61] | Oh and Suh (2013) | Manual | 210 |
| [62] | Pavlyuk (2016) | Manual | 100 |
| [63] | Ramanathan (2003) | Book | 2 |
| [64] | Shott and Lim (2015) | Manual | - |
| [65] | Sickles and Zelenyuk (2017) | Book | - |
| [66] | Wilson (2014)* | Manual | 2 |

# Comparative Analysis

| General purpose software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| Matlab (Sickles and Zelenyuk, 2017) | DEA/SFA models | Ibidem | Unlimited constraints and variables* | GUI | Text file; Scripts; Graphs; Projections; Tables; General Scores, Panel Data Estimators; Specific Results; Statistics, indexes | Free (for academic uses only) | Email and Phone technical support |
| R - Benchmarking - (Bogetoft and Otto, 2015) | FDH/DEA Models | Windows NT, Server, XP, Vista, 7, 8, 8.1 and 10; Mac OS X 10.6 or newer; Linux Ubuntu OpenSuse, Debian, Redhat and Ubuntu | Unlimited constraints and variables* | R Command-Line Prompt, RStudio, JGR, R Commander, RKWard, Deducer, Rattle, Red-R | Text file; Scripts; Graphs; Projections; Plots; Tables; Codes; General Scores, Specific Results; Partial Prices; Statistics, indexes; Weights, Lambdas; Peers and Slacks | Free (requires lpSolveAPI and ucminf packages) | User's Guide (59 pages); Reference Book; Email Support |
| R - FEAR - (Wilson, 2008) | FDH/DEA other robust Nonparametric Models | Ibidem | Ibidem | Ibidem | Ibidem | Free , but registration needed (requires KernSmooth package) | Introductory Whitepaper (14 pages); User's Guide (53 pages); Email Support |
| R - Frontier - (Coelli and Henningsen, 2015) | SFA Models | Ibidem | Ibidem | Ibidem | Ibidem | Free (requires micEcon and lmtest packages) | User's Guide (38 pages); Email Support |

# Comparative Analysis

| General purpose software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| R - Frontiles - (Daouia and Laurent, 2015) | Robust Nonparametric Eff. Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - Nonparaeff - (Oh and Suh, 2015) | FDH/DEA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - npsf - (Badunenko et al., 2017) | nonparametric and parametric efficiency Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - Productivity - (Dakpo et al., 2016) | DEA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |

# Comparative Analysis

| General purpose software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| R - semsfa - (Ferrara and Vidoli, 2015) | Semiparametric models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - SFA - (Straub, 2015) | SFA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - spfrontier - (Pavlyuk, 2016) | Spatial SFA | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - SSFA - (Fusco and Vidoli, 2015) | Spatial SFA | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |

# Comparative Analysis

| General purpose software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| R - TFDEA - (Shott and Lim, 2015) | DEA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - rDEA - (Simm and Besstremyannaya 2016)) | DEA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |
| R - DJL - (Lim 2016) | DEA Models | Ibidem | Ibidem | Ibidem | Ibidem | Ibidem | User's Guide; Email Support |

# Comparative Analysis

| General purpose software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| AMPL (Green, 1996) | DEA models | Windows (32 or 64 bit); Linux (32 or 64 bit); MacOS; NEOS Webserver | 300 DMUs, up to 90000 constraints, 300 input/output variables | Java; Matlab; C++; Visual Studio; AMPL Studio | Text file; Scripts; Graphs; Projections; Brief Summary of Detailed Results; Specific Results; Tables; Report on compilation errors | US$400 + US$300 (MINUS) for academics / US$4000 + US$9500 (CPLEX) + US$3000 (MINUS) for business | Free online Book and Reports |
| GAMS - dea.gms - (Kalvelagen, 2002) | DEA models | Windows Vista or newer (32 or 64 bit); Linux; MacOS X; Solaris (i86pc or SPARC 64 bit); IBM AIX | 300 DMUs and constraints, 300 input/output variables, 50 discrete variables, 2000 linear non-zeros, and 1000 non-linear non-zeros | GAMS IDE or external APIs (Excel; VBA; C; Visual Basic; Java; PHP, GIS; Matlab; Gnuplot; Web Server and others) | Text file; Scripts; Graphs; Projections; Maps; Summary Tables; Codes; Brief Detailed Results; Specific Results; Compilation Errors | US$640 for academics / US$3200 for business | Free online documentation; User's Guide (590 pages, US$13.22); Solver Manuals (508 pages, US$13.22) |
| Mathematica - DEA.m - Ley (1996) | DEA models | Windows 7, 8, 8.1 and 10 (32 or 64 bit); Mac OS X 10.9, 10.10, 10.11 and 10.12; Ubuntu 12.04–16.04, RHEL and CentOS 6–7, Debian 7–8, openSUSE 12.1–13.2, Leap 42.1 and Fedora 14–24; and Webserver (Browser or Mobile App) | Unlimited constraints and variables* | Graphical User Interface | Text file; Scripts; Graphs; Projections; Plots; Tables; Maps Sounds; Codes; General Scores, Specific Results; Charts; Statistics | US$1150 for academics / US$2360 for NPO (government) / US$2620 for business | Online Wolfram Language Documentation; Email and Phone technical support; Online FAQ; Video Tutorials and Training Classes; |
| Matlab - DEA Toobox – (Álvarez et al. 2016) | DEA models | Windows Server 2003, XP, Vista, 7, 8, 8.1 and 10; Mac OS X 10.9.5 and 10.10; Linux Ubuntu 16.04, SUSE 12, Red Hat 6 and 7 and Debian 7 and 8 | Unlimited constraints and variables* | GUI | Text file; Scripts; Graphs; Projections; Plots; Tables; Codes; General Scores, Specific Results; Statistics, indexes; Weights, Lambdas | Free (requires the Optimization Toolbox - €200 for academics and €1150 for business - and Statistics and Machine Learning Toolbox (optional for bootstrapping analysis) - €200 for academics and €1000 for business) | Documentation (White Paper - 40 pages); Online FAQ |

