



A Data Envelopment Analysis Toolbox for MATLAB

Inmaculada C. Álvarez

Universidad Autónoma de Madrid
Oviedo Efficiency Group

Javier Barbero

Universidad Autónoma de Madrid
Oviedo Efficiency Group

José L. Zoffio

Universidad Autónoma de Madrid
Erasmus Research Institute of Management
Erasmus University Rotterdam

Abstract

The **Data Envelopment Analysis Toolbox** is a new package for MATLAB that includes functions to calculate the main data envelopment analysis models. The package includes code for the standard radial input, output and additive measures, allowing for constant and variable returns to scale, as well as recent developments related to the directional distance function, and including both desirable and undesirable outputs when measuring efficiency and productivity; i.e., Malmquist and Malmquist-Luenberger indices. Bootstrapping to perform statistical analysis is also included. This paper describes the methodology and implementation of the functions, and reports numerical results using a reliable productivity database on US agriculture to illustrate their use.

Keywords: data envelopment analysis, distance functions, technical efficiency, MATLAB.

1. Introduction

Data envelopment analysis, DEA, has grown in importance over the past decades due to increase in the availability of data related to the performance of decision making units, DMUs, regardless their market, governmental, or non-for-profit orientation. DEA is a nonparametric method in operations research and economics that allows to practice benchmarking exercises among a set of comparable observations. This comparison is made through optimizing mathematical programming techniques that approximate the production technology and identify the most efficient facets or hyperplanes that, enveloping the data (hence the name), define the production or best-practice frontier, serving as reference for all observations. The meth-

ods yield individual efficiency scores and reference benchmarks for each observation, thereby providing real-life peers that serve as example for managers and decision-makers interested in improving their performance. Because of the flexibility of DEA, researchers in a number of fields have quickly acknowledged these techniques as an excellent methodology for modeling operational processes. The empirical orientation and absence of a priori assumptions have resulted in numerous studies involving best-practice identification in a wide array of sectors – for a comprehensive introduction and a complete list of applications see [Fried, Lovell, and Schmidt \(2008\)](#). Indeed, the popularity of DEA has increased exponentially in these years as recent volumes devoted to specific sectors and industries bear witness; e.g., [Blackburn, Brennan, and Ruggiero \(2014\)](#), [Ozcan \(2014\)](#), [Paradi, Sherman, and Tam \(2018\)](#) and [Khezri-motlagh and Chen \(2018\)](#). The present toolbox allows interested practitioners to implement these studies in their everyday efficiency improvement strategies.

Basic DEA methods are included in some standard software packages used by econometricians (e.g., [Stata, StataCorp 2015](#), with the user-written command by [Ji and Lee 2010](#); [LIMDEP, Econometric Software, Inc 2009](#)); available in dedicated non-commercial software accompanying academic handbooks – [Cooper, Seiford, and Tone \(2007\)](#), [Wilson \(2008\)](#), [Bogetoft and Otto \(2011\)](#) (these latter two implemented in R; [R Core Team 2020](#)); commercial software – including trials versions, [Emrouznejad and Cabanda \(2014\)](#); free-ware programs – [Sheel \(2000\)](#); and even tutorials for spreadsheets: [Sherman and Zhu \(2006\)](#) and [Zhu \(2014\)](#). Earlier versions of these programs have been reviewed, among others, by [Hollingsworth \(2004\)](#) and [Barr \(2004\)](#). More recently, [Daraio, Kerstens, Nepomuceno, and Sickles \(2019\)](#) provide the most recent survey of the available options. Their eligibility criteria include only software that is diffused as self-contained programs, or packages and toolboxes for general computing environments. To complement the above references with general purpose DEA software that implements newer proposals, but excluding very specialized contributions that code specific models normally linked to particular papers, it is worth referencing [Oh and Suh \(2013\)](#) and [Badunenko, Mozharovskiy, and Kolomiytseva \(2020\)](#).

While these software packages implement the main DEA models, there is a lack of a full set of functions for MATLAB ([The MathWorks, Inc. 2017](#)), including some recent theoretical contributions that are missed in the existing software.¹ Thus, besides implementing the most popular DEA models in the MATLAB environment, the toolbox solves new models such as those corresponding to the directional and the weighted additive measures, while allowing for a greater degree of flexibility in the formulations; e.g., users can supply their preferred directional and weights vectors. Also, the models implemented, such as the directional measure with undesirable outputs, or the Malmquist-Luenberger index, correspond to the most recent theoretical proposals that solve relevant inconsistencies. As the most recent contribution to statistical software – mostly in R, the toolbox is up to date with respect to the models it includes, including bootstrapping techniques and hypotheses testing that are almost inexistent in older alternatives.

¹In principle, it would be desirable that our toolbox were compatible with the current distribution of `Octave` (v5.2.0; [Eaton, Bateman, Hauberg, and Wehbring 2020](#)). However, this is not the case with the out-of-the-box distribution. The reason is that the `linprog` function for linear programming in MATLAB is more advanced and flexible (in terms of the optimization options) than the equivalent function provided in `Octave`. In addition, some of the functions used are not available in `Octave` (such as `optimoptions`) and need to be replaced with other functions that do not have the same functionality as in MATLAB. Hence, it is not possible to make the toolbox compatible with `Octave` while maintaining some of its desirable features. However, any reader interested in solving any of our DEA models in `Octave` may adapt the syntax of our functions so they run in this program.

Nevertheless, it is virtually impossible to keep track and implement all current proposals in the literature. We refer the interested reader to the most recent specialized publications in DEA topics. In particular [Zhu \(2015\)](#) includes a series of chapters on current DEA theory dealing with most advanced topics such as stochastic nonparametric frontier approaches, efficiency measurement with fuzzy data, models with production trade-offs and weight restrictions, etc. [Daraio and Simar \(2007\)](#) deal with statistically robust DEA methods, including bootstrapping techniques – partially implemented in R by [Simm and Besstremyannaya \(2020\)](#). Finally [Aparicio, Lovell, and Pastor \(2016\)](#) and [Hwang, Lee, and Zhu \(2016\)](#) include state of the art and recent advances in the field of efficiency and productivity theory. Indeed, issues related to enhancing the discriminatory power of DEA as well as its statistical robustness are still at the forefront of current research in DEA, e.g., [Angulo-Meza and Pereira Estellita Lins \(2002\)](#). However while these topics merit special attention, their implementation exceeds the scope of the present software, which should be seen as the baseline tool for further specific implementations based on the proposed functions.

