Data Envelopment Analysis: From Foundations to Modern Advancements

Other titles in Foundations and Trends® in Econometrics

Non-Experimental Data, Hypothesis Testing, and the Likelihood Principle: A Social Science Perspective
Tom Engsted and Jesper W. Schneider
ISBN:978-1-63828-324-9

Factor Extraction in Dynamic Factor Models: Kalman Filter Versus Principal Components
Esther Ruiz and Pilar Poncela
ISBN:978-1-63828-096-5

Performance Analysis: Economic Foundations and Trends
Valentin Zelenyuk
ISBN: 978-1-68083-866-4

Experimetrics: A Survey
Peter G. Moffatt
ISBN: 978-1-68083-792-6

Climate Econometrics: An Overview
Jennifer L. Castle and David F. Hendry
ISBN: 978-1-68083-708-7

Data Envelopment Analysis: From Foundations to Modern Advancements

Zhichao Wang
University of Queensland
zhichao.wang@uq.edu.au

Valentin Zelenyuk
University of Queensland
v.zelenyuk@uq.edu.au

the essence of knowledge
Boston — Delft

Foundations and Trends® in Econometrics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

Z. Wang and V. Zelenyuk. *Data Envelopment Analysis: From Foundations to Modern Advancements*. Foundations and Trends® in Econometrics, vol. 13, no. 3, pp. 170–282, 2024.

ISBN: 978-1-63828-447-5

Foundations and Trends® in Econometrics
Volume 13, Issue 3, 2024

Editorial Scope

Foundations and Trends® in Econometrics publishes survey and tutorial articles in the following topics:

- Econometric Models
- Simultaneous Equation Models
- Estimation Frameworks
- Biased Estimation
- Computational Problems
- Microeconometrics
- Treatment Modeling
- Discrete Choice Modeling
- Models for Count Data
- Duration Models
- Limited Dependent Variables
- Panel Data
- Time Series Analysis
- Latent Variable Models
- Qualitative Response Models
- Hypothesis Testing
- Econometric Theory
- Financial Econometrics
- Measurement Error in Survey Data
- Productivity Measurement and Analysis
- Semiparametric and Nonparametric Estimation
- Bootstrap Methods
- Nonstationary Time Series
- Robust Estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2024, Volume 13, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Contents

Data Envelopment Analysis: From Foundations to Modern Advancements

Zhichao Wang[1] and Valentin Zelenyuk[2]

[1] *School of Economics, University of Queensland, Australia; zhichao.wang@uq.edu.au*

[2] *School of Economics and Centre for Efficiency and Productivity Analysis, University of Queensland, Australia; v.zelenyuk@uq.edu.au*

ABSTRACT

Data envelopment analysis (DEA) is a mainstream method for efficiency and productivity analysis, widely applied in numerous fields, including the healthcare sector, banking, energy generation and distribution, and cross-country economic growth analysis. In this monograph, we aim to provide a compendious overview of DEA. We start with the DEA estimators in various scenarios, such as for estimating technology, cost, revenue, profit functions and related efficiency measures, and its popular variants based on different assumptions about the shape of technology. The statistical properties and extensions on DEA, such as analysis on covariates of efficiency, are also discussed and the practical tips for computations are provided.

Keywords: Efficiency; data envelopment analysis; productivity index; statistical properties; covariates of efficiency.

JEL Codes: C14; C24; C43; C61; D24; I11; I18.

Zhichao Wang and Valentin Zelenyuk (2024), "Data Envelopment Analysis: From Foundations to Modern Advancements", Foundations and Trends® in Econometrics: Vol. 13, No. 3, pp 170–282. DOI: 10.1561/0800000040.

1

Introduction

Efficiency analysis methods are widely developed and applied in numerous fields such as healthcare, banking, energy, and agriculture, among others. The data envelopment analysis (DEA) (Charnes *et al.*, 1978) and stochastic frontier analysis (SFA) (Aigner *et al.*, 1977; Meeusen and Broeck, 1977) are the two mainstream approaches with many variants that can be at least traced back to the seminal works of Debreu (1951), Shephard (1953), and Farrell (1957). A variety of studies have been conducted with DEA and its variants in the past decades, such as in the systematic reviews by Emrouznejad *et al.* (2008) and Emrouznejad and Yang (2018), as well as the recent chapter by Ray (2020) and those regarding the DEA applications in banking (Miller, 2020), transportation (Wheat *et al.*, 2020), telecommunications (Bhattacharyya, 2020), etc. Here, we complement these studies in many respects, as explained below.

DEA has been found useful for analyzing many sectors of the economy in many countries. In the area of healthcare efficiency analysis, DEA was found as the most widely applied approach (e.g., in the review by Hollingsworth, 2008; O'Neill *et al.*, 2008; and more recently Kohl *et al.*, 2019; See *et al.*, 2024; and Wang *et al.*, 2024), which can be traced back to at least Nunamaker (1983), Banker *et al.* (1986), and Grosskopf and Valdmanis (1987). DEA is also widely used for analyzing efficiency and productivity of banking, e.g., see Sherman and Gold (1985), English *et al.* (1993), Seiford and Zhu (1999), Alam (2001), Sathye (2003), Casu *et al.* (2013), and Du *et al.* (2018), to mention a few.[1] DEA is also among the key analytical tools in energy and environmental studies, for example, Färe *et al.* (1985a), Goto and Tsutsui (1998), Athanassopoulos *et al.* (1999), Reinhard *et al.* (2000), Agrell and Bogetoft (2005), Sueyoshi and Goto (2013), and Wu *et al.* (2019).[2] DEA also received a wide use and appreciation in agricultural research, e.g., O'Donnell (2012) and Mugera *et al.* (2016). This list can go on for many pages, yet we need to get to business, introducing the DEA for a wide audience.

The main goal of this monograph is to provide a compendious overview of DEA and its major variations mainly from a practical perspective.[3] The economic theory for the models behind the DEA estimators has been covered extensively in various sources, most recently in Zelenyuk (2021), whose notation we try to follow here (with some refinements) for the sake of consistency.

In the following sections, we start by introducing the canonical envelopment-type estimators in the production function in Section 2, their advancements in Section 3, and the estimation of cost, revenue, and profit efficiency in Section 4.1. The reader is then introduced to several more advanced streams of DEA literature: the productivity indexes with DEA (Section 5), the statistical properties (Section 6), including the recent development in aggregation, bias-correction, explanation of efficiency, etc., and the "two-stage DEA" for explaining the inefficiency (Section 7). Finally, we also provide demonstrations of some prevalent DEA estimators in R (R Core Team, 2023) and other software with the hospital data in Appendix C.

[1] See Henriques *et al.* (2020) and Zelenyuk and Zelenyuk (2021) for related discussions.

[2] Also see more studies as surveyed by Zhou *et al.* (2008), Zhang and Choi (2014), and Sueyoshi *et al.* (2017).

[3] For another major alternative approach, see Sickles *et al.* (2020) and Nguyen *et al.* (2022a) for overviews of the SFA paradigm and applications in statistical tools, Sickles *et al.* (2024) for utilizing SFA in healthcare, and Sickles and Zelenyuk (2019, Chapter 11–16) for more comprehensive treatment.

Appendices

A

Steps of the SZ07 Bootstrapping Method for Efficiency Aggregation

(1) **Estimate the aggregate efficiency**: Deploy the selected estimator on the original data set to estimate individual efficiency of all DMUs. Then obtain the aggregate estimate of each subgroup and of the population, denoted as $\widehat{\overline{\mathcal{TE}}}^{l}, l = 1, \ldots, L$ and $\widehat{\overline{\mathcal{TE}}}$ with (6.5) and (6.7), respectively.

(2) **Generate the b$^{\text{th}}$ bootstrap sample**: For each subgroup l of size n_l $(l = 1, \ldots, L)$, determine a sample size $m_l = \lfloor n_l^{\delta} \rfloor, \delta \in (0, 1)$ and resample m_l out of the n_l allocations independently, uniformly, and with replacement. The generated bootstrap sample for each group can be denoted as $\chi^{*,l}_{m_l,b} = \{(x_b^{*,l,j}, y_b^{*,l,j}) | j = 1, \ldots, m_l; l = 1, \ldots, L\}$, meanwhile the pool of bootstrap samples is $\chi^{*}_{m,b} = \{\chi^{*,l}_{m_l,b}, l = 1, \ldots, L; b = 1, \ldots, B\}, m = \sum_{l=1}^{L} m_l$.

(3) **Estimate the b$^{\text{th}}$ bootstrap aggregate efficiency:** For the b^{th} bootstrap subsample, estimate the aggregate efficiency (i.e., $\widehat{\overline{\mathcal{TE}}}^{*}_{b}$ and $\widehat{\overline{\mathcal{TE}}}^{*,l}_{b}, l = 1, \ldots, L$) as in the first step, but using the frontier constructed by the pooled bootstrap sample $\chi^{*}_{m,b}$.

(4) **Repeat:** Step 2 and 3 for B times and record B bootstrap estimates of the interested aggregate efficiency, e.g., $\{\widehat{\overline{\mathcal{TE}}}_b^*\}_{b=1}^B$ and $\{\widehat{\overline{\mathcal{TE}}}_b^{*,l}\}_{b=1}^B, l = 1, \ldots, L$.

(5) **Obtain the bootstrap goals:** Use the collected results to infer the true efficiency, e.g., construct the confidence intervals of $\overline{\mathcal{TE}}$ and perform statistical tests with the distribution of $\{\widehat{\overline{\mathcal{TE}}}_b^*\}_{b=1}^B$, as well as correct the bias of $\widehat{\overline{\mathcal{TE}}} - \overline{\mathcal{TE}}$ through $\{\widehat{\overline{\mathcal{TE}}}_b^* - \widehat{\overline{\mathcal{TE}}}\}_{b=1}^B$.

B

Steps of the Two-Stage Double Bootstrap Method in SW07

(1) **Estimate the efficiency**: Apply the selected DEA/FDH estimator on the original data set χ^n to obtain the individual efficiency estimates $\widehat{\mathcal{OTE}}(x^i, y^i|\chi^n)$ of all DMUs.

(2) **Estimate the parameters of the truncated regression:** Use the MLE to estimate the parameters specified in the function of efficiency and obtain $\widehat{\beta}$ and $\widehat{\sigma}_\varepsilon$.

(3) **Obtain the bootstrap estimates of efficiency:** Run the bootstrap process of B_1 times as follows.

(3.1) Draw $\widehat{\varepsilon}_b^i$ for $i = 1, \ldots, n$, from the distribution of $N(0, \widehat{\sigma}_\varepsilon^2)$ with the left truncation as $\widehat{\varepsilon}_b^i \geq 1 - g(z^i|\widehat{\beta})$.

(3.2) Estimate the bootstrap efficiency with the results above as $\mathcal{OTE}_b^*(x^i, y^i|\chi^n) = g(z^i|\widehat{\beta}) + \widehat{\varepsilon}_b^i$.

(3.3) For $i = 1, \ldots, n$, define $x_b^{i*} = x^i$, $y_b^{i*} = y^i[\widehat{\mathcal{OTE}}(x^i, y^i|\chi^n)/\mathcal{OTE}_b^*(x^i, y^i|\chi^n)]$, $z_b^{i*} = z^i$.

(3.4) Estimate $\widehat{\mathcal{OTE}}_b^*(x^i, y^i|\chi^n)$ using the same estimator as in the first step using the bootstrap data set $\chi_b^{n*} = \{(x_b^{i*}, y_b^{i*})|i = 1, \ldots, n\}$.

(4) **Correct the bias of the efficiency estimates:** For $i = 1, \ldots, n$, obtain $\widehat{\mathcal{OTE}}^{i}_{bc}(x^i, y^i|\chi^n) = \widehat{\mathcal{OTE}}(x^i, y^i|\chi^n) - \widehat{\mathrm{Bias}}(\widehat{\mathcal{OTE}}(x^i, y^i|\chi^n))$, where the bias is bootstrap-estimated in a procedure akin to the section 6.5.2.

(5) **Estimate the parameters based on the corrected efficiency:** Apply MLE on the truncated regression of $\widehat{\mathcal{OTE}}^{i}_{bc}(x^i, y^i|\chi^n)$ on z^i to obtain the estimates $(\widehat{\widehat{\beta}}, \widehat{\widehat{\sigma}}_\varepsilon)$.