# Comparative Analysis

| Dedicated Software | Dimensions | | | | | | |
|---|---|---|---|---|---|---|---|
| | Frontier models | System requirement | Variable and Constraints Limitation | User interface | Reports' structure | Cost | User support |
| DEAFrontier (Zhu, 2002; 2014) | FDH/DEA Models | Windows XP, 7, 8, 8.1 and 10; MS Excel 1997 - 2003 and Excel 2007 - 2016 | 200 DMUs and constraints, 200 input/output variables | MS Excel spreadsheet | Text file; Scripts; Graphs; Projections; Tables; General Scores and Specific Results; Weights, Lambdas and Slacks | US$699 + US$1000 (Excel Solver) for academics / US$1499 + US$1000 (Excel Solver) for business | 1-month Free Technical Support; Free User's Guide and Online FAQ; Book of models and applications (optional) (414 pages, US$139.22) |
| DEA-Solver-Pro (Cooper et al., 2006) | FDH/DEA Models | Windows 9x/NT early versions, 2000, XP, 2003, Vista, 7, 8, 8.1 and 10Excel 97 - 2016 | Up to 60000 DMUs and constraints, unlimited number of input/output items* | MS Excel spreadsheet | Text; Graphs; Charts; Figures; Projections; Tables; General Scores, Summary Statistics, indexes; and Specific Results, Weights, Lambdas and Slacks | US$800 for academics / US$1,600 for business | Free User's Guide (134 English version pages and 148 Japanese version pages); Email technical support; Training classes (optional) (US$200 for up to 20 students) |
| EMS (Scheel, 2000) | FDH/DEA Models | Windows 9x/NT early versions, 2000, XP, Vista, 7, 8, 8.1 and 10 | 5000 DMUs and constraints, 40 input/output variables | EMS Graphical User Interface (Can resort to Text Editor and/or Excel spreadsheet) | Text and Tables; General Scores and Specific Results, weights, Lambdas, Benchmarking, costs and slacks | Free | Free User's Guide (12 pages); Email technical support |
| PIM-DEAsoft (Thanassoulis, 2001) | DEA Models | Windows 2000, XP, 2003, Vista, 7, 8, 8.1 and 10 | 50 DMUs and constraints, unlimited number of input/output items* | PIM-DEASoft Graphical User Interface | Text File; Tables; Graphs; Charts; Plots; Figures; Projections; General Scores, Summary Statistics, Indexes; and Specific Results | £100 for academics / £200 for business | Email technical support; Training classes (optional); 3 months free maintenance; |

# READING DATA

# Data Structure

- Supports virtually any type of data
- Numbers, characters, logicals (TRUE/ FALSE)
- Arrays of virtually unlimited sizes
- Simplest: Vectors and Matrices
- Lists: Can Contain mixed type variables
- Data Frame: Rectangular Data Set

# Reading Data

## Before Importing

- The first row of the spreadsheet is usually reserved for the header, while the first column is used to identify the sampling unit;
- Try to avoid using names that contain symbols such as ?, $, %, ^, (, ), -, #, , ‘,’ , <, >, /, \, |, [, ], {, and }
- Delete any comments that you have made in your Excel file to avoid extra columns or NA's to be added to your file; and
- Make sure that any missing values in your data set are indicated with **NA**

# Reading Data

- Directly using a vector e.g.: x <- c(1,2,3…)
- Spreadsheet like data:
  - Small/medium size: use read.table(), read.csv(), read.csv2(), read.xls(), read.xlsx(), read.txt()
  - Large data: use scan()
- Read from other systems:
  - Use the library “foreign”: library(foreign)
  - Can import from SAS, SPSS, Epi Info
  - Can export to STATA
- Using matrix function to read data matrices
- Using data.frame to read mixed data