Arguably, a separate toolbox should be devoted to topics that are relevant by themselves and self-contained, such as weight restrictions and the inclusion of a priori information, or statistical methods such as robust, stochastic, and fuzzy DEA methods; e.g., see the recent reviews on stochastic DEA and fuzzy DEA by [Olesen and Petersen \(2016\)](#) and [Hatami-Marbini, Emrouznejad, and Tavana \(2011\)](#), respectively, and related applications by [Kao and Liu \(2009\)](#) and [Costantino, Dotoli, Epicoco, Falagario, and Sciancalepore \(2012b\)](#).

The **Data Envelopment Analysis Toolbox** introduces a complete set of baseline functions, covering a wide range of efficiency and productivity models, and reporting numerical results based on classical examples presented in the literature. The **Data Envelopment Analysis Toolbox** is available as free software, under the GNU General Public License version 3, and can be downloaded from <http://www.deatoolbox.com/>, with all the supplementary material (data, examples and source code) to replicate all the results presented in this paper. The toolbox is also hosted on an open source repository on GitHub.²

The paper is organized as follows. The following section presents the data structures characterizing the production possibility sets, the structure of the functions, results, etc., and briefly describes the data that is used to illustrate the toolbox. Section 3 covers the standard DEA models introduced by [Charnes, Cooper, and Rhodes \(1978\)](#) and [Banker, Charnes, and Cooper \(1984\)](#) corresponding to the radial input and output efficiency measures, allowing for constant and variable returns to scale, as well as newer proposals based on the flexible directional distance function. The non-oriented additive model is also presented as well as the super-efficiency model for all the previous efficiency measures. The complementary cross efficiency model that also allows to overcome the low discriminatory power of standard DEA is presented in Section 4. Malmquist productivity indices and their decomposition into efficiency change and technical change are shown in Section 5, while Section 6 deals with the measurement of economic efficiency, and its decomposition into technical and allocative factors. Section 7 is devoted to the measurement of efficiency with undesirable outputs, most notably environmental efficiency, followed by Section 8 presenting the Malmquist-Luenberger index. Statistical analyses and hypotheses testing using bootstrapping techniques are discussed in Section 9. Advanced options, including displaying and exporting results can be found in Section 10. Section 11 concludes.

²The address of the repository is <https://github.com/javierbarbero/DEAMATLAB>.

2. Data structures

Data envelopment analysis measures productive and economic performance of a set of $j = 1, 2, \dots, n$ observed DMUs (firms, activities, countries, individuals, etc.). These observations transform a vector of $i = 1, 2, \dots, m$ positive inputs $\mathbf{x} \in \mathbb{R}_{++}^m$ into a vector of $i = 1, 2, \dots, s$ positive outputs $\mathbf{y} \in \mathbb{R}_{++}^s$ using the technology represented by the following constant returns to scale production possibility set: $P_{\text{CRS}} = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\lambda, \mathbf{y} \leq Y\lambda, \lambda \geq \mathbf{0}\}$, where $X = (\mathbf{x})_j \in \mathbb{R}^{s \times n}$, $Y = (\mathbf{y})_j \in \mathbb{R}^{m \times n}$ and $\lambda = (\lambda_1, \dots, \lambda_n)^\top$ is a semipositive vector.

Data are managed as regular MATLAB vectors and matrices, constituting the inputs of the estimation functions. All estimation functions return a structure ‘`deaout`’ that contains fields with the estimation results as well as the input of the estimation function. Fields can be accessed directly using the dot notation and the whole structure can be used as an input to other functions that print or export results (e.g., `deadisp`).

Some of the fields of the ‘`deaout`’ structure are the following:³

- `X`, `Y` and `Yu`: Contain the inputs, outputs and undesirable outputs variables, respectively.
- `n` and `neval`: Number of DMUs, and number of evaluated DMUs.
- `m`, `s` and `r`: Number of inputs, outputs and undesirable outputs.
- `model`, `orient`, `rts`: Strings containing the model type, the orientation, and the returns to scale assumption.
- `eff`: Computed efficiency measure.
- `slackX`, `slackY`, `slackYu`: Computed input, output and undesirable output slacks.
- `names`: Names of the DMUs.

2.1. Dataset and statistical sources

We illustrate all the models presented in this toolbox resorting to a single dataset. The data has been collected and tabulated by the Economic Research Section (ERS) of the United States Department of Agriculture (USDA) in an effort to study long term productivity trends using prices-based indices. It corresponds to a subset of the state-level tables including price indices and implicit quantities of farm outputs and inputs (Table 23), which can be downloaded in full from <https://www.ers.usda.gov/data-products/agricultural-productivity-in-the-us/>.

To illustrate the cross-section efficiency models we focus on the last available year, 2004, while productivity trends corresponding to the Malmquist index are calculated using panel data from 2000 to 2004. Our dataset consists of three outputs (livestock, crops and other farm related output), four inputs (capital, land, labor and intermediate inputs), supplemented with the number of pesticide exposures reported by the Center for Disease Control and Prevention (CDCP): <https://ephtracking.cdc.gov/showPesticidesExposuresLanding>. This variable constituting an undesirable output is necessary to calculate ‘green’ Malmquist-Luenberger indices accounting for externalities.

³For a full list see the help of the function typing `help deaout` in MATLAB.

More information on the data, including a discussion on the agricultural productivity growth in the US can be found in the above link. Due to its comprehensiveness and reliability, this dataset has been used over the years by many authors, whose results are complemented with those obtained by solving the models included in this toolbox – see, e.g., Färe, Grosskopf, and Margaritis (2008), Ball, Färe, Grosskopf, and Nehring (2001) and Zofio and Lovell (2001).