(6) **Obtain the bootstrap estimates of the parameters:** Run the bootstrap process as follows for B_2 times.

(6.1) Draw $\widehat{\widehat{\varepsilon}}^{i}_{b}$ for $i = 1, \ldots, n$, from the distribution of $N(0, \widehat{\widehat{\sigma}}^{2}_{\varepsilon})$ with the left truncation as $\widehat{\widehat{\varepsilon}}^{i}_{b} \geq 1 - g(z^i|\widehat{\widehat{\beta}})$.

(6.2) Estimate the double-bootstrap efficiency as $\mathcal{OTE}^{**}_{b}(x^i, y^i|\chi^n) = g(z^i|\widehat{\widehat{\beta}}) + \widehat{\widehat{\varepsilon}}^{i}_{b}$.

(6.3) Apply MLE as in the fifth step on the truncated regression of $\mathcal{OTE}^{**}_{b}(x^i, y^i|\chi^n)$ on z^i to obtain the estimates $(\widehat{\widehat{\beta}}^{*}_{b}, \widehat{\widehat{\sigma}}^{*}_{\varepsilon,b})$.

(7) **Construct the confidence intervals of the parameters:** The refined estimates $(\widehat{\widehat{\beta}}, \widehat{\widehat{\sigma}}_\varepsilon)$ and the bootstrap estimates $(\widehat{\widehat{\beta}}^{*}_{b}, \widehat{\widehat{\sigma}}^{*}_{\varepsilon,b})$ can be used to construct a confidence interval and hence the other interested statistics accordingly. For a significance level of α, there are values (a_α, b_α) for the jth element of $\widehat{\widehat{\beta}}$ that $\Pr[a_\alpha \leq (\beta^j - \widehat{\widehat{\beta}}^{j}) \leq b_\alpha] = 1 - \alpha$. As the distribution of $(\beta^j - \widehat{\widehat{\beta}}^{j})$ is unavailable, the bootstrap estimates can approximate the values that $\Pr[a^*_\alpha \leq (\widehat{\widehat{\beta}}^{j} - \widehat{\widehat{\beta}}^{*j}_{b}) \leq b^*_\alpha] \approx 1 - \alpha$. With $B_2 \to \infty$, the confidence interval of β^j can be estimated as $[\widehat{\widehat{\beta}}^{j} + a^*_\alpha, \widehat{\widehat{\beta}}^{j} + b^*_\alpha]$. The confidence interval of the elements in σ_ε can be constructed analogously.

C

Programming and Computations

C.1 DEA Estimators in Statistical Tools

As one of the primary approach in efficiency analysis, a considerable amount of effort has been devoted into the programming and application of the DEA estimators. Specialized software has been developed for deploying DEA, e.g., DEAP (Coelli, 1996) and PIM-DEA (Emrouznejad and Thanassoulis, 2014). One can also apply DEA with Excel Solver amalgamated with add-ins, e.g., DEA Solver Pro (Cooper *et al.*, 2007) and DEAFrontier (Zhu, 2009).[1]

Among others, the DEA estimators can be conducted in a variety of prevalent statistical tools, e.g., R, MATLAB, Stata, etc. One option is utilizing the functional user-written commands to estimate the specified estimators directly. Prior to the illustrations of the widespread packages and commands, we first implement the LP problems of DEA step-by-step in a statistical programming environment. In this way, the code is also more flexible than the pre-programmed commands adapting to other variants of DEA.

For an illustration, we generate a sample of seven DMUs with one input and two outputs. As shown in Box C.1 and C.2, the output-oriented

[1] See Appendix C.3 for an illustration of utilizing the Excel Solver with a numerical example.

```
clear all
close all
%%%% Illustration of Output-Oriented DEA in Matlab

%% Generate X and Y
y1 = [1, 2, 1, 0.5, 1.5, 2, 0]';
y2 = [2, 1, 1, 1.5, 0.5, 0, 2]';
x1 = [1, 1, 1, 1  , 1  , 1, 1]';
Y = [y1, y2];
X = [x1];

%% Define parameters
rts = 'CRS';
% rts = 'VRS';

M = size(Y,2); N = size(X,2); n = size(X,1);
lb = [1; zeros(n,1)];  % Lower bounds for decision variables
ub = [];               % Upper bounds

theta_all = [];

% Choose between VRS and CRS
if strcmp('VRS', rts)
    Aeq = [0, ones(1,n)];  % Matrix for linear equality constraints;
    beq = [1];             % Vector for linear equality constraint;
 elseif strcmp('CRS', rts)
    Aeq = [];              % CRS: no extra equality constraint
    beq = [];
end

%% Estimate theta_j
for j=1:n
x_j = X(j,:);  % Select x and y for a DMU of interest, j
y_j = Y(j,:);

f = -[1; zeros(n,1)];  % Parameters of the objective function
Aineq = [zeros(N,1), X' ; y_j', -Y'];  % Matrix for linear inequality
     constraints
bineq = [x_j'; zeros(M,1)];            % Vector for linear inequality
     constraints

[x,fval,exitflag,output,lambda] = linprog(f,Aineq,bineq,Aeq,beq,lb,ub);
   theta_DEA = -fval;  % Minus, because "min" was used instead of "max"
theta_all = [theta_all; theta_DEA];  % Collect estimates
end

%% Print the results
disp('Firm #    Estimated efficiency')
disp([(1:n)', theta_all])
```

Box C.1: MATLAB code snippet for Output-Oriented DEA

```
### Illustration of Output-Oriented DEA in R ###

# Generate sample data
y1 = c(1, 2, 1, 0.5, 1.5, 2, 0)
y2 = c(2, 1, 1, 1.5, 0.5, 0, 2)
x1 = c(1, 1, 1, 1, 1, 1, 1)
Y = as.data.frame(cbind(y1, y2))
X = as.data.frame(x1)

# Define parameters
rts='CRS'
# rts='VRS'

M = length(Y)
N = length(X)
n = nrow(X)
# Choose between VRS and CRS
if (rts=='VRS'){
  Aeq = cbind(0, t(rep(1,n))) # Matrix for linear equality constraints;
  beq = 1                     # Vector for linear equality constraint;
}
if (rts=='CRS'){
  Aeq <- vector(mode = "numeric", length = 0)
  beq <- vector(mode = "numeric", length = 0)
}

library(lpSolve)
solution <- vector(mode = "numeric", length = 0)
# Estimate theta-i for each individual
for (i in 1:n){
xi = X[i,]
yi = Y[i,]
objx <- vector(mode = "numeric", length = 0)
f = t(-cbind(1, t(rep(0,n)))) # Objective function
Aineq = rbind(cbind(rep(0,N), t(X)),
            cbind(t(yi), -t(Y))) # Matrix for linear inequality
                 constraints
bineq = t(cbind(xi, t(rep(0,M))))  # Vector for linear inequality
     constraints
direction = c(rep("<=",N+M),"=") # Set the directions including the
     equivalent constraint
A = rbind(Aineq,Aeq)
b = rbind(bineq,beq)

lp = lp (direction = "min", objective.in = f, const.mat = A,
    const.dir = direction, const.rhs = b) # Note that the bounds of every
        variable has been assumed in lp
solution = rbind(solution, lp[["solution"]][1]) # Collect estimates
}
solution
```

Box C.2: R code snippet for Output-Oriented DEA

DEA estimator can be programmed in a form of LP problems by transforming the constraints. The optimization problem can be then solved with the LP solver in MATLAB (command **lingprog**) and R (function **lp** in package **lpSolve** (Berkelaar *et al.*, 2023)), respectively. For example, when assuming CRS, the efficiency scores estimated using the two snippets of code are identical as $\{1.00, 1.00, 1.50, 1.33, 1.33, 1.00, 1.00\}$.

C.2 Applications with User-Written Packages

C.2.1 Sample Data

A demonstration with a real data set may illustrate the DEA techniques, covered in this work, more explicitly in practice. We use the data from Queensland Health (QH), regarding 95 public hospitals in Queensland, Australia, during a four-year period (FY 2012/13 to FY 2015/16). The input-output allocation follows the seminal practice on hospital efficiency analysis (e.g., Grosskopf and Valdmanis, 1987; Rosko, 2001) and the recent studies on Queensland public hospital (e.g., Nguyen and Zelenyuk (2021a,b; Wang and Zelenyuk, 2024b).[2] It contains three inputs, representing the aggregated labor, consumable goods, and capital, and one output, the aggregated inpatient and outpatient services.[3] Moreover, three environmental variables, i.e., the location, size, and teaching functions, are considered in the regression of efficiency estimates. The descriptive statistics of these variables are as summarized in Table C.1.

C.2.2 Practical Applications of Technical Efficiency

The user-written commands are developed in a range of popular statistical tools, which are practical for the commonly applied DEA estimators. We focus on the R environment (R Core Team, 2023) in this monograph,

[2]See also Wang and Zelenyuk (2024a) for overview and illustrations of the efficiency analysis techniques in healthcare.

[3]The aggregation of the labor input and output variables follows the process in Nguyen and Zelenyuk (2021a) and Wang and Zelenyuk (2024b), based on a Principal Component Analysis (PCA) type approach introduced by Daraio and Simar (2007c).

Table C.1: Descriptive statistics of Queensland public hospitals, FY 2012/13 to FY 2015/16

Variable	Description	Mean	Std Dev	Min	Max
Input					
LABOR	Aggregated labor input	0.76	1.57	0.01	8.71
BED	Number of beds	74.92	133.78	3.00	680.00
SUPP*	Consumable expenditure	7.83	19.20	0.03	164.00
Output					
OUT	Inpatient and outpatient service	0.54	1.03	0.01	5.05
Variable	**Description**	**Frequency**	**Percentage**		
Environmental					
REMOTE	Located in remote areas	108	28.42%		
SMALL	Small hospitals	300	78.95%		
TEACH	Teaching hospitals	70	18.42%		

Note: *AUD 1,000,000 in constant price of FY2012/2013.

where the conventional DEA models can be constructed with some powerful packages, e.g., **Benchmarking** by Bogetoft and Otto (2022).[4, 5] The technical efficiency can be estimated with the DEA estimators discussed above as illustrated in the snippet of R code in Box C.3.[6, 7]

Both the output-oriented and input-oriented technical efficiency are estimated with the DEA estimators in CRS, VRS, or NIRS, or the FDH estimators through the functions in **Benchmarking**, respectively. The function **dea** in **Benchmarking** is feasible of all these scenarios, where for a more comparable efficiency score between $[0, 1]$, the output-oriented results are reciprocally transformed.

[4]Some other packages in R, e.g., **FEAR** by Wilson (2020)**, rDEA** by Simm and Besstremyannaya (2020), and **deaR** by Coll-Serrano *et al.* (2022), also provide extensive functions, including the advanced applications in DEA, such as the bias-corrected efficiency, Malmquist productivity index, bootstrapping DEA, etc., more detail of which will be discussed in later sections.

[5]DEA is also well developed in MATLAB, as introduced by Sickles and Zelenyuk (2019). Interested readers can also follow the code they provided at: https://sites.google.com/site/productivityefficiency/home. One can also use the commands by Ji and Lee (2010) to apply DEA estimators in Stata.

[6]See Appendix D for the complete R code of the illustrations on the sample data, including the data process procedures and plotting.