# Accessing Variables

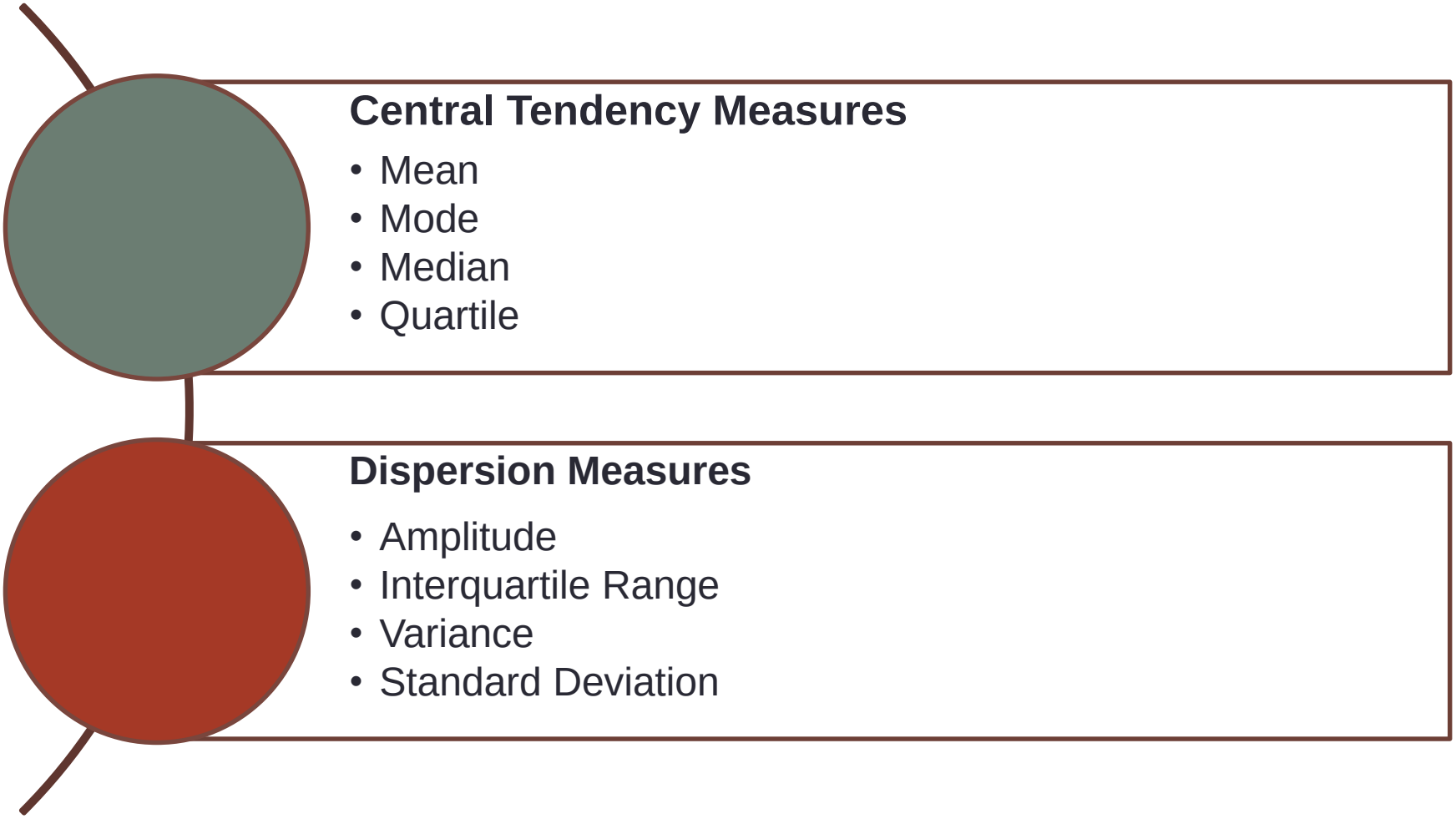
- edit(<mydataobject>)

- Subscripts essential tools
  - x[1] identifies first element in vector x
  - y[1,] identifies first row in matrix y
  - y[,1] identifies first column in matrix y

- $ sign for lists and data frames
  - mydata$age gets age variable of mydata
  - attach(mydata) -> extract by variable name

# DESCRIPTIVE STATISTICS & VISUALIZATIONS

# Descriptive Statistics



**Central Tendency Measures**

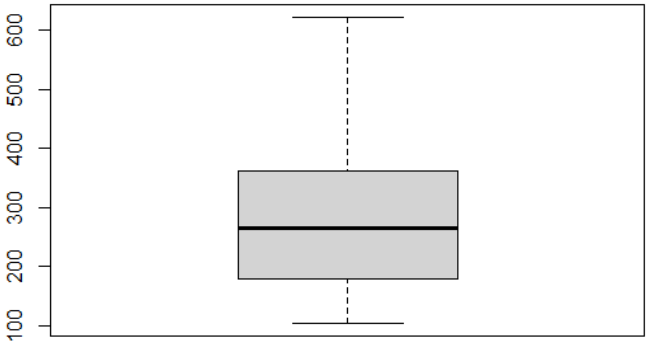
- Mean
- Mode
- Median
- Quartile

**Dispersion Measures**

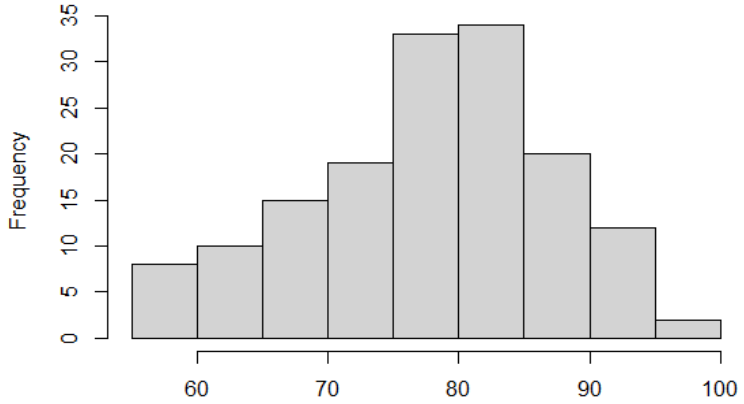
- Amplitude
- Interquartile Range
- Variance
- Standard Deviation