3. Basic DEA models

3.1. Radial input oriented model: Constant and variable returns to scale

Based on the data matrix (X, Y) , we measure the input oriented efficiency of each observation o by solving n times the following linear programming problem – known as the Charnes, Cooper, and Rhodes, CCR, model:⁴

$$\begin{aligned} \min_{\theta, \lambda} \quad & \theta & (1) \\ \text{subject to} \quad & & \\ & \theta \mathbf{x}_o \geq X\lambda & \\ & Y\lambda \geq \mathbf{y}_o & \\ & \lambda \geq \mathbf{0}. & \end{aligned}$$

The optimal solution to this program – characterizing a technology with constant returns to scale, CRS, is denoted by θ_{CRS}^* . The constraints require the observation $(\theta_{\text{CRS}} \mathbf{x}_o, \mathbf{y}_o)$ to belong to P_{CRS} , while the objective seeks the minimum θ_{CRS} that reduces the input vector \mathbf{x}_o radially to $\theta_{\text{CRS}} \mathbf{x}_o$ while remaining in P_{CRS} ; i.e., that projects the observation to the production frontier. The frontier corresponds to the supporting hyperplane defined by the linear combination of the observations that serve as reference benchmark for the evaluated unit. Those observations whose associated λ multipliers are greater than zero define the enveloping hyperplane. A feasible solution signaling radial efficiency is $\theta_{\text{CRS}}^* = 1$. Therefore if $\theta_{\text{CRS}}^* < 1$, the observation is *radially* inefficient and $(\lambda X, \lambda Y)$ outperforms $(\mathbf{x}_o, \mathbf{y}_o)$. With regard to this property, we define the additional input excesses and output shortfalls by the following slack vectors: $\mathbf{s}^- \in \mathbb{R}^m$ and $\mathbf{s}^+ \in \mathbb{R}^s$, respectively. Therefore: $\mathbf{s}^- = \theta_{\text{CRS}}^* \mathbf{x}_o - X\lambda$, and $\mathbf{s}^+ = Y\lambda - \mathbf{y}_o$ with $\mathbf{s}^- \geq \mathbf{0}$ and $\mathbf{s}^+ \geq \mathbf{0}$ for any feasible solution (θ, λ) .

To obtain the possible input excesses and output shortfalls, the following second stage program that incorporates the optimal value θ_{CRS}^* and corrects radial inefficiency is solved:

$$\begin{aligned} \max_{\lambda, \mathbf{s}^-, \mathbf{s}^+} \quad & \omega = \mathbf{e}\mathbf{s}^- + \mathbf{e}\mathbf{s}^+ & (2) \\ \text{subject to} \quad & & \\ & \mathbf{s}^- = \theta_{\text{CRS}}^* \mathbf{x}_o - X\lambda & \\ & \mathbf{s}^+ = Y\lambda - \mathbf{y}_o & \\ & \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, & \end{aligned}$$

where $\mathbf{e} = (1, \dots, 1)^\top$ so $\mathbf{e}\mathbf{s}^- = \sum_{i=1}^m \mathbf{s}_i^-$ and $\mathbf{e}\mathbf{s}^+ = \sum_{i=1}^s \mathbf{s}_i^+$.

⁴This program corresponds to the so-called “envelopment form” of the formulation introduced by Charnes *et al.* (1978). Correspondences with the dual approaches (“multipliers form”) can be found in Cooper *et al.* (2007).

As a result, an observation is *technically* efficient if the optimal solution $(\theta_{\text{CRS}}^*, \lambda^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ of the two above programs satisfies $\theta_{\text{CRS}}^* = 1, \mathbf{s}^{-*} = 0$, and $\mathbf{s}^{+*} = 0$, so no equiproportional contraction of inputs, and individual inputs reductions and outputs increases are possible (Pareto-Koopmans efficiency).⁵

The measurement of technical efficiency assuming variable returns to scale, VRS, as introduced by Banker *et al.* (1984) – known as the Banker, Charnes and Cooper, BCC, model, considers the following production possibility set $P_{\text{VRS}} = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X\lambda, \mathbf{y} \leq Y\lambda, \mathbf{e}\lambda = \mathbf{1}, \lambda \geq \mathbf{0}\}$. Therefore, the only difference with the CRS model is the adjunction of the condition $\sum_{j=1}^n \lambda_j = 1$. Calculating the VRS efficiency, along with the subsequent second stage program analogous to (2), yields the corresponding optimal solution $(\theta_{\text{VRS}}^*, \lambda^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$. As before, an observation is efficient with respect to the VRS technology if the optimal solution of the two programs satisfies $\theta_{\text{VRS}}^* = 1, \mathbf{s}^{-*} = 0$, and $\mathbf{s}^{+*} = 0$.

Finally, the scale efficiency, *SE*, of each observation is calculated as the ratio of the VRS to the CRS scores: $SE = \theta_{\text{CRS}}^* / \theta_{\text{VRS}}^*$. As a result the radial technical efficiency of an observation can be decomposed into its variable returns to scale efficiency (“pure” technical efficiency, *PTE*) and scale efficiency: $TE = \theta_{\text{CRS}}^* = PTE \times SE = \theta_{\text{VRS}}^* \times SE$.

The radial input oriented model can be computed in MATLAB using the `dea(X, Y, ...)` function with the `orient` parameter set to `io` (input oriented). The returns to scale assumption can be specified by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale). With the optional parameter `names` we can specify a cell string with the name of the decision making units, which in this case are the names of the American states.⁶

Results are returned in a ‘`deaout`’ structure and can be accessed directly (see Section 2) or displayed using the `deadisp` function. The second parameter of the `deadisp` function allows to specify the information that is displayed on screen. If not specified, the information displayed will depend on the calculated model.⁷

```
load 'DataAgriculture'
X = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y = [LS04, CROPS04, FARMOUT04];
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
io = dea(X, Y, 'orient', 'io', 'names', statenames);
deadisp(io, 'names/eff/slack.X/slack.Y');
```

Data Envelopment Analysis (DEA)

DMUs: 48

⁵It is possible to solve both programs in a single stage formulation employing a “non-Archimedean” infinitesimal constant ϵ , e.g., Ali and Seiford (1993, p. 140). However, this may result in computational inaccuracies and erroneous results.

⁶If the optional parameter `names` is not used, observations (DMUs) are numbered from 1 to n .

⁷See Section 10.3 for advanced uses of the `deadisp` function. Besides the US agricultural data, a second set of data used by Ali and Seiford (1993) is available at the toolbox website. A previous version of this toolbox illustrated the input, output, directional and additive models with these data.