[7]See also Sickles *et al.* (2020) for a guidance of the DEA applications in the R environment.

```
################### Technical Efficiency ###################

attach(data)

# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Output-oriented
dea.crs.out.bench = 1/dea(X, Y, RTS="crs", ORIENTATION="out")$eff # CRS
dea.vrs.out.bench = 1/dea(X, Y, RTS="vrs", ORIENTATION="out")$eff # VRS
dea.nirs.out.bench = 1/dea(X, Y, RTS="drs", ORIENTATION="out")$eff # NIRS
dea.fdh.out.bench = 1/dea(X, Y, RTS="fdh", ORIENTATION="out")$eff # FDH

# Input-oriented
dea.crs.in.bench = dea(X, Y, RTS="crs", ORIENTATION="in")$eff
dea.vrs.in.bench = dea(X, Y, RTS="vrs", ORIENTATION="in")$eff
dea.nirs.in.bench = dea(X, Y, RTS="drs", ORIENTATION="in")$eff
dea.fdh.in.bench = dea(X, Y, RTS="fdh", ORIENTATION="in")$eff

# Summarize the estimations
Effi <- as.data.frame(cbind(dea.crs.out.bench, dea.vrs.out.bench,
                            dea.nirs.out.bench, dea.fdh.out.bench,
                            dea.crs.in.bench, dea.vrs.in.bench,
                            dea.nirs.in.bench, dea.fdh.in.bench))
summary(Effi)
```

Box C.3: R code snippet for technical efficiency

The summarized statistics and the distribution of the efficiency estimates are as reported in Figures C.1 and C.2. The estimates with different returns to scale specifications lead to significantly different

```
> summary(Effi)
 dea.crs.out.bench dea.vrs.out.bench dea.nirs.out.bench dea.fdh.out.bench
 Min.   :0.1182    Min.   :0.1454    Min.   :0.1322     Min.   :0.1733
 1st Qu.:0.2703    1st Qu.:0.3698    1st Qu.:0.3529     1st Qu.:0.5444
 Median :0.3565    Median :0.6192    Median :0.5842     Median :0.8330
 Mean   :0.3666    Mean   :0.5924    Mean   :0.5728     Mean   :0.7477
 3rd Qu.:0.4281    3rd Qu.:0.8102    3rd Qu.:0.7977     3rd Qu.:1.0000
 Max.   :1.0000    Max.   :1.0000    Max.   :1.0000     Max.   :1.0000
 dea.crs.in.bench dea.vrs.in.bench dea.nirs.in.bench dea.fdh.in.bench
 Min.   :0.1182   Min.   :0.1391   Min.   :0.1182    Min.   :0.1556
 1st Qu.:0.2703   1st Qu.:0.4300   1st Qu.:0.2824    1st Qu.:0.6376
 Median :0.3565   Median :0.5872   Median :0.4204    Median :0.8571
 Mean   :0.3666   Mean   :0.6068   Mean   :0.4950    Mean   :0.7966
 3rd Qu.:0.4281   3rd Qu.:0.7656   3rd Qu.:0.7217    3rd Qu.:1.0000
 Max.   :1.0000   Max.   :1.0000   Max.   :1.0000    Max.   :1.0000
```

Figure C.1: Screenshot of the summarized statistics of the technical efficiency estimates of different estimators in R

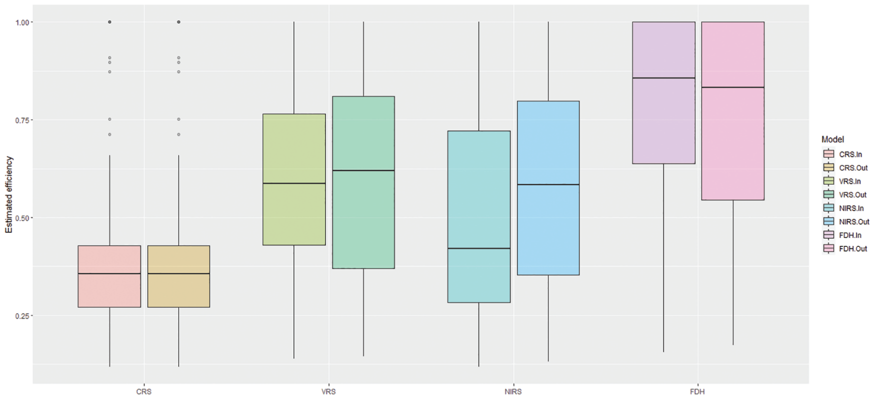


Figure C.2: Boxplot of technical efficiency estimates

results. The mean efficiency estimated by DEA-CRS is 36.66%, which may be due to the outliers in the sample, as indicated in the boxplot. Meanwhile in comparison, when NIRS assumption is applied, i.e., the outputs would not change equiproportionally as the inputs, the mean efficiency is estimated at 57.28%. The estimates of the FDH estimators are higher than the DEA estimators in general, and indicate a significant proportion of efficient units. Besides, as mentioned above, the reciprocal relationship between $\mathcal{OTE}$ and $\mathcal{ITE}$ is also exhibited. As the results of one orientation is in reciprocal, the output-oriented and input-oriented DEA-CRS estimates are identical to each other.

C.2.3 Practical Applications of Other Efficiency Measures

The estimators of cost, revenue, and profit efficiency are also feasible in the **Benchmarking** package (Bogetoft and Otto, 2022), which can be deployed with the **cost.opt**, **revenue.opt**, and **profit.opt** function, respectively. The applications are as illustrated in Box C.4.[8, 9] Besides the input-output allocations, vectors of prices are required for both inputs and outputs. Since the price variables are not available in our

[8]Inputs and outputs are predefined. See Appendix D for the complete R code including the data process procedures.

[9]CRS assumption is applied for an illustration, while the other commonly deployed assumptions on the returns to scale are also available.

```
################### Cost/Revenue/Profit Efficiency ###################

attach(data)

# Generate artificial matrix of prices
w <- t(as.matrix(c(1,2,3)))
p <- as.matrix(4)

# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Cost efficiency
xopt = cost.opt(X, Y, w, RTS='crs') #CRS
cobs <- X %*% t(w) # Observed Cost
copt <- xopt$x %*% t(w) # Optimal Cost
dea.crs.cost = copt/cobs # cost efficiency

# Revenue efficiency
yopt = revenue.opt(X, Y, p, RTS='crs') #CRS
yobs <- Y %*% p # Observed Revenue
yopt <- yopt$y %*% p # Optimal Revenue
dea.crs.revenue = yobs/yopt # Revenue efficiency

# Profit efficiency
popt = profit.opt(X, Y, w, p, RTS='crs') #CRS
pobs <- Y %*% p -X %*% t(w) # Observed Profit
popt <- popt$y %*% p - popt$x %*% t(w) # Optimal Profit
dea.crs.revenue = pobs/popt # Profit efficiency

# Other RTS assumptions can be analogously applied
```

Box C.4: R code snippet for cost, revenue, and profit efficiency

sample data, we generate simulated data for vectors w and p to represent the prices of the three inputs and one output, respectively.[10] The functions return the optimal level of input or/and output regarding the efficiency type based on the price vectors. Consequently, the efficiency is estimated as a ratio between the optimal and real allocations as discussed in Section 4.1.

C.2.4 Practical Applications of MPI

The MPI can be directly estimated using the **malmq** function in **Benchmarking** (Bogetoft and Otto, 2022). As illustrated in Box C.5,

[10] Based on these artificial prices, the efficiency estimates are for examples only.

```
################### MPI ###################

attach(data)

# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Use period 1 and 2 as an example
X0 <- as.matrix(X[Year==1,])
X1 <- as.matrix(X[Year==2,])

Y0 <- as.matrix(Y[Year==1,])
Y1 <- as.matrix(Y[Year==2,])

ID0 <- as.matrix(id[Year==1])
ID1 <- as.matrix(id[Year==2])

mpi.crs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "in
    ")$m
mpi.vrs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "vrs", ORIENTATION = "in
    ")$m
mpi.nirs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "drs", ORIENTATION = "
    in")$m
mpi.crs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "
    out")$m
mpi.vrs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "vrs", ORIENTATION = "
    out")$m
mpi.nirs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "drs", ORIENTATION = "
    out")$m

# Summarize the estimations
MPI <- as.data.frame(cbind(mpi.crs.in, mpi.vrs.in,
                            mpi.nirs.in, mpi.crs.out,
                            mpi.vrs.out, mpi.nirs.out))
summary(MPI)

# Estimate with DEA estimators (F re et al. 1992)
dea00<-dea(X0, Y0, RTS="crs", ORIENTATION="out", XREF=X0, YREF=Y0)$eff
dea11<-dea(X1, Y1, RTS="crs", ORIENTATION="out", XREF=X1, YREF=Y1)$eff
dea10<-dea(X0, Y0, RTS="crs", ORIENTATION="out", XREF=X1, YREF=Y1)$eff
dea01<-dea(X1, Y1, RTS="crs", ORIENTATION="out", XREF=X0, YREF=Y0)$eff

mpi.dea<-sqrt(dea01/dea00*dea11/dea10)
summary(mpi.dea) # Same results as above
summary(mpi.crs.out)
```

Box C.5: R code snippet for MPI

using the first two periods (FY 2012/13 and FY 2013/14) as an example, the MPIs of the 95 hospitals in Queensland are estimated under different returns to scale assumptions (CRS, VRS, and NIRS) and in input- and output-orientation, respectively.

The summarized statistics of the outcome is as shown in Figure C.3. The mean MPI of the hospitals is similar across different specifications, which indicate a moderate rise of productivity in general between the two periods. The highest MPI in different models is around 1.7, while the minimum value of input-oriented VRS and NIRS models are 0, which may be due to the specified model not conforming the data generating process in reality. Nevertheless, the Pearson correlation between each pair of the models is higher than 0.7.

As indicated above, DEA estimators can be applied in the estimation of the distance functions (Färe *et al.*, 1992). In the second part of the R code snippet, we deploy the DEA estimators by the aforementioned **dea** function in **Benchmarking**, and estimate the MPI accordingly. Using the output-oriented VRS specification as an example, the results are identical to the estimates of the **malmq** function. Note that as indicated above, the Farrell-type technical efficiency is reciprocal of the Shephard's distance function, and hence the results of the **dea** function is not reciprocally transformed as when they are used to estimate the technical efficiency in Box C.5.

For a more intuitive visualization of the productivity change, we can estimate the MPI of hospitals between each pair of adjacent periods.[11] As shown in the heatmap in Figure C.4, the MPIs are aggregated by the 14 Local Hospital Networks.[12] The hospitals in HHS 327 between FY 2013/14 and FY 2014/15 exhibit the highest productivity improvement, followed by HHS 315 and 317 in the earlier period. In contrast, hospitals in HHS 313, 317, and 326 in the last period of study are estimated with a decreasing productivity in general with a MPI estimate close to 0.5.

[11] The input-oriented MPI under CRS assumption is displayed as an demonstration. See Appendix D for the complete R code of the illustrations, including the plotting of the heatmap.

[12] Local Hospital Networks in Queensland are often referred to as Hospital and Health Services (HHSs). There are 15 geographical HHSs, and 14 of them are recorded in our sample.

```
> summary(MPI)
  mpi.crs.in        mpi.vrs.in        mpi.nirs.in       mpi.crs.out       mpi.vrs.out       mpi.nirs.out
 Min.   :0.6092   Min.   :0.0000   Min.   :0.0000   Min.   :0.6202   Min.   :0.3428   Min.   :0.6199
 1st Qu.:0.9017   1st Qu.:0.9184   1st Qu.:0.8929   1st Qu.:0.9186   1st Qu.:0.9362   1st Qu.:0.9319
 Median :0.9787   Median :1.0118   Median :0.9947   Median :1.0218   Median :1.0002   Median :1.0054
 Mean   :1.0005   Mean   :1.0151   Mean   :0.9965   Mean   :1.0359   Mean   :1.0329   Mean   :1.0339
 3rd Qu.:1.0887   3rd Qu.:1.0926   3rd Qu.:1.0890   3rd Qu.:1.1090   3rd Qu.:1.0959   3rd Qu.:1.1008
 Max.   :1.6124   Max.   :1.6195   Max.   :1.7214   Max.   :1.6414   Max.   :1.7236   Max.   :1.6272
                                                                      NA's   :4
> summary(mpi.dea) # Same results as above
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.6202  0.9186  1.0218  1.0359  1.1090  1.6414
> summary(mpi.crs.out)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.6202  0.9186  1.0218  1.0359  1.1090  1.6414
```

Figure C.3: Screenshot of the summarized statistics of MPI estimated by different functions in R

```
################### Bias-correction ###################

attach(data)

# Use the functions by Wilson (2020)
library(FEAR)

# Bias-corrected efficiency (CRS-Output-oriented)
Bootstrap.fear = boot.sw98(Xt, Yt, NREP = 2000, RTS = 3, ORIENTATION = 2,
     alpha = 0.05, CI.TYPE=2)
cdea = 1/Bootstrap.fear$dhat.bc
```

Box C.6: R code snippet for bias correction

Consequently, MPI is a useful method in evaluating the hospital performance, especially in a perspective of measuring the productivity change. A cross-sectional comparison of the MPI of hospitals in different groups (i.e., different regions, types, sizes, etc.) would also be one attractive application (i.e., Aragón *et al.* (2019)) as well as evaluating the growth rate of different groups in a study period, when a panel data is available.