# Visualizations

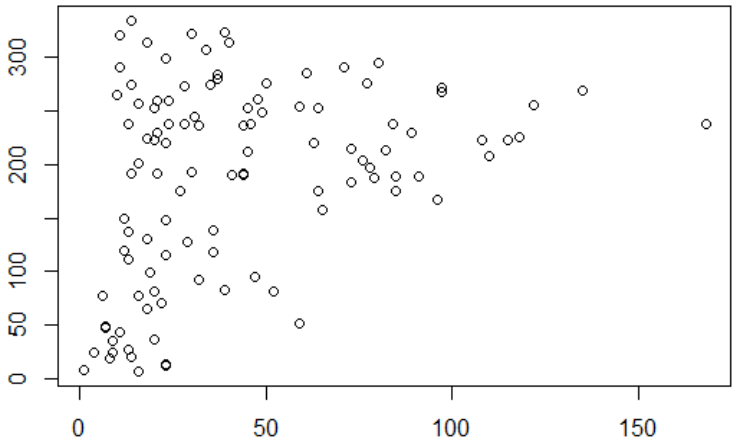
- Boxplots



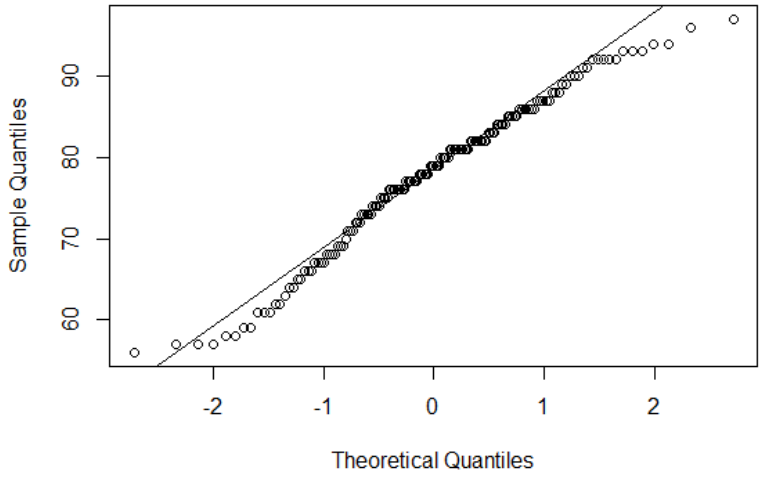

- Histograms




- Dispersion Graphs



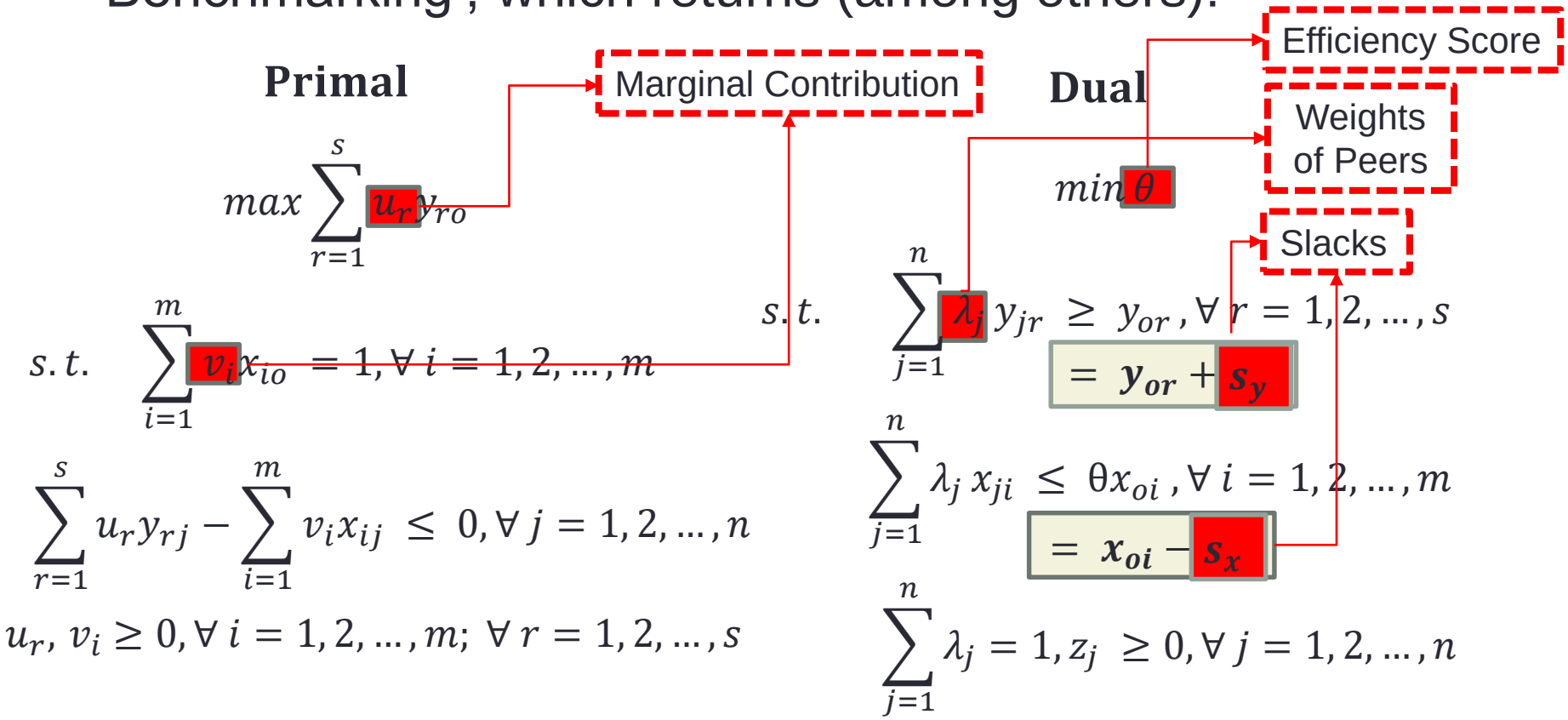

- QQ-Plots



# SEMINAL MODELS (CCR & BCC)

# Seminal Models

- Data Envelopment Analysis Models are implemented im many R package. The most common option is ‘Benchmarking’, which returns (among others):