Inputs: 4 Outputs: 3
 Model: radial
 Orientation: io (Input oriented)
 Returns to scale: crs (Constant)

```
-----
DMU|  Theta| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 0.9077| 58.2778| 54.5172| 0.0000| 0.0000| 0.0000| 331.2950| 0.0000|
AR| 0.9888| 0.0000| 94.0581| 0.0000| 0.0000| 0.0000| 0.0000| 121.3747|
AZ| 0.9698| 0.0000| 278.0711| 0.0000| 0.0000| 0.0000| 0.0000| 34.4711|
.
.
WI| 0.8452| 353.4822| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 332.8943|
WV| 0.6612| 0.0000| 34.2292| 55.9801| 0.0000| 0.0000| 29.2884| 0.0000|
WY| 0.5911| 5.9018| 465.9112| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
-----
```

```
io_vrs = dea(X, Y, 'orient', 'io', 'rts', 'vrs', 'names', statenames);
deadisp(io, 'names/eff/slack.X/slack.Y');
```

```
-----
Data Envelopment Analysis (DEA)
```

DMUs: 48
 Inputs: 4 Outputs: 3
 Model: radial
 Orientation: io (Input oriented)
 Returns to scale: vrs (Variable)

```
-----
DMU|  Theta| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 0.9666| 60.9790| 55.0145| 0.0000| 0.0000| 0.0000| 582.8943| 31.0395|
AR| 0.9891| 0.0000| 91.4690| 0.0000| 0.0000| 0.0000| 0.0000| 105.0558|
AZ| 0.9723| 0.0000| 275.5144| 0.0000| 0.0000| 0.0000| 0.0000| 10.0449|
.
.
WI| 1.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
WV| 0.6724| 2.2679| 45.0415| 92.7799| 0.0000| 0.0000| 57.0121| 0.0000|
WY| 0.5912| 5.1165| 465.4577| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
-----
```

The scale efficiency can be calculated using the `deascaleeff(X, Y, ...)` function. The function parameters are the same as those of the `dea` function, although the `rts` parameter specified will be omitted since both are needed in order to compute scale efficiency.


```
io_scale = deascale(X, Y, 'orient', 'io', 'names', statenames);
deadisp(io_scale);
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: io (Input oriented)
Returns to scale: scaleeff (Scale efficiency)
```

```
-----
DMU|      CRS|      VRS| ScaleEff|
-----
AL|  0.9077|  0.9666|  0.9390|
AR|  0.9888|  0.9891|  0.9998|
AZ|  0.9698|  0.9723|  0.9975|
.
.
WI|  0.8452|  1.0000|  0.8452|
WV|  0.6612|  0.6724|  0.9834|
WY|  0.5911|  0.5912|  0.9997|
-----
```

Results show the relative difference between the constant returns to scale (CRS) and variable returns to scale (VRS) efficiency scores. Since the VRS model envelops the data more tightly as a result of the additional constraint $\mathbf{e}\lambda = \mathbf{1}$, these scores are greater than their CRS counterparts, and therefore scale efficiency is also equal or less than one. For the selected states we observe that while none of them is technically efficient in the constant returns to scale model, with scores less than one, Wisconsin (WI) is efficient under the VRS assumption with a unitary score, and a relative scale efficiency of $SE = 0.8452$. Therefore, this state presents a suboptimal scale in terms of relative input and output quantities when compared to other counterparts that are almost efficient under CRS; e.g., Arkansas (AR). As for the input and output slacks, corresponding to no radial reductions and increased, we observe that there is an excessive usage of the second input (land), whose values are greater than zero across the majority of states. Other intermediate inputs and livestock output do not exhibit slacks at all.

3.2. Radial output oriented model: Constant and variable returns to scale

It is possible to measure the output oriented technical efficiency of each observation by solving the following linear program, counterpart to (1):

$$\begin{aligned} \max_{\phi, \lambda} \quad & \phi & (3) \\ \text{subject to} \quad & & \\ & X\lambda \leq \mathbf{x}_o & \\ & \phi \mathbf{y}_o \leq Y\lambda & \\ & \lambda \geq \mathbf{0}. & \end{aligned}$$

In this case, the optimal solution is denoted by ϕ_{CRS}^* with the constraints ensuring that $(\mathbf{x}_o, \phi_{\text{CRS}}^* \mathbf{y}_o)$ belongs to P_{CRS} . Now the objective seeks the maximum ϕ_{CRS} that increases the output vector \mathbf{y}_o radially to $\phi_{\text{CRS}}^* \mathbf{y}_o$ while remaining in P_{CRS} . A feasible solution signaling radial efficiency is $\phi_{\text{CRS}}^* = 1$. Therefore if $\phi_{\text{CRS}}^* > 1$, the observation is *radially* inefficient and $(\lambda X, \lambda Y)$ outperforms $(\mathbf{x}_o, \mathbf{y}_o)$. Again, there might be further input excesses and output shortfalls: $\mathbf{s}^- = \mathbf{x}_o - X\lambda$, and $\mathbf{s}^+ = Y\lambda - \phi_{\text{CRS}}^* \mathbf{y}_o$ with $\mathbf{s}^- \geq 0$ and $\mathbf{s}^+ \geq 0$ for any feasible solution (ϕ, λ) . To calculate these slacks in a second stage, the corresponding program incorporating the optimal value ϕ_{CRS}^* is needed:

$$\begin{aligned} & \max_{\lambda, \mathbf{s}^-, \mathbf{s}^+} && \omega = \mathbf{e}\mathbf{s}^- + \mathbf{e}\mathbf{s}^+ && (4) \\ & \text{subject to} && && \\ & && \mathbf{s}^- = \mathbf{x}_o - X\lambda && \\ & && \mathbf{s}^+ = Y\lambda - \phi_{\text{CRS}}^* \mathbf{y}_o && \\ & && \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}. && \end{aligned}$$

Finally, it is also possible to calculate the technical efficiency with respect to P_{VRS} by solving the programs for the radial component and its associated input and output slacks analogous to (3) and (4), but adding the VRS constraint $\sum_{j=1}^n \lambda_j = 1$. If $\phi_{\text{VRS}}^* = 1$ the observation is radially efficient, while it is technically efficient if $\mathbf{s}^{-*} = 0$ and $\mathbf{s}^{+*} = 0$ in the second stage. The scale efficiency defines in this case as $SE = \phi_{\text{CRS}}^* / \phi_{\text{VRS}}^*$ and radial efficiency can be now decomposed into pure technical efficiency, PTE , and scale efficiency: $TE = \phi_{\text{CRS}}^* = PTE \times SE = \phi_{\text{VRS}}^* \times SE$. The radial output oriented model is computed in MATLAB using the same `dea(X, Y, ...)` function with the `orient` parameter set to `oo` (output oriented). Again, the returns to scale assumption can be specified by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale).

```
oo = dea(X, Y, 'orient', 'oo', 'names', statenames);
deadisp(oo, 'names/eff/slack.X/slack.Y');
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: oo (Output oriented)
Returns to scale: crs (Constant)
```

```
-----
DMU|   Phi| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 1.1017| 64.2047| 60.0617| 0.0000| 0.0000| 0.0000|364.9884| 0.0000|
AR| 1.0113| 0.0000| 95.1193| 0.0000| 0.0000| 0.0000| 0.0000|122.7440|
AZ| 1.0311| 0.0000|286.7192| 0.0000| 0.0000| 0.0000| 0.0000| 35.5432|
.
.
```

```

WI| 1.1831|418.2209| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|393.8624|
WV| 1.5124| 0.0000| 51.7667| 84.6617| 0.0000| 0.0000| 44.2945| 0.0000|
WY| 1.6919| 9.9850|788.2563| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
-----