C.2.5 Practical Applications of Bias Correction, Kernel Density Estimation, and Aggregation

Some aforementioned techniques in Section 6 can be implemented in R, by cooperating with some user-written packages. For example as shown in Box C.6, the seminal bootstrap method adapted in the DEA/FDH estimators for bias-correction by Simar and Wilson (1998) can be

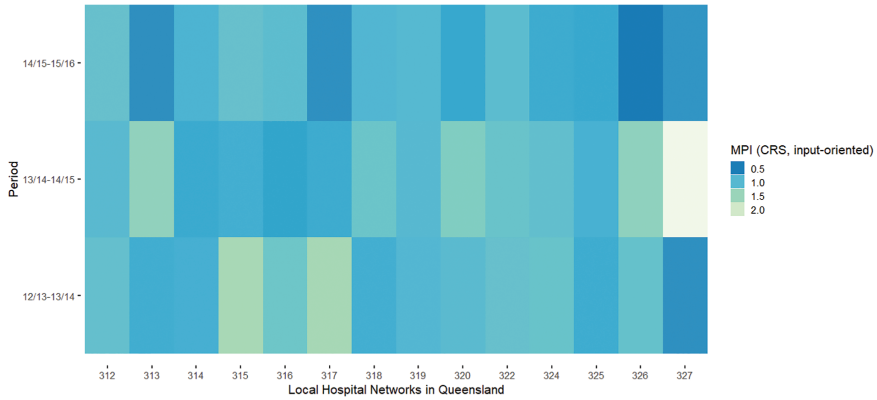


Figure C.4: Heatmap of MPI of HHS across periods

deployed by the **boot.sw98** function in **FEAR** (Wilson, 2020), or the **dea.robust** function in **rDEA** (Simm and Besstremyannaya, 2020), or more alternative functions.[13] The results of different functions are close to each other, i.e., correlations higher than 90%, but not identical due to the random sampling process.

Taking the output-oriented DEA-CRS estimator as an example, as shown in the kernel density plot in Figure C.5,[14] the efficiency bias-corrected by the **boot.sw98** function is relatively higher than the original estimates of the **dea** function. This phenomenon is as expected due to that the estimator of the output-oriented efficiency is downward biased ($1 \leq \widehat{\mathcal{OTE}}(x, y|\widehat{\Psi}) \leq \mathcal{OTE}(x, y|\Psi) \leq \infty$).

The aggregate of efficiency is also feasible in R. Due to unavailable price variables in our sample, we generate a simple vector of price p. As shown in Box C.8, the aggregated efficiency of the population following (6.5) and (6.7) ($L = 1$) is applicable with few lines of code using the basic commands in R. Taking output-oriented DEA-CRS as an example, we obtain an efficiency level of 0.4038 in aggregate, which is slightly higher than the simple mean of the individual estimates (0.3666).

[13] An alternative is the **dea.boot** function in **Benchmarking**, which is however slower in computation than **boot.sw98** (Bogetoft and Otto, 2022).

[14] Using Gaussian kernel and bandwidth selected by cross-validation as shown in Box C.7.

```
################### Kernel density plot ###################

# Compare the original and bias-corrected estimates
dea = dea(Xt, Yt, RTS=3, ORIENTATION=2)
cdea = cbind(cdea,rep(1,380))
dea = cbind(dea,rep(0,380))

correction = as.data.frame(rbind(cdea,dea))
colnames(correction) = c("DEA","method")

correction$Estimates[correction$method==1]='Bias-corrected'
correction$Estimates[correction$method==0]='Original'

attach(correction)
require("ggplot2")
.df <- na.omit(data.frame(x = correction$DEA))
.nbins <- pretty(range(.df$x), n = nclass.FD(.df$x), min.n = 1)
.dea <- ggplot(data = .df, aes(x = x, y = ..density..)) +
  # Epanechnikov kernel and CV bandwidth
  geom_density(
    kernel = "gaussian",
    bw = "ucv",
    alpha = 0.5,
    aes(color = Estimates, fill = Estimates)
  ) +
  scale_y_continuous(expand = c(0.01, 0)) +
  xlab("Estimated inefficiency") +
  ylab("Estimated Density") +
  labs(colour = "Estimates",
       shape = "Estimates",
       fill = "Estimates") +
  RcmdrPlugin.KMggplot2::theme_simple(base_size = 14, base_family = "sans")
print(.dea)
rm(.df, .nbins)

# Use the functions by Simm and Besstremyannaya (2020)
library(rDEA)
Bootstrap.rDEA = dea.robust(X, Y, W=NULL, model="output", RTS="constant", B
     =2000, alpha=0.05, bw="bw.ucv")
```

Box C.7: R code snippet for kernel density plot

C.2.6 Practical Applications of Two-Stage DEA

The SW07 methods are predominantly applied in the literature, which can be conveniently applied in various mainstream environments using the user-written packages. For example in Stata, the two algorithms in SW07 can be conducted through the **simarwilson** command by (Badunenko and Tauchmann, 2019) with the **teradial** command

```
################## Aggregate ##################

# Generate artificial matrix of prices
p <- as.matrix(4)

# Calculate weight
weight = p%*%Yt/sum(p%*%rowSums(Yt))

# Aggregate
aggregate = sum(dea.crs.out.fear%*%t(weight))
aggregate
mean(dea.crs.out.bench)
```

Box C.8: R code snippet for aggregate efficiency

(Badunenko and Mozharovskyi, 2016) in estimating the efficiency scores in prior.[15]

In the focused R environment in this monograph, the steps of SW07 can be achieved by incorporating multiple functions. For example, the bias-corrected efficiency score can be estimated with the functions illustrated in the last section. Meanwhile, the bootstrap truncated regression can be conducted with the **truncSP** package by Karlsson and Lindmark (2014). The **treg** function in **FEAR** (Wilson, 2020) is also applicable in the estimation of a truncated regression with MLE. Nevertheless, the second algorithm in SW07, as discussed above, can be directly deployed with the **dea.env.robust** function in the **rDEA** package (Simm and Besstremyannaya, 2020). As illustrated in Box C.9, in our sample, we define a vector of environmental variable as $Z = \{\text{TEACH, SMALL, REMOTE}\}$, which exhibits the teaching functions, the size, and the location of the hospitals, respectively. Taking the output-oriented technical efficiency under CRS as an example, by setting $B_1 = 100, B_2 = 2000, \alpha = 0.05$ (the bootstrap time of efficiency and parameter estimation and the significance level), we obtain a result as in Figure C.6.[16] As the Farrell-type technical efficiency is employed, a higher score reflects a lower level of efficiency (further from the frontier). Hence, the negative coefficient of the teaching functions

[15] See also the MATLAB code for the two-stage DEA provided by Sickles and Zelenyuk (2019) at https://sites.google.com/site/productivityefficiency/home.

[16] See more detailed steps of SW07 in Appendix A.

```
################## SW07 ##################

attach(data)

# Use the functions by Simm and Besstremyannaya (2020)
library(rDEA)

# Define environmental variables
Z = as.matrix(cbind(TEACH, Small, Remote))

# Output-oriented & CRS
sw07 = dea.env.robust(X, Y, W=NULL, Z, "output", RTS="constant", L1=100, L2
     =2000, alpha=0.05)
sw07$beta_hat_hat
sw07$beta_ci
```

Box C.9: R code snippet for SW07

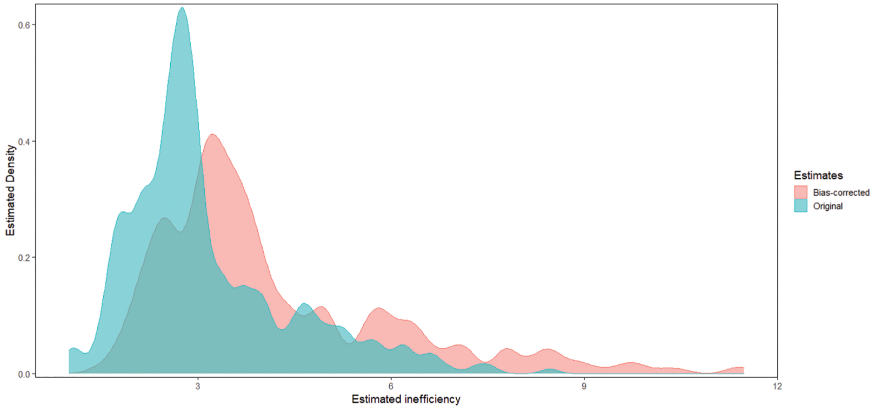


Figure C.5: Estimated kernel densities of the original and bias-corrected efficiency level

indicate that teaching hospitals tend to perform more efficiently, whereas the positive coefficients of the other two environmental variables indicate that hospitals in smaller sizes or in remote areas exhibit a tendency of being estimated as more inefficient. As shown in the confidence interval estimated below, the coefficients are at least significant at a 5% level of significance.

```
> sw07$beta_hat_hat
(Intercept)       TEACH       Small      Remote
  1.1840621  -1.3537605   2.1275255   0.5960517
> sw07$beta_ci
                  2.5%       97.5%
(Intercept) 0.15897451  2.4377391
TEACH      -2.36600439 -0.2055785
Small       0.92978447  3.1491399
Remote      0.09072817  1.0893119
```

Figure C.6: Screenshot of the results of SW07 method in R

C.3 Illustrations with Excel Solver

The DEA estimators, e.g., in (2.12), (2.20), (2.23), and (2.29), are essentially LP problems, while the FDH estimator is a mixed {0,1}-integer and LP problem. Therefore, the Excel Solver is naturally an alternative option for solving the LP problems due to its accessibility, adaptability, and functional stability. Utilizing the numerical example in Section 2.2.5, the DEA and FDH estimators in (2.30) and (2.31) can be modeled in Excel as illustrated in the left panel from Figure C.7 to Figure C.10.

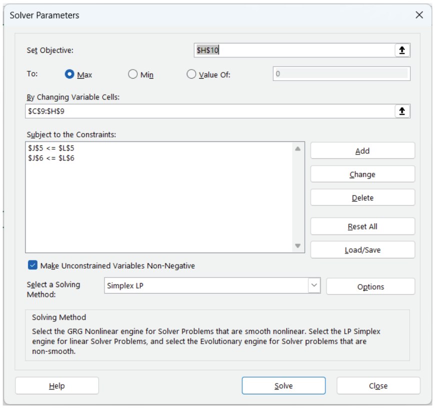
Data

	z^1	z^2	z^3	z^4	z^5	θ	LHS	sign	RHS
Maximize						1	2		
output - y	-1	-2	-3	-4	-2	2	0	<=	0
input - x	1	1	2.5	5	2		2	<=	2

Results

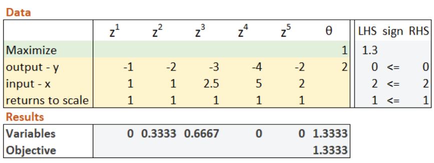
Variables	**0**	**2**	**0**	**0**	**0**	**2**
Objective						**2**

(a) DEA modeling



(b) Settings in Excel Solver

Figure C.7: Illustration of Excel Solver with the numerical sample in Table 2.1 under CRS



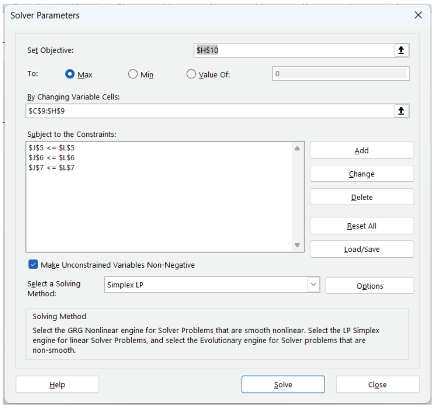
Data

	z^1	z^2	z^3	z^4	z^5	θ	LHS	sign	RHS
Maximize						1	1.3		
output - y	-1	-2	-3	-4	-2	2	0	<=	0
input - x	1	1	2.5	5	2		2	<=	2
returns to scale	1	1	1	1	1		1	<=	1

Results

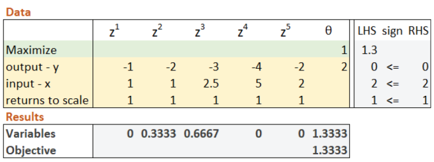
Variables	0	0.3333	0.6667	0	0	1.3333
Objective						1.3333

(a) DEA modeling



(b) Settings in Excel Solver

Figure C.8: Illustration of Excel Solver with the numerical sample in Table 2.1 under NIRS



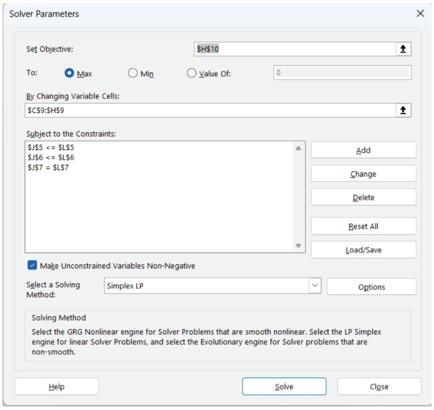
Data

	z^1	z^2	z^3	z^4	z^5	θ	LHS	sign	RHS
Maximize						1	1.3		
output - y	-1	-2	-3	-4	-2	2	0	<=	0
input - x	1	1	2.5	5	2		2	<=	2
returns to scale	1	1	1	1	1		1	<=	1

Results

Variables	0	0.3333	0.6667	0	0	1.3333
Objective						1.3333

(a) DEA modeling



(b) Settings in Excel Solver

Figure C.9: Illustration of Excel Solver with the numerical sample in Table 2.1 under VRS

Each row presents a constraint of an input/output or the assumption of returns to scale. By changing the intensity variables and the efficiency score, the Excel Solver would optimize the efficiency following the corresponding constraints as illustrated in the right panel from Figure C.7 to Figure C.10.