# Seminal Models

## Implementation in R

- Usage
  - dea(X, Y, RTS="vrs", ORIENTATION="in", XREF=NULL, YREF=NULL, FRONT.IDX=NULL, SLACK=FALSE, DUAL=FALSE, DIRECT=NULL, param=NULL, TRANSPOSE=FALSE, FAST=FALSE, LP=FALSE, CONTROL=NULL, LPK=NULL)
- Arguments
  - **X**: Inputs of firms to be evaluated, a K x m matrix of observations of K firms with m inputs (firm x input)
  - **Y**: Outputs of firms to be evaluated, a K x n matrix of observations of K firms with n outputs (firm x input). DIRECT Directional efficiency
  - **DIRECT**: is either a scalar, an array, or a matrix with nonnegative elements. If the argument is a scalar, the direction is (1,1,...,1) times the scalar; the value of the efficiency depends on the scalar as well as on the unit of measurements. If the argument an array, this is used for the direction for every firm; the length of the array must correspond to the number of inputs and/or outputs depending on the ORIENTATION. If the argument is a matrix then different directions are used for each firm

# Seminal Models

## Implementation in R

- Arguments
  - **RTS:** Text string or a number defining the underlying DEA technology / returns to scale assumption
    - 0) **fdh**: Free disposability hull, no convexity assumption
    - 1) **vrs:** Variable returns to scale, convexity and free disposability
    - 2) **drs:** Decreasing returns to scale (down-scaling, but not up-scaling), convexity, and free disposability
    - 3) **crs:** Constant returns to scale, convexity and free disposability
    - 4) **irs:** Increasing returns to scale (up-scaling, but not down-scaling), convexity, and free disposability
    - 5) **irs2:** Increasing returns to scale (up-scaling, but not down-scaling), additivity, and free disposability
    - 6) **add:** Additivity (scaling up and down, but only with integers), and free disposability
    - 7) **fdh+:** A combination of free disposability and restricted or local constant return to scale
    - 10) **vrs+:** As vrs, but with restrictions on the individual lambdas via param

# Seminal Models

## Implementation in R

- Arguments
  - **ORIENTATION:** Input efficiency "in" (1), output efficiency "out" (2), and graph efficiency "graph" (3). For use with DIRECT, an additional option is "in-out" (0).
  - **XREF** Inputs of the firms determining the technology, defaults to X.
  - **YREF** Outputs of the firms determining the technology, defaults to Y.
  - **FRONT.IDX** Index for firms determining the technology
  - **SLACK** Calculate slack by an intern call of the function slack
  - **param** Possible parameters. At the moment only used for RTS="fdh+" to set low and high values for restrictions on lambda

# Seminal Models

## Implementation in R

- Arguments
  - **DUAL:** Calculate dual variables, i.e. shadow prices; not calculated for orientation graph as that is not an LP problem.
  - **FAST:** Only calculate efficiencies and just return them as a vector, i.e. no lambda or other output. The return when using FAST cannot be used as input for slack and peers.
  - **CONTROL:** Possible controls to lpSolveAPI
  - **LP:** Only for debugging. If LP=TRUE then input and output for the LP program are written to standard output for each unit.
  - **Object, x:**An object of class Farrell (returned by the function dea) – R code uses ‘object’ and ‘x’ alternating for generic methods.

# TESTING RETURNS TO SCALE AND CONVEXITY

# Testing Returns to Scale and Convexity

- Based on Kneip et al. (2016) tests of convexity and returns to scale, and Daraio et al. (2018) approach to randomly split the original sample into two independent subsamples to ensure independence between the two means under comparison.

- Each test involves comparing a sample mean of efficiency estimates that impose the conditions of the null hypothesis against a sample mean of efficiency estimates where the null is not imposed.

# Testing Returns to Scale and Convexity

- Hyphothesis Test (Returns to Scale):

$$H_0\colon \mu \underset{n^k}{\longrightarrow} \theta_{CRS} | k = 2/(p + q)$$
$$H_1\colon \mu \underset{n^k}{\longrightarrow} \theta_{VRS} | k = 2/(p + q)$$

- Hyphothesis Test (Convexity):

$$H_0\colon \mu \underset{n^k}{\longrightarrow} \theta_{FDH} | k = 1/(p + q+1)$$
$$H_1\colon \mu \underset{n^k}{\longrightarrow} \theta_{VRS} | \begin{cases} k = 2/(p + q+1)\ under\ \textit{strict convexity} \\ k = 2/(p + q)\ under\ \textit{weak convexity} \end{cases}$$