```

```

oo_vrs = dea(X, Y, 'orient', 'oo', 'rts', 'vrs', 'names', statenames);
deadisp(oo_vrs, 'names/eff/slack.X/slack.Y');

```

```

-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: oo (Output oriented)
Returns to scale: vrs (Variable)

```

```

-----
DMU|    Phi| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 1.0309| 63.6312| 58.1154| 0.0000| 0.0000| 0.0000|609.0434| 32.1028|
AR| 1.0111| 0.0000| 92.6889| 0.0000| 0.0000| 0.0000| 0.0000|107.5199|
AZ| 1.0286| 0.0000|283.2863| 0.0000| 0.0000| 0.0000| 0.0000| 10.1863|
.
.
WI| 1.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
WV| 1.4707| 23.1506| 71.4949|205.0725| 0.0000| 0.0000| 56.5374| 0.0000|
WY| 1.6911| 10.6927|788.3816| 1.9521| 0.0000| 0.0000| 0.0000| 0.0000|
-----

```

```

oo_scale = deascale(X, Y, 'orient', 'oo', 'names', statenames);
deadisp(oo_scale);

```

```

-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: oo (Output oriented)
Returns to scale: scaleeff (Scale efficiency)

```

```

-----
DMU|    CRS|    VRS| ScaleEff|
-----
AL| 1.1017| 1.0309| 1.0687|
AR| 1.0113| 1.0111| 1.0002|

```

AZ	1.0311	1.0286	1.0024
.			
WI	1.1831	1.0000	1.1831
WV	1.5124	1.4707	1.0283
WY	1.6919	1.6911	1.0004

Regarding the output oriented results, we first highlight that from DEA theory the output constant returns to scale scores are the inverse of their input counterparts; i.e., $\phi_{\text{CRS}}^* = 1/\theta_{\text{CRS}}^*$, and therefore none of our selected states are efficient. Additionally, it is also the case that if an observation is efficient under variable returns to scale from an input orientation, it is efficient from the output perspective (and for the forthcoming directional and additive models). Consequently Wisconsin (WI) is also VRS efficient from an output perspective. However scale efficiency can be substantially different depending on the orientation. The consistency of the results between the input and output models is also inferred from the positive values of the slacks corresponding to the second input (land).

3.3. The directional model: Constant and variable returns to scale

Chambers, Chung, and Färe (1996) introduced a measure of efficiency that projects observation $(\mathbf{x}_o, \mathbf{y}_o)$ in a pre-assigned direction $\mathbf{g} = (-\mathbf{g}_x^-, \mathbf{g}_y^+) \neq \mathbf{0}_{m+s}$, $\mathbf{g}_x^- \in \mathbb{R}^m$ and $\mathbf{g}_y^+ \in \mathbb{R}^s$, in a proportion β . The associated linear program is:

$$\begin{aligned} \max_{\beta, \lambda} \quad & \beta & (5) \\ \text{subject to} \quad & & \\ & X\lambda \leq \mathbf{x}_o - \beta\mathbf{g}_x^- & \\ & Y\lambda \geq \mathbf{y}_o + \beta\mathbf{g}_y^+ & \\ & \lambda \geq \mathbf{0}. & \end{aligned}$$

In this occasion the optimal solution to this program corresponds to β_{CRS}^* . Now $\beta_{\text{CRS}}^* = 0$ signals *directional* efficiency. Therefore if $\beta_{\text{CRS}}^* > 0$, the observation is inefficient and $(\lambda X, \lambda Y)$ outperforms $(\mathbf{x}_o, \mathbf{y}_o)$, with $(\mathbf{x}_o - \beta_{\text{CRS}}^* \mathbf{g}_x^-, \mathbf{y}_o + \beta_{\text{CRS}}^* \mathbf{g}_y^+) \in P_{\text{CRS}}$. It is again possible that further input excesses and output shortfalls exist. The slacks correspond to $\mathbf{s}^- = \mathbf{x}_o - \beta\mathbf{g}_x^- - X\lambda$, and $\mathbf{s}^+ = Y\lambda - \mathbf{y}_o + \beta\mathbf{g}_y^+$, respectively, with $\mathbf{s}^- \geq \mathbf{0}$ and $\mathbf{s}^+ \geq \mathbf{0}$ for any feasible solution (β, λ) . As a result, a second stage is once again needed to calculate these slacks. The next program incorporating the optimal value β_{CRS}^* allows determination of these values:

$$\begin{aligned} \max_{\lambda, \mathbf{s}^-, \mathbf{s}^+} \quad & \omega = \mathbf{e}\mathbf{s}^- + \mathbf{e}\mathbf{s}^+ & (6) \\ \text{subject to} \quad & & \\ & \mathbf{s}^- = \mathbf{x}_o - \beta\mathbf{g}_x^- - X\lambda & \\ & \mathbf{s}^+ = Y\lambda - \mathbf{y}_o + \beta\mathbf{g}_y^+ & \\ & \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}. & \end{aligned}$$

As in the inputs and outputs oriented models, one may also calculate technical efficiency with respect to P_{VRS} . This requires solving equivalent programs to (5) and (6) adding the VRS

constraint $\sum_{j=1}^n \lambda_j = 1$. If $\beta_{\text{VRS}}^* = 0$ and $\mathbf{s}^{-*} = \mathbf{0}$ and $\mathbf{s}^{+*} = \mathbf{0}$, the observation is *technically* efficient. Consequently, we now define scale efficiency as $SE = \beta_{\text{CRS}}^* - \beta_{\text{VRS}}^*$ and directional efficiency is decomposed into pure technical efficiency, *PTE*, and the scale efficiency term: $TE = \beta_{\text{CRS}}^* = PTE + SE = \beta_{\text{VRS}}^* + SE$.