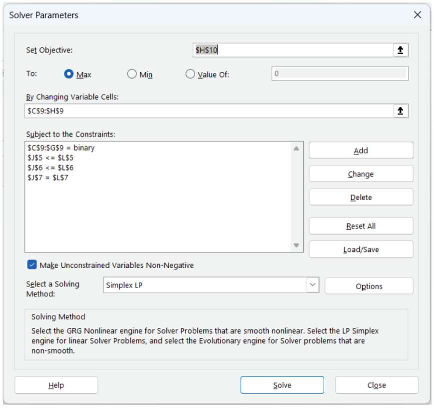
Data

	z^1	z^2	z^3	z^4	z^5	θ	LHS	sign	RHS
Maximize						1	1		
output - y	-1	-2	-3	-4	-2	2	0	<=	0
input - x	1	1	2.5	5	2		1	<=	2
returns to scale	1	1	1	1	1		1	=	1

Results

Variables	0	1	0	0	0	1
Objective						1

(a) FDH modeling



(b) Settings in Excel Solver

Figure C.10: Illustration of Excel Solver with the numerical sample in Table 2.1 under FDH

As can be seen, the Excel Solver provides the same estimates of the efficiency scores and intensity variables as we demonstrated in Section 2.2.5.

D

Full R Code for the Illustrations

```
##################################################################
#####      Illustration of DEA on Queensland hospitals      #####
##################################################################

rm(list=ls())
graphics.off()


################### Data Process ###################

# Read data
data <- read.csv("QLD.csv")
names(data)[names(data)=="HOSID"] <- "id"
names(data)[names(data)=="Yeardummy"] <- "Year"

# Convert to panel data
library(plm)
paneldata<- pdata.frame(data, c("id","Year"))

attach(data)
# Input/Output for "Benchmarking"
X = as.matrix(cbind(BEDS, Agglabours, SUPP))
Y = as.matrix(Aggout)

# Input/Output for "FEAR"
Xt = t(X)
Yt = t(Y)

################### Technical Efficiency ###################

attach(data)
```

```
# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Output-oriented
dea.crs.out.bench = 1/dea(X, Y, RTS="crs", ORIENTATION="out")$eff # CRS
dea.vrs.out.bench = 1/dea(X, Y, RTS="vrs", ORIENTATION="out")$eff # VRS
dea.nirs.out.bench = 1/dea(X, Y, RTS="drs", ORIENTATION="out")$eff # NIRS
dea.fdh.out.bench = 1/dea(X, Y, RTS="fdh", ORIENTATION="out")$eff # FDH

# Input-oriented
dea.crs.in.bench = dea(X, Y, RTS="crs", ORIENTATION="in")$eff
dea.vrs.in.bench = dea(X, Y, RTS="vrs", ORIENTATION="in")$eff
dea.nirs.in.bench = dea(X, Y, RTS="drs", ORIENTATION="in")$eff
dea.fdh.in.bench = dea(X, Y, RTS="fdh", ORIENTATION="in")$eff

# Summarize the estimations
Effi <- as.data.frame(cbind(dea.crs.out.bench, dea.vrs.out.bench,
                            dea.nirs.out.bench, dea.fdh.out.bench,
                            dea.crs.in.bench, dea.vrs.in.bench,
                            dea.nirs.in.bench, dea.fdh.in.bench))
summary(Effi)

################### Cost/Revenue/Profit Efficiency ###################

attach(data)

# Generate artificial matrix of prices
w <- t(as.matrix(c(1,2,3)))
p <- as.matrix(4)

# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Cost efficiency
xopt = cost.opt(X, Y, w, RTS='crs') #CRS
cobs <- X %*% t(w) # Observed Cost
copt <- xopt$x %*% t(w) # Optimal Cost
dea.crs.cost = copt/cobs # cost efficiency

# Revenue efficiency
yopt = revenue.opt(X, Y, p, RTS='crs') #CRS
yobs <- Y %*% p # Observed Revenue
yopt <- yopt$y %*% p # Optimal Revenue
dea.crs.revenue = yobs/yopt # Revenue efficiency

# Profit efficiency
popt = profit.opt(X, Y, w, p, RTS='crs') #CRS
pobs <- Y %*% p -X %*% t(w) # Observed Profit
popt <- popt$y %*% p - popt$x %*% t(w) # Optimal Profit
dea.crs.revenue = pobs/popt # Profit efficiency

# Other RTS assumptions can be analogously applied.

################### MPI ###################
```

```
attach(data)

# Use the functions by Bogetoft and Otto (2022)
library(Benchmarking)

# Use period 1 and 2 as an example
X0 <- as.matrix(X[Year==1,])
X1 <- as.matrix(X[Year==2,])

Y0 <- as.matrix(Y[Year==1,])
Y1 <- as.matrix(Y[Year==2,])

ID0 <- as.matrix(id[Year==1])
ID1 <- as.matrix(id[Year==2])

mpi.crs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "in
     ")$m
mpi.vrs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "vrs", ORIENTATION = "in
     ")$m
mpi.nirs.in = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "drs", ORIENTATION = "
     in")$m
mpi.crs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "
     out")$m
mpi.vrs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "vrs", ORIENTATION = "
     out")$m
mpi.nirs.out = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "drs", ORIENTATION = "
     out")$m

# Summarize the estimations
MPI <- as.data.frame(cbind(mpi.crs.in, mpi.vrs.in,
                            mpi.nirs.in, mpi.crs.out,
                            mpi.vrs.out, mpi.nirs.out))
summary(MPI)

# Estimate with DEA estimators (Fare et al. 1992)
# Same results by manual MPI (when implosion, using: (X2,Y2))
dea00<-dea(X0, Y0, RTS="crs", ORIENTATION="out", XREF=X0, YREF=Y0)$eff
dea11<-dea(X1, Y1, RTS="crs", ORIENTATION="out", XREF=X1, YREF=Y1)$eff
dea10<-dea(X0, Y0, RTS="crs", ORIENTATION="out", XREF=X1, YREF=Y1)$eff
dea01<-dea(X1, Y1, RTS="crs", ORIENTATION="out", XREF=X0, YREF=Y0)$eff

mpi.dea<-sqrt(dea01/dea00*dea11/dea10)
summary(mpi.dea) # Same results as above
summary(mpi.crs.out)

# Heatmap for MPI
IDs = cbind(data$id, data$NetworkID, rep(1,nrow(data)))[1:95,]
X0 <- as.matrix(X[Year==1,])
X1 <- as.matrix(X[Year==2,])
Y0 <- as.matrix(Y[Year==1,])
Y1 <- as.matrix(Y[Year==2,])
ID0 <- as.matrix(id[Year==1])
ID1 <- as.matrix(id[Year==2])
mpi.crs.in.1 = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "
     in")$m
mpi.crs.in.1 = cbind(mpi.crs.in.1, IDs)
```

```
IDs = cbind(data$id, data$NetworkID, rep(2,nrow(data)))[1:95,]
X0 <- as.matrix(X[Year==2,])
X1 <- as.matrix(X[Year==3,])
Y0 <- as.matrix(Y[Year==2,])
Y1 <- as.matrix(Y[Year==3,])
ID0 <- as.matrix(id[Year==2])
ID1 <- as.matrix(id[Year==3])
mpi.crs.in.2 = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "
    in")$m
mpi.crs.in.2 = cbind(mpi.crs.in.2, IDs)

IDs = cbind(data$id, data$NetworkID, rep(3,nrow(data)))[1:95,]
X0 <- as.matrix(X[Year==3,])
X1 <- as.matrix(X[Year==4,])
Y0 <- as.matrix(Y[Year==3,])
Y1 <- as.matrix(Y[Year==4,])
ID0 <- as.matrix(id[Year==3])
ID1 <- as.matrix(id[Year==4])
mpi.crs.in.3 = malmq(X0, Y0, ID0, X1, Y1, ID1, RTS = "crs", ORIENTATION = "
    in")$m
mpi.crs.in.3 = cbind(mpi.crs.in.3, IDs)

mpi.heat = as.data.frame(rbind(mpi.crs.in.1, mpi.crs.in.2, mpi.crs.in.3))
colnames(mpi.heat) <- c("mpi.crs.in", "id", "HHS", "Period")

require("ggplot2")
require("hrbrthemes")
ggplot(mpi.heat, aes(as.character(HHS), Period, fill= mpi.crs.in)) +
  geom_tile() +
  xlab("Local Hospital Networks in Queensland")+
  ylab("Period")+
  scale_fill_distiller(palette = "GnBu")+
  scale_y_discrete(limit = c("12/13-13/14","13/14-14/5","14/15-15/16"))+
  guides(fill=guide_legend(title="MPI (CRS, input-oriented)"))+
  theme_bw(base_size = 16)+
  theme(panel.border = element_blank(), panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), axis.line = element_line(colour
              = "white"))

################## Bias-correction ###################

attach(data)

# Use the functions by Wilson (2020)
library(FEAR)

# Bias-corrected efficiency (CRS-Output-oriented)
Bootstrap.fear = boot.sw98(Xt, Yt, NREP = 2000, RTS = 3, ORIENTATION = 2,
    alpha = 0.05, CI.TYPE=2)
cdea = 1/Bootstrap.fear$dhat.bc

# Compare the original and bias-corrected estimates
dea = dea(Xt, Yt, RTS=3, ORIENTATION=2)
cdea = cbind(cdea,rep(1,380))
dea = cbind(dea,rep(0,380))
```

```
correction = as.data.frame(rbind(cdea,dea))
colnames(correction) = c("DEA","method")

correction$Estimates[correction$method==1]='Bias-corrected'
correction$Estimates[correction$method==0]='Original'

################## Kernel density plot ##################

attach(correction)
require("ggplot2")
.df <- na.omit(data.frame(x = correction$DEA))
.nbins <- pretty(range(.df$x), n = nclass.FD(.df$x), min.n = 1)
.dea <- ggplot(data = .df, aes(x = x, y = ..density..)) +
  # Epanechnikov kernel and CV bandwidth
  geom_density(
    kernel = "gaussian",
    bw = "ucv",
    alpha = 0.5,
    aes(color = Estimates, fill = Estimates)
  ) +
  scale_y_continuous(expand = c(0.01, 0)) +
  xlab("Estimated inefficiency") +
  ylab("Estimated Density") +
  labs(colour = "Estimates",
       shape = "Estimates",
       fill = "Estimates") +
  RcmdrPlugin.KMggplot2::theme_simple(base_size = 14, base_family = "sans")
print(.dea)
rm(.df, .nbins)

# Use the functions by Simm and Besstremyannaya (2020)
library(rDEA)
Bootstrap.rDEA = dea.robust(X, Y, W=NULL, model="output", RTS="constant", B
     =2000, alpha=0.05, bw="bw.ucv")

################## Aggregate ##################

# Generate artificial matrix of prices
p <- as.matrix(4)

# Calculate weight
weight = p%*%Yt/sum(p%*%rowSums(Yt))

# Aggregate
aggregate = sum(dea.crs.out.bench%*%t(weight))
aggregate
mean(dea.crs.out.bench)

################## SW07 ##################

attach(data)

# Use the functions by Simm and Besstremyannaya (2020)
library(rDEA)
```

```
# Define environmental variables
Z = as.matrix(cbind(TEACH, Small, Remote))

# Output-oriented & CRS
sw07 = dea.env.robust(X, Y, W=NULL, Z, "output", RTS="constant", L1=100, L2
    =2000, alpha=0.05)
sw07$beta_hat_hat
sw07$beta_ci
```

References

Adler, N. and B. Golany (2001). "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe". *European Journal of Operational Research*. 132(2): 260–273.