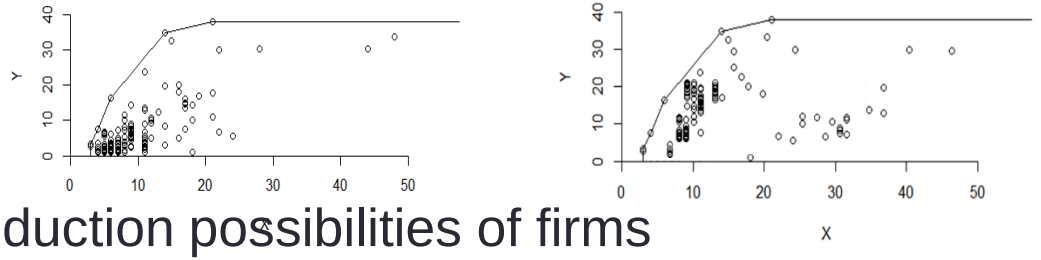
- Reject the Null Hypothesis if p-value is below the critical value (0.05)

# TESTING THE SEPARABILITY CONDITION

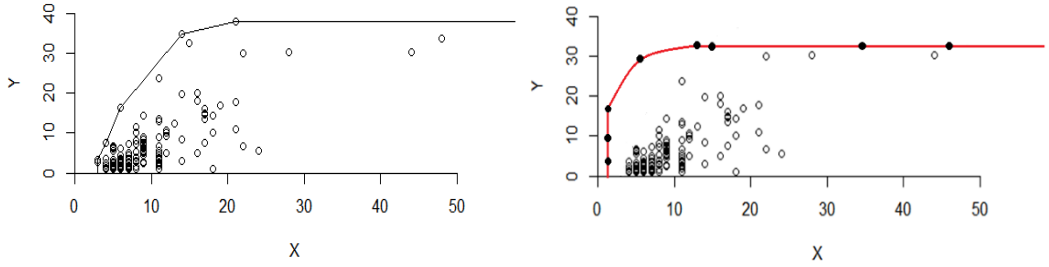
# Testing the Separability Condition

- The presence of environmental variables raises important questions for practitioners, such as the question of precisely how the environmental variables might affect the production process

- Environmental variables might affect
  - Only the distribution of efficiency among firms




  - The production possibilities of firms



# Testing the Separability Condition

- The separability condition assumes that environmental variables only affect the distribution of efficiency and do not affect production possibilities (Simar and Wilson, 2007).

- Separability Condition: $\Psi^z = \Psi$ for all $z \in Z$, where $\Psi = \{(X, Y) \mid X \text{ can produce } Y\}$.

- Non-Separability Condition: $\Psi^z \neq \Psi$ for some $z \in Z$.

# Testing the Separability Condition

- If the separability condition does not hold, unconditional DEA and FDH estimators have no useful interpretation: not only are second-stage regressions lack meaning when the separability condition is violated, but also the (unconditional) first-stage efficiency estimates is not meaningful (Dario et al. 2018).

- When it holds, it is meaningful to analyze the behavior of $\lambda(x, y)$ as a function of $Z$ by using an appropriate regression model (Simar and Wilson, 2007, 2011)

# BOOTSTRAP

# Bootstrap

- The bootstrap method described by Simar and Wilson (1998) is used to estimate confidence intervals for Shephard (1970) input or output distance functions.

- Efficiency estimates are computed using the routine dea

- Farrell output efficiency measure (which is the reciprocal of the Shephard output distance function) is recommended in many cases when estimated bias is larger than the distance function estimate (setting OUTPUT.FARRELL=TRUE)

# DIRECTIONAL EFFICIENCY

# Directional Distance Functions

- Representation of the production technology with multiple inputs and outputs.
- DDF accounts for both the reduction in the inputs or the expansion in the output simultaneously in the direction $g_{(x_i, y_r)}$