The directional model is oriented both in the input and output dimensions, while the choice of directional vector corresponds to the researcher. Customarily, to keep consistency with the radial models, the observed amounts of inputs and outputs set the direction: $\mathbf{g} = (-\mathbf{g}_x^-, \mathbf{g}_y^+) = (-\mathbf{x}_o, \mathbf{y}_o)$, coinciding with the generalized Farrell measure (Briec 1997). In this case it can be shown that the directional model nests the input and output oriented models. Indeed, if $(-\mathbf{g}_x^-, \mathbf{g}_y^+) = (-\mathbf{x}_o, \mathbf{0})$, then $\beta^* = 1 - \theta^*$, while if $(-\mathbf{g}_x^-, \mathbf{g}_y^+) = (\mathbf{0}, \mathbf{y}_o)$, $\beta^* = \phi^* - 1$. However, other choices are available; particularly $(-\mathbf{g}_x^-, \mathbf{g}_y^+) = (-\mathbf{1}, \mathbf{1})$ or the mean of the data: $(-\mathbf{g}_x^-, \mathbf{g}_y^+) = (-\bar{\mathbf{x}}_o, \bar{\mathbf{y}}_o)$, which are neutral with respect to the orientation, as it does not use the individual weights corresponding to the observed amounts of inputs and outputs.⁸ When deciding on the directional model, the researcher must declare whether the direction corresponds to the observed input or output mixes, the unitary vectors, or her own choice of directional vector. In this case she must introduce the directional input and output matrices. The directional model can be computed in MATLAB using the `dea(X, Y, ...)` function with the `orient` parameter set to `ddf` (directional). The input and output directions are specified in the `Gx` and `Gy` parameters as a matrix or as a scalar (usually, 0 or 1). If omitted, `X` and `Y` will be used for `Gx` and `Gy` respectively. The returns to scale assumption can be specified by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale).

```
ddf = dea(X, Y, 'orient', 'ddf', 'Gx', X, 'Gy', Y, 'names', statenames);
deadisp(ddf, 'names/eff/slack.X/slack.Y');
```

Data Envelopment Analysis (DEA)

DMUs: 48

Inputs: 4 Outputs: 3

Model: radial

Orientation: ddf (Directional distance function)

Returns to scale: crs (Constant)

```
-----
DMU|   Beta| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL|  0.0484| 61.0978| 57.1553|  0.0000|  0.0000|  0.0000|347.3265|  0.0000|
AR|  0.0056|  0.0000| 94.5857|  0.0000|  0.0000|  0.0000|  0.0000|122.0555|
AZ|  0.0153|  0.0000|282.3289|  0.0000|  0.0000|  0.0000|  0.0000| 34.9990|
.
```

⁸Other directions are possible, including elaborated transformations driven by the data as proposed by Daraio and Simar (2016), or those projecting observations to economic optima as introduced by Zofío, Pastor, and Aparicio (2013), e.g., maximum profit – as shown in Section 6.

```

.
WI| 0.0839|383.1361| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|360.8211|
WV| 0.2039| 0.0000| 41.2097| 67.3964| 0.0000| 0.0000| 35.2613| 0.0000|
WY| 0.2570| 7.4187|585.6593| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
-----

```

```

ddf_vrs = dea(X, Y, 'orient', 'ddf', 'rts', 'vrs', 'Gx', X, 'Gy', Y, ...
'names', statenames);
deadisp(ddf_vrs, 'names/eff/slack.X/slack.Y');

```

```

-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: ddf (Directional distance function)
Returns to scale: vrs (Variable)

```

```

-----
DMU|   Beta| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 0.0161| 62.3563| 56.6249| 0.0000| 0.0000| 0.0000|596.4738| 31.5917|
AR| 0.0055| 0.0000| 92.0751| 0.0000| 0.0000| 0.0000| 0.0000|106.2801|
AZ| 0.0141| 0.0000|279.3368| 0.0000| 0.0000| 0.0000| 0.0000| 10.1145|
.
.
WI| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
WV| 0.1932| 10.8377| 55.8973|138.8620| 0.0000| 0.0000| 56.8173| 0.0000|
WY| 0.2570| 6.9367|585.3507| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
-----

```

```

ddf_scale = deascale(X, Y, 'orient', 'ddf', 'Gx', X, 'Gy', Y, ...
'names', statenames);
deadisp(ddf_scale);

```

```

-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial
Orientation: ddf (Directional distance function)
Returns to scale: scaleeff (Scale efficiency)

```

```

-----
DMU|      CRS|      VRS| ScaleEff|

```

AL	0.0484	0.0161	0.0323
AR	0.0056	0.0055	0.0001
AZ	0.0153	0.0141	0.0012
.			
.			
WI	0.0839	0.0000	0.0839
WV	0.2039	0.1932	0.0108
WY	0.2570	0.2570	0.0001

The results corresponding to the directional distance function are not bounded by one as in the previous input and output models. Nevertheless, if any observation is efficient in the former case with a score equal to one, either under CRS or VRS, it is also efficient under the directional approach because it defines the production frontier; i.e., inputs and outputs cannot be simultaneously reduced or increased, respectively. That is why Wisconsin (WI) presents a zero valued score under VRS. These scores can be compared among themselves, and in this case the greater the value, the more inefficient is the state. Correspondingly, as in the previous models, the most inefficient state among the selected ones is Wyoming (WY). The values of the input and output slacks also present a similar pattern to that previously commented.