Adler, N. and E. Yazhemsky (2010). "Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction". *European Journal of Operational Research*. 202(1): 273–284.

Afriat, S. N. (1972). "Efficiency estimation of production functions". *International Economic Review*. 13(3): 568–598.

Agrell, P. J. and P. Bogetoft (2005). "Economic and environmental efficiency of district heating plants". *Energy Policy*. 33(10): 1351–1362.

Aigner, D., C. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". *Journal of Econometrics*. 6(1): 21–37.

Alam, I. M. S. (2001). "A nonparametric approach for assessing productivity dynamics of large US banks". *Journal of Money, Credit and Banking*. 33(1): 121–139.

Aragón, A. M. J., A. Castelli, M. Chalkley, and J. Gaughan (2019). "Can productivity growth measures identify best performing hospitals? Evidence from the English National Health Service". *Health Economics*. 28(3): 364–372.

Athanassopoulos, A. D., N. Lambroukos, and L. Seiford (1999). "Data envelopment scenario analysis for setting targets to electricity generating plants". *European Journal of Operational Research*. 115(3): 413–428.

Atici, K. B. and V. V. Podinovski (2015). "Using data envelopment analysis for the assessment of technical efficiency of units with different specialisations: An application to agriculture". *Omega*. 54: 72–83.

Bădin, L., C. Daraio, and L. Simar (2014). "Explaining inefficiency in nonparametric production models: The state of the art". *Annals of Operations Research*. 214(1): 5–30.

Badunenko, O. and P. Mozharovskyi (2016). "Nonparametric frontier analysis using Stata". *The Stata Journal*. 16(3): 550–589.

Badunenko, O. and H. Tauchmann (2019). "Simar and Wilson two-stage efficiency analysis for Stata". *The Stata Journal*. 19(4): 950–988.

Banker, R. D. (1993). "Maximum likelihood, consistency and data envelopment analysis: A statistical foundation". *Management Science*. 39(10): 1265–1273.

Banker, R. D., A. Charnes, and W. W. Cooper (1984). "Some models for estimating technical and scale inefficiencies in data envelopment analysis". *Management Science*. 30(9): 1078–1092.

Banker, R. D., R. F. Conrad, and R. P. Strauss (1986). "A comparative application of data envelopment analysis and translog methods: An illustrative study of hospital production". *Management Science*. 32(1): 30–44.

Banker, R. D. and R. M. Thrall (1992). "Estimation of returns to scale using data envelopment analysis". *European Journal of Operational Research*. 62(1): 74–84.

Berkelaar, M. *et al.* (2023). *lpSolve: Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs*. R package version 5.6.17, URL: https://CRAN.R-project.org/package=lpSolve.

Bhattacharyya, A. (2020). "Production economics in the telecommunications industry". In: *Handbook of Production Economics*. Ed. by S. C. Ray, R. Chambers, and S. Kumbhakar. Singapore: Springer. 1699–1749.

Bjurek, H. (1996). "The Malmquist total factor productivity index". English. *The Scandinavian Journal of Economics*. 98(2): 303–313.

Bogetoft, P. and L. Otto (2022). *Benchmarking with DEA and SFA*. R package version 0.31, URL: https://cran.r-project.org/web/packages/Benchmarking/index.html.

Byrnes, P., R. Färe, S. Grosskopf, and C. Knox Lovell (1988). "The effect of unions on productivity: US surface mining of coal". *Management Science*. 34(9): 1037–1053.

Casu, B., A. Ferrari, and T. Zhao (2013). "Regulatory reform and productivity change in Indian banking". *The Review of Economics and Statistics*. 95(3): 1066–1077.

Caves, D. W., L. R. Christensen, and W. E. Diewert (1982). "The economic theory of index numbers and the measurement of input, output, and productivity". *Econometrica*. 50(6): 1393–1414.

Chambers, R. G., Y. Chung, and R. Färe (1996). "Benefit and distance functions". *Journal of Economic Theory*. 70(2): 407–419.

Chambers, R., Y. Chung, and R. Färe (1998). "Profit, directional distance functions, and Nerlovian efficiency". English. *Journal of Optimization Theory and Applications*. 98(2): 351–364.

Charnes, A., W. W. Cooper, Z. M. Huang, and D. B. Sun (1990). "Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks". *Journal of Econometrics*. 46(1–2): 73–91.

Charnes, A., W. Cooper, and E. Rhodes (1978). "Measuring the efficiency of decision making units". *European Journal of Operational Research*. 2(6): 429–444.

Chen, Y., M. G. Tsionas, and V. Zelenyuk (2021). "LASSO + DEA for small and big wide data". *Omega*. 102: 102419.

Chilingerian, J. A. (1995). "Evaluating physician efficiency in hospitals: A multivariate analysis of best practices". *European Journal of Operational Research*. 80(3): 548–574.

Chirikos, T. N. and A. M. Sear (1994). "Technical efficiency and the competitive behavior of hospitals". *Socio-Economic Planning Sciences*. 28(4): 219–227.

Chua, C. L., A. Palangkaraya, and J. Yong (2010). “A two-stage estimation of hospital quality using mortality outcome measures: An application using hospital administrative data”. *Health Economics*. 19(12): 1404–1424.

Coelli, T. (1996). “A guide to DEAP version 2.1: A data envelopment analysis (computer) program”. *Tech. rep.* No. 08. Centre for Efficiency and Productivity Analysis, University of New England, Australia. 1–49.

Coll-Serrano, V., V. Bolos, and R. B. Suarez (2022). *deaR: Conventional and fuzzy data envelopment analysis*. R package version 1.3.3, URL: https://CRAN.R-project.org/package=deaR.

Cooper, W. W., L. M. Seiford, and K. Tone (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-solver Software*. Vol. 2. New York, NY: Springer.

Dakpo, K. H., P. Jeanneaux, and L. Latruffe (2017). “Modelling pollution-generating technologies in performance benchmarking: Recent developments, limits and future prospects in the nonparametric framework”. *European Journal of Operational Research*. 250(2): 347–359.

Daraio, C. and L. Simar (2005). “Introducing environmental variables in nonparametric frontier models: A probabilistic approach”. *Journal of productivity analysis*. 24(1): 93–121.

Daraio, C. and L. Simar (2007a). *Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications*. New York, NY: Springer.

Daraio, C. and L. Simar (2007b). “Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach”. *Journal of Productivity Analysis*. 28(1): 13–32.

Daraio, C. and L. Simar (2007c). “Economies of scale, scope and experience in the Italian motor-vehicle sector”. In: *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Ed. by C. Daraio and L. Simar. New York, NY: Springer. Chap. 6. 135–165.

Debreu, G. (1951). “The coefficient of resource utilization”. *Econometrica*. 19(3): 273–292.

Deprins, D., L. Simar, and H. Tulkens (1984). "Measuring labour efficiency in post offices". In: *The Performance of Public Enterprises: Concepts and Measurement*. Ed. by M. Marchand, P. Pestieau, and H. Tulkens. Amsterdam, NL: Springer. 243–267.

Diewert, W. E. (1983). "The Measurement of Waste within the Production Sector of an Open Economy". *The Scandinavian Journal of Economics*. 85(2): 159–179.

Diewert, W. E. (1980). "Capital and the theory of productivity measurement". *The American Economic Review*. 70(2): 260–267.

Diewert, W. E. (1992). "Fisher ideal output, input, and productivity indexes revisited". *Journal of Productivity Analysis*. 3(3): 211–248.

Dorfman, R., P. A. Samuelson, and R. M. Solow (1987). *Linear programming and economic analysis*. Courier Corporation.

Du, K., A. C. Worthington, and V. Zelenyuk (2018). "Data envelopment analysis, truncated regression and double-bootstrap for panel data with application to Chinese banking". *European Journal of Operational Research*. 265(2): 748–764.

Dyson, R. G. and E. Thanassoulis (1988). "Reducing weight flexibility in data envelopment analysis". *Journal of the Operational Research Society*. 39(6): 563–576.

Emrouznejad, A., B. Parker, and G. Tavares (2008). "Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA". *Journal of Socio-Economics Planning Science*. 42(3): 151–157.

Emrouznejad, A. and E. Thanassoulis (2014). "Introduction to performance improvement management software (PIM-DEA)". In: *Handbook of research on strategic performance management and measurement using data envelopment analysis*. IGI Global. Chap. 5. 256–275.

Emrouznejad, A. and G.-l. Yang (2018). "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016". *Socio-Eeconomic Planning Sciences*. 61: 4–8.

English, M., S. Grosskopf, K. Hayes, and S. Yaisawarng (1993). "Output allocative and technical efficiency of banks". *Journal of Banking & Finance*. 17(2-3): 349–366.

Färe, R. and S. Grosskopf (1996). *Intertemporal Production Frontiers: With Dynamic DEA*. Norwell, MA: Kluwer Academic Publishers.

Färe, R. and S. Grosskopf (2000). "Network DEA". *Socio-Economic Planning Sciences*. 34: 35–49.

Färe, R. and S. Grosskopf (2004). "Modeling Undesirable Factors in Efficiency Evaluation: Comment". *European Journal of Operational Research*. 157(1): 242–245.

Färe, R. (1991). "Measuring Farrell efficiency for a firm with intermediate inputs". *Academia Economic Papers*. 19(2): 329–340.

Färe, R., H. Fukuyama, S. Grosskopf, and V. Zelenyuk (2015). "Decomposing Profit Efficiency using a Slack-based Directional Distance Function". *European Journal of Operational Research*. 247(1): 335–337.

Färe, R., R. Grabowski, S. Grosskopf, and S. Kraft (1997). "Efficiency of a fixed but allocatable input: A non-parametric approach". *Economics Letters*. 56(2): 187–193.

Färe, R. and S. Grosskopf (1985). "A Nonparametric Cost Approach to Scale Efficiency". *The Scandinavian Journal of Economics*. 87(4): 594–604.

Färe, R. and S. Grosskopf (1997). "Efficiency and productivity in rich and poor countries". In: *Dynamics, economic growth, and international trade*. Ed. by B. Jensen and K. Wong. University of Michigan Press, Studies in International Economics. Ann Arbor. 243–63.

Färe, R. and S. Grosskopf (2003). "Nonparametric Productivity Analysis with Undesirable Outputs: Comment". *American Journal of Agricultural Economics*. 85(4): 1070–1074.

Färe, R. and S. Grosskopf (2009). "A Comment on Weak Disposability in Nonparametric Production Analysis". *American Journal of Agricultural Economics*. 91(2): 535–538.

Färe, R. and S. Grosskopf (2010). "Directional distance functions and slacks-based measures of efficiency". *European Journal of Operational Research*. 200(1): 320–322.

Färe, R., S. Grosskopf, B. Lindgren, and P. Roos (1992). "Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach". *Journal of Productivity Analysis*. 3(1-2): 85–101.

Färe, R., S. Grosskopf, and J. Logan (1983). "The relative efficiency of Illinois electric utilities". *Resources and Energy*. 5(4): 349–367.

Färe, R., S. Grosskopf, and J. Logan (1985a). "The relative performance of publicly-owned and privately-owned electric utilities". *Journal of Public Economics*. 26(1): 89–106.

Färe, R., S. Grosskopf, and C. A. K. Lovell (1985b). *The Measurement of Efficiency of Production*. Boston, MA: Kluwer-Nijhoff Publishing.

Färe, R., S. Grosskopf, and V. Zelenyuk (2007). "Finding Common Ground: Efficiency Indices". In: *Aggregation, Efficiency and Measurement*. Ed. by R. Färe, S. Grosskopf, and D. Primont. Boston, MA: Springer. 83–95.

Färe, R., X. He, S. Li, and V. Zelenyuk (2019). "A unifying framework for farrell profit efficiency measurement". *Operations Research*. 67(1):183-197.