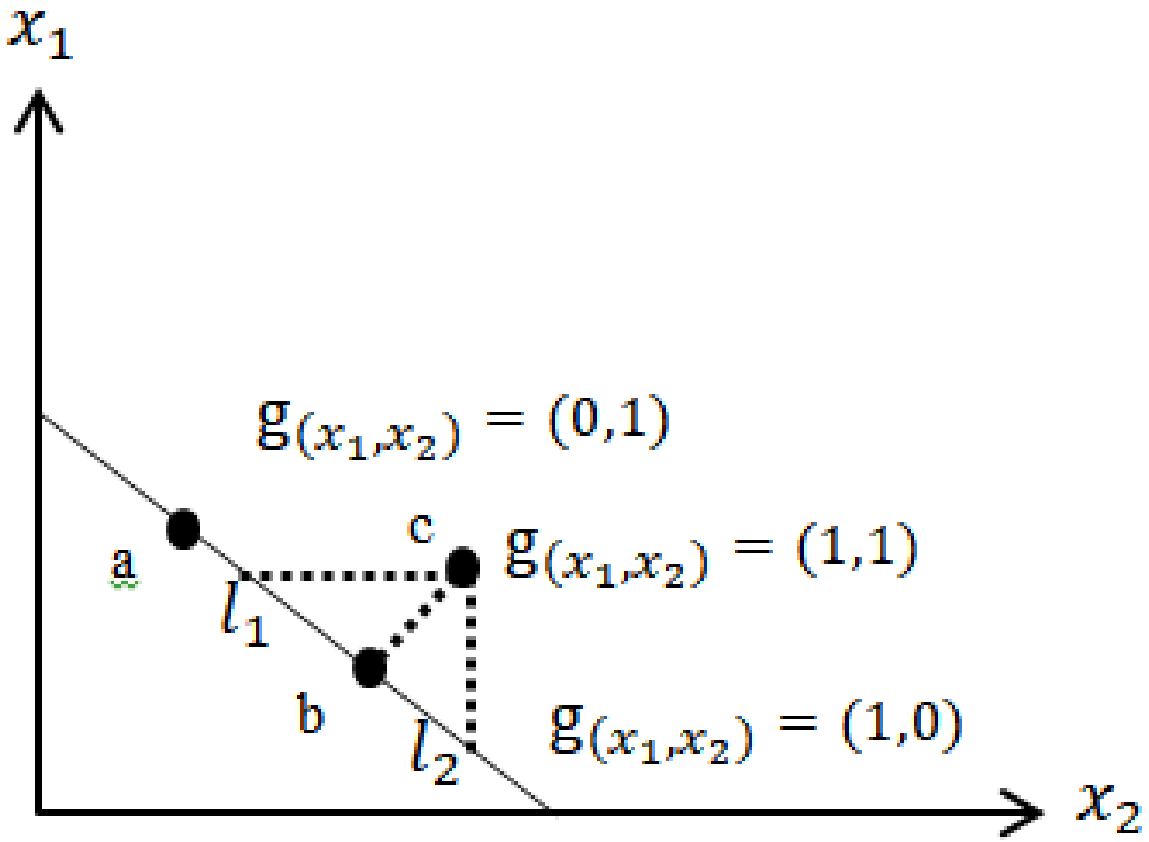
$$\mathrm{D}_t\left(\mathrm{x}, \mathrm{y}, g_x, g_y\right) = \sup\{\beta : (\mathrm{x} - \beta g_x, y + \beta g_y) \in \mathrm{T}\}.$$

$$Where\ \mathrm{T} = \{(\mathrm{x}, \mathrm{y}) : \mathrm{x}\ \text{can produce}\ \mathrm{y}\}$$

- Instead of a proportional contraction/expansion of the input/output vector, we allow non-proportional resource reductions or product increasing

# Directional Distance Functions

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



# Directional Distance Functions

## Implementation

- Implementing Directional Efficiency can be made by definition of polar coordinates, directional angles or targets (MEA), or by frontier scaling or slacks decompositions through Linear Programming:

$$\mathrm{D}_t\,(\mathrm{x},\mathrm{y},\mathrm{gx},\mathrm{gy}) = \; max\,\beta$$

$$s.t. \quad \sum_{j=1}^{k} z_j\, y_{jr} \;\geq\; y_{or} + \beta\, g_{y_r} \;\; r = 1, 2, \dots, s$$

$$\sum_{j=1}^{k} z_j\, x_{ji} \;\leq\; x_{oi} - \; \beta\, g_{x_i} \quad i = 1, 2, \dots, m$$

$$\sum_{j=1}^{k} z_j = 1, z_j \;\geq 0, \qquad j = 1, 2, \dots, k$$

# Directional Distance Functions

## Implementation in R

- Usage
  - dea.direct(X, Y, DIRECT, RTS = "vrs", ORIENTATION = "in", XREF = NULL, YREF = NULL, FRONT.IDX = NULL, SLACK = FALSE, param=NULL, TRANSPOSE = FALSE)
- Arguments
  - **X**: Inputs of firms to be evaluated, a K x m matrix of observations of K firms with m inputs (firm x input)
  - **Y**: Outputs of firms to be evaluated, a K x n matrix of observations of K firms with n outputs (firm x input). DIRECT Directional efficiency
  - **DIRECT**: is either a scalar, an array, or a matrix with nonnegative elements. If the argument is a scalar, the direction is (1,1,...,1) times the scalar; the value of the efficiency depends on the scalar as well as on the unit of measurements. If the argument an array, this is used for the direction for every firm; the length of the array must correspond to the number of inputs and/or outputs depending on the ORIENTATION. If the argument is a matrix then different directions are used for each firm

# Directional Distance Functions

## Implementation in R

- Arguments
  - **RTS:** Text string or a number defining the underlying DEA technology / returns to scale assumption
    - 0) **fdh**: Free disposability hull, no convexity assumption
    - 1) **vrs:** Variable returns to scale, convexity and free disposability
    - 2) **drs:** Decreasing returns to scale (down-scaling, but not up-scaling), convexity, and free disposability
    - 3) **crs:** Constant returns to scale, convexity and free disposability
    - 4) **irs:** Increasing returns to scale (up-scaling, but not down-scaling), convexity, and free disposability
    - 6) **add:** Additivity (scaling up and down, but only with integers), and free disposability
    - 7) **fdh+:** A combination of free disposability and restricted or local constant return to scale

# Directional Distance Functions

## Implementation in R

- Arguments
  - **ORIENTATION:** Input efficiency "in" (1), output efficiency "out" (2), and graph efficiency "graph" (3). For use with DIRECT, an additional option is "in-out" (0).
  - **XREF** Inputs of the firms determining the technology, defaults to X.
  - **YREF** Outputs of the firms determining the technology, defaults to Y.
  - **FRONT.IDX** Index for firms determining the technology.
  - **SLACK** Slacks calculated after taking the efficiency into consideration
  - **param** Possible parameters. At the moment only used for RTS="fdh+" to set low and high values for restrictions on lambda; see the section details and examples in dea for its use. Future versions might also use param for other purposes