3.4. The additive model

The additive model measures technical efficiency based solely on input excesses and output shortfalls. It does not calculate efficiency scores corresponding to the radial or directional interpretation of technical efficiency *a la* Farrell (1957), but characterizes efficiency in terms of the input and output slacks: $\mathbf{s}^- \in \mathbb{R}^m$ and $\mathbf{s}^+ \in \mathbb{R}^s$, respectively. The toolbox implements the weighted additive formulation of Lovell and Pastor (1995) and Pastor, Lovell, and Aparicio (2011), whose associated linear program is:

$$\begin{aligned}
 & \max_{\lambda, \mathbf{s}^-, \mathbf{s}^+} && \omega = \rho_{\mathbf{x}}^- \mathbf{s}^- + \rho_{\mathbf{y}}^+ \mathbf{s}^+ && (7) \\
 & \text{subject to} && && \\
 & && X\lambda + \mathbf{s}^- = \mathbf{x}_o && \\
 & && Y\lambda - \mathbf{s}^+ = \mathbf{y}_o && \\
 & && \mathbf{e}\lambda = \mathbf{1} && \\
 & && \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, &&
 \end{aligned}$$

where $(\rho_{\mathbf{x}}^-, \rho_{\mathbf{y}}^+) \in \mathbb{R}_+^m \times \mathbb{R}_+^s$ are the inputs and outputs weight vectors whose elements can vary across DMUs. Therefore, assigning unitary values, program (7) corresponds to the standard additive model, while it is worth noting that it encompasses a wide class of different DEA models known as general efficiency measures (GEMs). Particularly, for the measure of inefficiency proportions (MIP): $(\rho_{\mathbf{x}}^-, \rho_{\mathbf{y}}^+) = (1/\mathbf{x}_o, 1/\mathbf{y}_o)$; for the range adjusted measure (RAM): $(\rho^-, \rho^+) = (1/(m+s)R^-, 1/(m+s)R^+)$, where R^- and R^+ are the variables' ranges; while for the bounded adjusted measure (BAM): $(\rho_{\mathbf{x}}^-, \rho_{\mathbf{y}}^+) = (1/(m+s)(\mathbf{x}_o - \mathbf{x}), 1/(m+s)(\mathbf{y}_o - \mathbf{y}))$, where \mathbf{x} and \mathbf{y} are the minimum observed values. Additionally, it is possible to consider

other fixed values with the weights representing value judgments. For observation $(\mathbf{x}_o, \mathbf{y}_o)$ the objective seeks the maximum feasible reduction in its inputs and increase in its outputs while remaining in P_{VRS} . An observation is *technically* efficient if the optimal solution $(\lambda^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ of the the program is $\mathbf{s}^{-*} = 0$, and $\mathbf{s}^{+*} = 0$. Otherwise individual input reductions and output increases would be feasible, and the larger the sum of the slacks, ω_{VRS}^* , the larger the inefficiency. The relevance of these transformations is that they make the additive measures independent of the units of measurement, which is a desirable property.

The function `deaaddit(X, Y, ...)` solves the weighted additive model in MATLAB. The returns to scale assumption can be specified by setting the `rts` parameter to `vrs` (variable returns to scale). Inputs and outputs weights are specified in the `rhoX` and `rhoY` parameters. The default weights correspond to the MIP model if not included.

```
add_vrs = deaaddit(X, Y, 'rts', 'vrs', 'names', statenames);
deadisp(add_vrs, 'names/eff/slack.X/slack.Y');
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: additive
Orientation: none
Returns to scale: vrs (Variable)
```

```
-----
DMU|   Eff| slackX1| slackX2| slackX3| slackX4| slackY1| slackY2| slackY3|
-----
AL| 1.2677| 78.7255| 68.5399| 0.0000| 0.0000| 0.0000|718.8611| 29.8641|
AR| 0.6874| 37.0255|118.8919| 0.0000| 0.0000| 0.0000| 0.0000|136.2832|
AZ| 1.1851| 25.0847|304.5817| 33.4556| 0.0000| 16.9063| 0.0000| 30.7548|
.
.
WI| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000| 0.0000|
WV| 4.7765| 66.6089| 66.4211|217.3747| 0.0000| 0.0000|490.7063| 21.8707|
WY| 4.9848| 67.8369|783.7540| 0.0000| 0.0000| 65.8284|810.6456| 48.5739|
-----
```

For illustration purposes, the range adjusted measure (RAM) model can be computed by specifying the appropriate input and output slacks weights:

```
n = size(X, 1);
m = size(X, 2);
s = size(Y, 2);
rhoX = repelem(1 ./ ((m + s) * range(X, 1)), n, 1);
rhoY = repelem(1 ./ ((m + s) * range(Y, 1)), n, 1);
add_ram = deaaddit(X, Y, 'rts', 'vrs', 'rhoX', rhoX, 'rhoY', rhoY, ...
    'names', statenames);
deadisp(add_ram, 'names/eff/slack.X/slack.Y');
```

 Data Envelopment Analysis (DEA)

DMUs: 48
 Inputs: 4 Outputs: 3
 Model: additive
 Orientation: none
 Returns to scale: vrs (Variable)

DMU	Eff	slackX1	slackX2	slackX3	slackX4	slackY1	slackY2	slackY3
AL	0.0168	62.7419	40.8181	5.3285	168.2713	0.0000	554.2238	47.4548
AR	0.0169	37.0255	118.8919	0.0000	0.0000	0.0000	0.0000	136.2832
AZ	0.0149	25.0847	304.5817	33.4556	0.0000	16.9063	0.0000	30.7548
.
WI	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
WV	0.0199	80.4503	101.1508	385.9306	0.0000	117.8994	2.8581	0.0000
WY	0.0442	85.7319	838.6588	257.0071	19.0255	247.3773	0.0000	11.3209

Again, if an observation is identified as efficient by belonging to the the production frontier under any of the previous models, including zero valued input or output slacks, it shows an inefficiency score equal to zero in the additive models. This is the case of Wisconsin (WI). Note also that since the slacks weighting scheme is introduced in the objective function, the value of the optimal input and output slacks may remain the same. This is the case if the evaluated state is projected to the strongly efficient frontier, i.e., once slacks are accounted for, as it is the case of Arkansas (AR) in the above results. This not the case however for the scores of inefficient states. Taking Arizona (AZ) as the exemplifying case, its score is $\omega = 0.1851$ if the measure of inefficiency proportions (MIP) is calculated, while if the ranges of the variables are used to weight the slacks (RAM), the score reduces to $\omega = 0.0149$. It is relevant to remark that the use of alternative weights does not only change the value of the inefficiency scores, but may also change the relative rankings, even if the weights are the same for all observations, i.e., as in the RAM case. Hence, researchers should bear in mind that the choice of weights is not neutral, because it may result in alternative efficiency rankings.

3.5. Super-efficiency models

One interesting model that allows to differentiate across technically efficient observations is that proposed by [Andersen and Petersen \(1993\)](#). As [Angulo-Meza and Pereira Estellita Lins \(2002\)](#) highlight, one of the relevant drawbacks of standard DEA models is their inability to rank weakly efficient units differently as they all are assigned the same unitary efficiency score. The super-efficiency scores allow discriminating across them by calculating an individual score that is different across observations. These scores are obtained by individually solving for each observation any of the previous models, but excluding them from the reference dataset, which therefore reduces to $n - 1$ observations; i.e., $\hat{P}_{\text{CRS}}(\mathbf{x}_o, \mathbf{y}_o) =$

$\{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{x}_j, \mathbf{y} \leq \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{y}_j, \lambda \geq \mathbf{0}\}$. The magnitude of the super-efficiency score, as well as the number of observations whose efficiency changes as a result of each individual exclusion determine the importance of each efficient observation in the complete dataset.