Färe, R. and D. Primont (1995). *Multi-Output Production and Duality: Theory and Applications*. New York, NY: Kluwer Academic Publishers.

Färe, R. and G. Whittaker (1995). "An intermediate input model of dairy production using complex survey data". *Journal of Agricultural Economics*. 46(2): 201–213.

Färe, R. and V. Zelenyuk (2003). "On aggregate Farrell efficiencies". *European Journal of Operational Research*. 146(3): 615–620.

Färe, R. and V. Zelenyuk (2021). "On aggregation of multi-factor productivity indexes". *Journal of Productivity Analysis*. 55(2): 107–133.

Farrell, M. J. (1957). "The measurement of productive efficiency". *Journal of the Royal Statistical Society: Series A (General)*. 120(3): 253–281.

Førsund, F. R. and L. Hjalmarsson (2004). "Calculating scale elasticity in DEA models". *Journal of the Operational Research Society*. 55(10): 1023–1038.

Friesner, D., R. Mittelhammer, and R. Rosenman (2013). "Inferring the incidence of industry inefficiency from DEA estimates". *European Journal of Operational Research*. 224(2): 414–424.

Gass, S. I. (2003). *Linear programming: methods and applications*. Courier Corporation.

Goto, M. and M. Tsutsui (1998). "Comparison of productive and cost efficiencies among Japanese and US electric utilities". *Omega*. 26(2): 177–194.

Greene, W. H. (1980). "Maximum likelihood estimation of econometric frontier functions". *Journal of Econometrics*. 13(1): 27–56.

Greene, W. H. (1990). "A Gamma-distributed stochastic frontier model". *Journal of Econometrics*. 46(1): 141–163.

Grosskopf, S. and V. Valdmanis (1987). "Measuring hospital performance: A non-parametric approach". *Journal of Health Economics*. 6(2): 89–107.

Henderson, D. J. and R. R. Russell (2005). "Human Capital and Convergence: A Production-Frontier Approach". *International Economic Review*. 46(4): 1167–1205.

Henriques, I. C., V. A. Sobreiro, H. Kimura, and E. B. Mariano (2020). "Two-stage DEA in banks: Terminological controversies and future directions". *Expert Systems with Applications*. 161: 113632.

Hollingsworth, B. (2008). "The measurement of efficiency and productivity of health care delivery". *Health Economics*. 17(10): 1107–1128.

Jeong, S.-O. and L. Simar (2006). "Linearly interpolated FDH efficiency score for nonconvex frontiers". *Journal of Multivariate Analysis*. 97(10): 2141–2161.

Ji, Y. and C. Lee (2010). "Data envelopment analysis in Stata". *Stata Journal*. 10(2): 267–280.

Kao, C. (2014). "Network data envelopment analysis: A review". *European Journal of Operational Research*. 239(1): 1–16.

Karlsson, M. and A. Lindmark (2014). "truncSP: An R package for estimation of semi-parametric truncated linear regression models". *Journal of Statistical Software*. 57(14): 1–19.

Kazemi Matin, R. and T. Kuosmanen (2009). "Theory of integer-valued data envelopment analysis under alternative returns to scale axioms". *Omega*. 37(5): 988–995.

Kneip, A., B. U. Park, and L. Simar (1998). "A note on the convergence of nonparametric DEA estimators for production efficiency scores". *Econometric Theory*. 14(6): 783–793.

Kneip, A., L. Simar, and P. W. Wilson (2008). "Asymptotics and Consistent Bootstraps for DEA Estimators in Nonparametric Frontier Models". *Econometric Theory*. 24(6): 1663–1697.

Kneip, A., L. Simar, and P. W. Wilson (2015). "When bias kills the variance: Central Limit Theorems for DEA and FDH efficiency scores". *Econometric Theory*. 31(2): 394.

Kneip, A., L. Simar, and P. W. Wilson (2016). "Testing hypotheses in nonparametric models of production". *Journal of Business & Economic Statistics*. 34(3): 435–456.

Kneip, A., L. Simar, and P. W. Wilson (2021). "Inference in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices". *Econometric Theory*. 37(3): 537–572.

Kneip, A., L. Simar, and P. W. Wilson (2011). "A Computationally Efficient, Consistent Bootstrap for Inference with Non-parametric DEA Estimators". English. *Computational Economics*. 38(4): 483–515.

Kohl, S., J. Schoenfelder, A. Fügener, and J. O. Brunner (2019). "The use of Data Envelopment Analysis (DEA) in healthcare with a focus on hospitals". *Health Care Management Science*. 22(2): 245–286.

Koopmans, T. (1951). *Activity Analysis of Production and Allocation*. New York, NY: Wiley.

Kooreman, P. (1994). "Nursing home care in The Netherlands: a nonparametric efficiency analysis". *Journal of Health Economics*. 13(3): 301–316.

Korostelëv, A. P., L. Simar, and A. B. Tsybakov (1995). "Efficient estimation of monotone boundaries". *The Annals of Statistics*: 476–489.

Krivonozhko, V. E., F. R. Førsund, and A. V. Lychev (2014). "Measurement of returns to scale using non-radial DEA models". *European Journal of Operational Research*. 232(3): 664–670.

Kumar, S. and R. R. Russell (2002). "Technological Change, Technological Catch-Up, and Capital Deepening: Relative Contributions to Growth and Convergence". *American Economic Review*. 92(3): 527–548.

Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2007). “Nonparametric stochastic frontiers: A local maximum likelihood approach”. *Journal of Econometrics*. 137(1): 1–27.

Kumbhakar, S. C., C. F. Parmeter, and V. Zelenyuk (2022a). “Stochastic frontier analysis: Foundations and advances I”. In: *Handbook of production economics*. Ed. by S. C. Ray, R. G. Chambers, and S. C. Kumbhakar. Singapore: Singapore: Springer. Chap. 8. 331–370.

Kumbhakar, S. C., C. F. Parmeter, and V. Zelenyuk (2022b). “Stochastic frontier analysis: Foundations and advances II”. In: *Handbook of production economics*. Ed. by S. C. Ray, R. G. Chambers, and S. C. Kumbhakar. Singapore: Singapore: Springer. Chap. 9. 371–408.

Kuosmanen, T. (2005). “Weak Disposability in Non-parametric Production Analysis with Undesirable Outputs”. *American Journal of Agricultural Economics*. 87(4): 1077–1082.

Kuosmanen, T. (2008). “Representation theorem for convex nonparametric least squares”. *The Econometrics Journal*. 11(2): 308–325.

Kuosmanen, T. and A. L. Johnson (2010). “Data envelopment analysis as nonparametric least-squares regression”. *Operations Research*. 58(1): 149–160.

Kuosmanen, T. and R. Kazemi Matin (2009). “Theory of integer-valued data envelopment analysis”. *European Journal of Operational Research*. 192(2): 658–667.

Kuosmanen, T. and M. Kortelainen (2012). “Stochastic Non-smooth Envelopment of Data: Semi-parametric Frontier Estimation Subject to Shape Constraints”. *Journal of Productivity Analysis*. 38(1): 11–28.

Lee, C.-Y. and J.-Y. Cai (2020). “LASSO variable selection in data envelopment analysis with small datasets”. *Omega*. 91: 102019.

Li, Q. (1996). “Nonparametric testing of closeness between two unknown distribution functions”. *Econometric Reviews*. 15(3): 261–274.

Malmquist, S. (1953). “Index numbers and indifference surfaces”. *Trabajos de Estadistica*. 4(2): 209–242.

Mayer, A. and V. Zelenyuk (2014). “Aggregation of Malmquist productivity indexes allowing for reallocation of resources”. *European Journal of Operational Research*. 238(3): 774–785.

Meeusen, W. and J. van den Broeck (1977). “Efficiency estimation from Cobb-Douglas production functions with composed error”. *International Economic Review*. 18(2): 435–444.

Miller, S. M. (2020). “Empirical Analysis of Production Economics: Applications to Banking”. In: *Handbook of Production Economics*. Ed. by S. C. Ray, R. Chambers, and S. Kumbhakar. Singapore: Springer. 1165–1191.

Mitropoulos, P., M. A. Talias, and I. Mitropoulos (2015). “Combining stochastic DEA with Bayesian analysis to obtain statistical properties of the efficiency scores: An application to Greek public hospitals”. *European Journal of Operational Research*. 243(1): 302–311.

Mugera, A. W., M. R. Langemeier, and A. Ojede (2016). “Contributions of productivity and relative price changes to farm-level profitability change”. *American Journal of Agricultural Economics*. 98(4): 1210–1229.

Nguyen, B. H., R. C. Sickles, and V. Zelenyuk (2022a). “Efficiency Analysis with stochastic frontier models using popular statistical softwares”. In: *Advances in Economic Measurement*. A Volume in Honour of D.S. Prasada Rao. Ed. by D. Chotikapanich, A. Rambaldi, and N. Rohde. Singapore: Palgrave-Macmillan. 129–171.

Nguyen, B. H., L. Simar, and V. Zelenyuk (2022b). “Data sharpening for improving central limit theorem approximations for data envelopment analysis–type efficiency estimators”. *European Journal of Operational Research*.

Nguyen, B. H. and V. Zelenyuk (2021a). “Aggregate Efficiency of Industry and its Groups: The case of Queensland Public Hospitals”. *Empirical Economics*. 60(6): 2795–2836.

Nguyen, B. H. and V. Zelenyuk (2021b). “Robust efficiency analysis of public hospitals in Queensland, Australia”. In: *Advances in Contemporary Statistics and Econometrics*. Ed. by A. Daouia and A. Ruiz-Gazen. Springer. 221–242.

Nunamaker, T. R. (1983). “Measuring routine nursing service efficiency: A comparison of cost per patient day and Data Envelopment Analysis models.” *Health Services Research*. 18(2 Pt 1): 183.

O'Donnell, C. J. (2012). "Nonparametric Estimates of the Components of Productivity and Profitability Change in U.S. Agriculture". English. *American Journal of Agricultural Economics*. 94(4): 873–890.

Olson, J. A., P. Schmidt, and D. M. Waldman (1980). "A Monte Carlo study of estimators of the stochastic frontier production function". *Journal of Econometrics*. 13(1): 67–82.

O'Neill, L., M. Rauner, K. Heidenberger, and M. Kraus (2008). "A cross-national comparison and taxonomy of DEA-based hospital efficiency studies". *Socio-Economic Planning Sciences*. 42(3): 158–189.

Park, B. U., L. Simar, and C. Weiner (2000). "The FDH Estimator for Productivity Efficiency Scores: Asymptotic Properties". *Econometric Theory*. 16(6): 855–877.

Park, B. U., S.-O. Jeong, L. Simar, *et al.* (2010). "Asymptotic Distribution of Conical-Hull Estimators of Directional Edges". English. *Annals of Statistics*. 38(3): 1320–1340.

Parmeter, C. F. and V. Zelenyuk (2019). "Combining the virtues of stochastic frontier and data envelopment analysis". *Operations Research*. 67(6): 1628–1658.

Pham, M. D., L. Simar, and V. Zelenyuk (2023). "Statistical inference for aggregation of Malmquist productivity indices". *Operations Research*. 72(2): 1615–1629.

Pham, M. D. and V. Zelenyuk (2019). "Weak disposability in nonparametric production analysis: A new taxonomy of reference technology sets". *European Journal of Operational Research*. 274(1): 186–198.

Podinovski, V. V. (2004). "Production trade-offs and weight restrictions in data envelopment analysis". *Journal of the Operational Research Society*. 55(12): 1311–1322.

Podinovski, V. V. (2015). "DEA models with production trade-offs and weight restrictions". In: *Data Envelopment Analysis: A Handbook of Models and Methods*. Ed. by J. Zhu. New York, NY: Springer. 105–144.

Podinovski, V. V. and T. Bouzdine-Chameeva (2013). "Weight restrictions and free production in data envelopment analysis". *Operations Research*. 61(2): 426–437.

Podinovski, V. V., F. R. Førsund, and V. E. Krivonozhko (2009). "A simple derivation of scale elasticity in data envelopment analysis". *European Journal of Operational Research*. 197(1): 149–153.

Podinovski, V. V. and T. Kuosmanen (2011). "Modelling weak disposability in data envelopment analysis under relaxed convexity assumptions". *European Journal of Operational Research*. 211(3): 577–585.