# Directional Distance Functions

## Choosing the Direction

- The most common choice is for the unit vector $\boldsymbol{g}_{(\boldsymbol{x},\,\mathbf{y})} = (1,1)$ and data direction $\boldsymbol{g}_{(\boldsymbol{x},\,\mathbf{y})} = (x,y)$
- Other feasible choices are:
  - Define the direction for the expansion/contraction as proportions of the mean (Fukuyama and Weber 2017)
  - Data-driven direction (Daraio and Simar 2014, 2016)
  - $g_{(x,\,y)} = g_{(x_R\,,y_R)}$ (Briec 1997; Fukuyama and Weber 2009)
  - Orthogonal direction to the output set (Peyrache and Daraio 2012).
  - Determine the directional vector endogenously (Färe, Grosskopf and Whittaker 2013)
  - Subjective (Non-compensatory) Directions (Nepomuceno, Daraio & Costa, 2020)

# Directional Distance Functions

## Applications

Annals of Operations Research
https://doi.org/10.1007/s10479-020-03767-6

S.I. : CLAIO 2018

### Exogenous crimes and the assessment of public safety efficiency and effectiveness

Thyago Celso Cavalcante Nepomuceno[1,2] · Katarina Tatiana Marques Santiago[2] · Cinzia Daraio[1] · Ana Paula Cabral Seixas Costa[2]



**Abstract**
This work discusses the issue on how to include data about property and violent crimes in the production technology for the assessment of police technical efficiency. It applies recent advances in Directional Distances and Nonparametric Estimators. We claim that crime is an external variable not under the control of the decision units in view of the fact that it is exogenously determined. The results from the Conditional Directional Distance Analysis can be relevant to cities with high property misdemeanors and homicide rates. Our analysis may be helpful to obtain a more robust and fair classification of police and justice units under similar circumstances, determine the empirical effect of crime on police productivity, their optimal input–output relationship, explore potential associations and compensation effects, and rewarding efficient policy makers in the prevention of crime based on measures of police efficiency and effectiveness.

**Keywords** Data envelopment analysis (DEA) · Directional distance functions (DDF) · Free disposal hull (FDH) · Exogenous factors · Crime · Conditional frontier · Police efficiency · Policing · Pernambuco · Brazil

https://www.researchgate.net/publication/343840283_Exogenous_crimes_and_the_assessment_of_public_safety_efficiency_and_effectiveness

APPLIED ECONOMICS LETTERS
https://doi.org/10.1080/13504851.2019.1616051

ARTICLE

### Combining multi-criteria and directional distances to decompose non-compensatory measures of sustainable banking efficiency

Thyago Celso Cavalcante Nepomuceno[a,b], Cinzia Daraio[a] and Ana Paula Cabral Seixas Costa[b]

[a]Sapienza University of Rome, Roma, Italy; [b]Universidade Federal de Pernambuco, Pernambuco, Brazil

**ABSTRACT**
Non compensatory choices are widespread in the economics, strategic management and decision making. Nevertheless, many assessments of productivity still fail to consider non-compensatory preference structures in the measure of the technical inefficiency. This paper proposes a preference elicitation schema, typical of Multi-criteria decision analysis, for the selection of the directional vector in the assessment of a sustainable productivity. The direction choice is based on the weighted aggregation of concordance indexes for each decision criteria on each individual input, such that it represents an index of relative importance according to the decision maker's perspective. The methodology can be used to aid resource allocation and saving, identify benchmarks for efficient practices and more generally for planning environmental policies in many services and industrial organizations. We illustrate the method with an environmental efficiency evaluation of Brazilian Federal Saving Bank branches.

**KEYWORDS**
Directional distance functions; data envelopment analysis; multi-criteria decision analysis; banks; environmental efficiency

**JEL CLASSIFICATION**
C67; C44; D81; G21; Q29

#### I. Introduction

Since the introduction of the first linear formulations to measure the technical inefficiency and productivity of Decision Making Units (DMUs) from Charnes, Cooper, and Rhodes (1978), the so called Data Envelopment Analysis (DEA), countless models, applications and software tools have popularized the field (Daraio et al. 2019a, 2019b). The assessment of the efficiency although is subject to a kind of paradox. On the one hand, we have an objective efficiency measurement which is based on DEA, in which the analyst does not choose a direction along which to gauge the efficiency: the direction is imposed by the DEA linear programme according to which DMUs have to contract inputs (or expand outputs) in a radial proportionate way. On the other hand, we have a Directional Distance framework (Färe and Grosskopf 2004) in which the analyst is free to choose the preferred direction to expand the outputs or to reduce the inputs. The path towards with the outranking Multi-criteria (MCDA) elicitation methods to handle non-compensatory choices. The mathematical approach proposed here has the benefit to offer a detailed representation of the multi-attribute production possibilities to account for indifference, preference and veto thresholds, which may support policy makers to obtain insights on their own preferences and values. The application on the Brazilian Federal Saving Bank highlights the contribution of the proposed approach to classify sustainable units, identify sustainable practices, processes and determine the optimal input-output relationship.

#### II. Directional distance function

Directional Distances are a very flexible nonparametric way to gauge the performance of Decision Making Units. They allow measuring the technical efficiency of DMUs by the choice of a direction

https://www.researchgate.net/publication/333085363_Combining_multi-criteria_and_directional_distances_to_decompose_non-compensatory_measures_of_sustainable_banking_efficiency