Puig-Junoy, J. (1998). "Technical efficiency in the clinical management of critically ill patients". *Health Economics*. 7(3): 263–277.

R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rączka, J. (2001). "Explaining the performance of heat plants in Poland". *Energy Economics*. 23(4): 355–370.

Ray, S. C. (1991). "Resource-use efficiency in public schools: A study of Connecticut data". *Management Science*. 37(12): 1620–1628.

Ray, S. C. (2020). "Choice of Inputs and Outputs for Production Analysis". In: *Handbook of Production Economics*. Ed. by S. C. Ray, R. Chambers, and S. Kumbhakar. Singapore: Springer. 1083–1116.

Reinhard, S., C. K. Lovell, and G. J. Thijssen (2000). "Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA". *European Journal of Operational Research*. 121(2): 287–303.

Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function". *Annals of Mathematical Statistics*. 27(3): 832–837.

Rosko, M. D. (2001). "Cost efficiency of US hospitals: A stochastic frontier approach". *Health Economics*. 10(6): 539–551.

Russell, R. R. (1990). "Continuity of measures of technical efficiency". *Journal of Economic Theory*. 51(2): 255–267.

Sathye, M. (2003). "Efficiency of banks in a developing economy: The case of India". *European Journal of Operational Research*. 148(3): 662–671.

See, K., S. Grosskopf, V. Valdmanis, and V. Zelenyuk (2024). "What do we know from the vast literature on efficiency and productivity in healthcare? A review and bibliometric analysis". In: *The Cambridge Handbook of Productivity, Efficiency and Effectiveness in Health Care*. Ed. by S. Grosskopf, V. Valdmanis, and V. Zelenyuk. Cambridge, UK: Cambridge University Press.

Seiford, L. M. and J. Zhu (1999). "Profitability and marketability of the top 55 US commercial banks". *Management Science*. 45(9): 1270–1288.

Sheather, S. J. and M. C. Jones (1991). "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation". *Journal of the Royal Statistical Society: Series B (Methodological)*. 53(3): 683–690.

Shephard, R. W. (1974). "Indirect Production Functions". In: *Mathematical Systems in Economics*. Vol. 10. Hain, Meisenheim am Glan.

Shephard, R. W., D. Gale, and H. W. Kuhn (1970). *Theory of Cost and Production Functions. Princeton studies in mathematical economics*. No. 4. Princeton, NJ: Princeton University Press.

Shephard, R. (1953). *Cost and production functions*. Princeton, NJ: Princeton University Press.

Sherman, H. D. and F. Gold (1985). "Bank branch operating efficiency: Evaluation with Data Envelopment Analysis". *Journal of Banking and Finance*. 9(2): 297–315.

Sickles, R. C., W. Song, and V. Zelenyuk (2020). "Econometric analysis of productivity: Theory and implementation in R". In: *Financial, Macro and Micro Econometrics Using R*. Ed. by H. D. Vinod and C. Rao. Vol. 42. *Handbook of Statistics*. Elsevier. 267–297.

Sickles, R. C., Z. Wang, and V. Zelenyuk (2024). "Stochastic frontier analysis for healthcare, with illustrations in R". In: *The Cambridge Handbook of Productivity, Efficiency and Effectiveness in Health Care*. Ed. by S. Grosskopf, V. Valdmanis, and V. Zelenyuk. Cambridge, UK: Cambridge University Press.

Sickles, R. C. and V. Zelenyuk (2019). *Measurement of Productivity and Efficiency: Theory and Practice*. New York, NY: Cambridge University Press.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Vol. 26. London, UK: Chapman and Hall.

Simar, L., I. Van Keilegom, and V. Zelenyuk (2017). "Nonparametric Least Squares Methods for Stochastic Frontier Models". *Journal of Productivity Analysis*. 47(3): 189–204.

Simar, L. and P. W. Wilson (2013). "Estimation and inference in nonparametric frontier models: Recent developments and perspectives". *Foundations and Trends in Econometrics*. 5(3–4): 183–337.

Simar, L. (2007). "How to improve the performances of DEA/FDH estimators in the presence of noise?" *Journal of Productivity Analysis*. 28(3): 183–201.

Simar, L. and P. W. Wilson (1998). "Sensitivity analysis to efficiency scores: How to bootstrap in Nonparametric frontier models". *Management Science*. 44(1): 49–61.

Simar, L. and P. W. Wilson (2007). "Estimation and inference in two-stage, semi-parametric models of production processes". *Journal of Econometrics*. 136(1): 31–64.

Simar, L. and P. W. Wilson (2002). "Non-parametric tests of returns to scale". *European Journal of Operational Research*. 139(1): 115–132.

Simar, L. and P. W. Wilson (2015). "Statistical approaches for nonparametric frontier models: A guided tour". *International Statistical Review*. 83(1): 77–110.

Simar, L. and P. W. Wilson (2019). "Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices". *European Journal of Operational Research*. 277(2): 756–769.

Simar, L. and P. W. Wilson (2020). "Technical, allocative and overall efficiency: Estimation and inference". *European Journal of Operational Research*. 282(3): 1164–1176.

Simar, L. and V. Zelenyuk (2020). "Improving finite sample approximation by central limit theorems for estimates from Data Envelopment Analysis". *European Journal of Operational Research*. 284(3): 1002–1015.

Simar, L. and V. Zelenyuk (2006). "On testing equality of distributions of technical efficiency scores". *Econometric Reviews*. 25(4): 497–522.

Simar, L. and V. Zelenyuk (2007). “Statistical inference for aggregates of Farrell-type efficiencies”. *Journal of Applied Econometrics*. 22(7): 1367–1394.

Simar, L. and V. Zelenyuk (2011). “Stochastic FDH/DEA estimators for frontier analysis”. *Journal of Productivity Analysis*. 36(1): 1–20.

Simar, L. and V. Zelenyuk (2018). “Central limit theorems for aggregate efficiency”. *Operations Research*. 66(1): 137–149.

Simar, L., V. Zelenyuk, and S. Zhao (2023). “Further improvements of finite sample approximation of central limit theorems for envelopment estimators”. *Journal of Productivity Analysis*. 59(2): 189–194.

Simar, L., V. Zelenyuk, and S. Zhao (2024a). “Inference for aggregate efficiency: Theory and guidelines for practitioners”. *European Journal of Operational Research*. 316(1): 240–254.

Simar, L., V. Zelenyuk, and S. Zhao (2024b). “Statistical inference for Hicks–Moorsteen productivity indices”. *Annals of Operations Research*. Forthcoming, DOI: 10.1007/s10479-024-06288-8.

Simm, J. and G. Besstremyannaya (2020). *rDEA: Robust data envelopment analysis (DEA) for R*. R package version 1.2-6, URL: https://CRAN.R-project.org/package=rDEA.

Stanton, K. R. (2002). “Trends in relationship lending and factors affecting relationship lending efficiency”. *Journal of Banking & Finance*. 26(1): 127–152.

Stevenson, R. E. (1980). “Likelihood Functions for Generalized Stochastic Frontier Estimation”. *Journal of Econometrics*. 13(1): 57–66.

Sueyoshi, T. and M. Goto (2013). “A comparative study among fossil fuel power plants in PJM and California ISO by DEA environmental assessment”. *Energy Economics*. 40: 130–145.

Sueyoshi, T., Y. Yuan, and M. Goto (2017). “A literature study for DEA applied to energy and environment”. *Energy Economics*. 62: 104–124.

Thompson, R. G., L. N. Langemeier, C.-T. Lee, E. Lee, and R. M. Thrall (1990). “The role of multiplier bounds in efficiency analysis with application to Kansas farming”. *Journal of Econometrics*. 46(1-2): 93–108.

Tibshirani, R. (1996). “Regression Shrinkage and Selection Via the Lasso”. *Journal of the Royal Statistical Society: Series B (Methodological)*. 58(1): 267–288.

Tone, K. and M. Tsutsui (2010). “Erratum to "Network DEA: A slacks-based measure approach"”. *European Journal of Operational Research*. 202(1): 308–309.

Tone, K. and M. Tsutsui (2014). “Dynamic DEA with network structure: A slacks-based measure approach”. *Omega*. 42(1): 124–131.

Tsionas, M. G. (2020). “A coherent approach to Bayesian data envelopment analysis”. *European Journal of Operational Research*. 281(2): 439–448.

Tsionas, M., C. F. Parmeter, and V. Zelenyuk (2023). “Bayesian artificial neural networks for frontier efficiency analysis”. *Journal of Econometrics*. 236(2): 105491.

Vanderbei, R. J. *et al.* (2020). *Linear programming*. Springer.

von Neumann, J. (1938). “On infinite tensor products”. *Composite Mathematics*. 6: 1–77.

Wang, Z., B. H. Nguyen, and V. Zelenyuk (2024). “Performance analysis of hospitals in Australia and its peers: A systematic and critical review”. *Journal of Productivity Analysis*. 62(2): 139–173.

Wang, Z. and V. Zelenyuk (2024a). “Overview of performance analytics for healthcare with examples in R”. In: *The Cambridge Handbook of Productivity, Efficiency and Effectiveness in Health Care*. Ed. by S. Grosskopf, V. Valdmanis, and V. Zelenyuk. Cambridge, UK: Cambridge University Press.

Wang, Z. and V. Zelenyuk (2024b). “Random versus explained inefficiency in stochastic frontier analysis: The case of Queensland hospitals”. In: *Advances in Econometrics: Essays in honor of Subal Kumbhakar*. Ed. by C. F. Parmeter, M. Tsionas, and H.-J. Wang. Vol. 46. Emerald Publishing Limited. 371–413.

Wheat, P., K. Odolinski, and A. Smith (2020). “Applications of Production Theory in Transportation”. In: *Handbook of Production Economics*. Ed. by S. C. Ray, R. Chambers, and S. Kumbhakar. Singapore: Springer. 1491–1524.

Wilson, P. W. (2018). "Dimension reduction in nonparametric models of production". *European Journal of Operational Research*. 267(1): 349–367.

Wilson, P. W. (2020). *FEAR: Frontier efficiency analysis with R*. R package version 3.1, URL: https://pww.people.clemson.edu/Software/FEAR/fear.html.

Wu, J., M. Li, Q. Zhu, Z. Zhou, and L. Liang (2019). "Energy and environmental efficiency measurement of China's industrial sectors: A DEA model with non-homogeneous inputs and outputs". *Energy Economics*. 78: 468–480.

Zelenyuk, N. and V. Zelenyuk (2021). "Bank performance analysis". In: *Handbook of Research Methods and Applications in Empirical Microeconomics*. Edward Elgar Publishing. 238–279.

Zelenyuk, V. (2006). "Aggregation of Malmquist productivity indexes". *European Journal of Operational Research*. 174(2): 1076–1086.

Zelenyuk, V. (2013). "A scale elasticity measure for directional distance function and its dual: Theory and DEA estimation". *European Journal of Operational Research*. 228(3): 592–600.

Zelenyuk, V. (2014). "Scale efficiency and homotheticity: Equivalence of primal and dual measures". *Journal of Productivity Analysis*. 42(1): 15–24.

Zelenyuk, V. (2020). "Aggregation of inputs and outputs prior to Data Envelopment Analysis under big data". *European Journal of Operational Research*. 282(1): 172–187.

Zelenyuk, V. (2021). "Performance analysis: Economic foundations and trends". *Foundations and Trends® in Econometrics*. 11(3): 153–229.

Zelenyuk, V. and S. Zhao (2024). "Russell and slack-based measures of efficiency: A unifying framework". *European Journal of Operational Research*. 318(3): 867–876.

Zhang, N. and Y. Choi (2014). "A note on the evolution of directional distance function and its development in energy and environmental studies 1997–2013". *Renewable and Sustainable Energy Reviews*. 33: 50–59.

Zhou, P., B. W. Ang, and K. L. Poh (2008). "A survey of data envelopment analysis in energy and environmental studies". *European Journal of Operational Research*. 189(1): 1–18.

Zhu, J. (1998). "Data envelopment analysis vs. principal component analysis: An illustrative study of economic performance of Chinese cities". *European Journal of Operational Research*. 111(1): 50–61.

Zhu, J. (2009). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Vol. 2. Springer.

Zou, H. and T. Hastie (2005). "Regularization and variable selection via the elastic net". *Journal of the rRoyal Statistical Society: Series B (Statistical Methodology)*. 67(2): 301–320.