Therefore, if an observation is inefficient, its super-efficiency score is the same as that previously calculated, while in the efficient case it is greater than one for the radially input oriented model, less than one for the output counterpart, and negative for the directional model – as inputs are to be increased and outputs reduced to reach the reference benchmarks.⁹

For these oriented models we show in what follows the formulation corresponding to the input orientation under CRS, while its output and directional counterparts are omitted for the sake of brevity. For the same reason, while the MATLAB toolbox internally solves a two stage process, the problem can be equivalently expressed according to the following single stage non-Archimedean formulation:

$$\begin{aligned} & \min_{\tilde{\theta}, \lambda, \mathbf{s}^-, \mathbf{s}^+} && \tilde{\theta} - \varepsilon(\mathbf{1}\mathbf{s}^- + \mathbf{1}\mathbf{s}^+) && (8) \\ & \text{subject to} && && \\ & && \tilde{\theta}\mathbf{x}_o = \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{x}_j + \mathbf{s}^- && \\ & && \mathbf{y}_o = \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{y}_j - \mathbf{s}^+ && \\ & && \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}, && \end{aligned}$$

where $\mathbf{s}^- \in \mathbb{R}^m$ and $\mathbf{s}^+ \in \mathbb{R}^s$, and ε is an infinitesimal constant. The optimal solution – obtained from a first stage equivalent to (1) but excluding the DMU under evaluation from the reference set – is again denoted by $\tilde{\theta}_{\text{CRS}}^*$. In this case, if $\tilde{\theta}_{\text{CRS}}^* > 1$, the observation is super-efficient and the larger the score and the increase in the values of the remaining observations with respect to the original calculations in (1), the more relevant is the observation that has been removed when defining the production frontier. Also, if an observation is inefficient when solving (1), its efficiency score does not change when removing it from the production possibility set, simply because it never defined the production frontier in the first place, and its original benchmarks remain the same. Implementing the super-efficiency method in the directional distance function model follows similar steps, while variable returns to scale can be computed by adding the corresponding constraint: $\sum_{j=1, \neq o}^{n-1} \lambda_j = 1$.

The radial super-efficiency model corresponds to the `deasuper(X, Y, ...)` function in MATLAB with the `orient` parameter set to the desired orientation (`io`, `oo`, or `ddf`). Once again, the returns to scale assumption can be specified by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale).

```
super = deasuper(X, Y, 'orient', 'io', 'rts', 'vrs', 'names', statenames);
deadisp(super, 'names/eff/slack.X/slack.Y');
```

Data Envelopment Analysis (DEA)

⁹When solving these models under the VRS assumption, it is likely that some unfeasible results are obtained.

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial-supereff
Orientation: io (Input oriented)
Returns to scale: vrs (Variable)

```

```

-----
DMU|  Theta| slackX1| slackX2| slackX3| slackX4| slackY1|  slackY2| slackY3|
-----
AL| 0.9666| 60.9790| 55.0145|  0.0000|  0.0000|  0.0000| 582.8943| 31.0395|
AR| 0.9891|  0.0000| 91.4690|  0.0000|  0.0000|  0.0000|  0.0000|105.0558|
AZ| 0.9723|  0.0000|275.5144|  0.0000|  0.0000|  0.0000|  0.0000| 10.0449|
.
.
WI| 1.0273|625.3266|  0.0000|179.6911|  0.0000|  0.0000|2448.7623|510.6159|
WV| 0.6724|  2.2679| 45.0415| 92.7799|  0.0000|  0.0000|  57.0121|  0.0000|
WY| 0.5912|  5.1165|465.4577|  0.0000|  0.0000|  0.0000|  0.0000|  0.0000|
-----

```

```

superddf = deasuper(X, Y, 'orient', 'ddf', 'Gx', X, 'Gy', Y, 'rts', 'vrs', ...
    'names', statenames);
deadisp(superddf, 'names/eff/slack.X/slack.Y');

```

```

-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: directional-supereff
Orientation: ddf (Directional distance function)
Returns to scale: vrs (Variable)

```

```

-----
DMU|  Beta| slackX1| slackX2| slackX3| slackX4| slackY1|  slackY2| slackY3|
-----
AL| 0.0161| 62.3563| 56.6249|  0.0000|  0.0000|  0.0000| 596.4738| 31.5917|
AR| 0.0055|  0.0000| 92.0751|  0.0000|  0.0000|  0.0000|  0.0000|106.2801|
AZ| 0.0141|  0.0000|279.3368|  0.0000|  0.0000|  0.0000|  0.0000| 10.1145|
.
.
WI|-0.0104|611.4014|  0.0000|187.9909|  0.0000|  0.0000|2331.9308|501.9102|
WV| 0.1932| 10.8377| 55.8973|138.8620|  0.0000|  0.0000|  56.8173|  0.0000|
WY| 0.2570|  6.9367|585.3507|  0.0000|  0.0000|  0.0000|  0.0000|  0.0000|
-----

```

The results obtained for the super-efficient radial input oriented model under VRS show that once the observations are removed from the production possibility set, their efficiency scores

may exceed the unitary value. This is the case of Wisconsin (WI), with $\tilde{\theta}_{\text{VRS}}^* = 1.0273 > 1$, showing that input quantities must be actually *increased* to reach the production frontier. On the contrary, for the rest of the selected states, their efficiency scores remain unchanged from those reported for (1) since they are inefficient. Similarly, for the super-efficient directional distance function model. The optimal inefficiency score for Wisconsin (WI) is now *negative*: $\tilde{\beta}_{\text{VRS}}^* = -0.0104 < 0$, and therefore inputs must be increased and outputs reduced to reach the production frontier. Again, the values for the remaining states do not change from those obtained when solving (5). Note also that the super-efficiency model yields its own set of optimal input and output slacks.

As for the super-efficiency calculations in the additive model, we follow Du, Liang, and Zhu (2010), and solve the corresponding counterpart to (7) – for the case of the standard additive model:

$$\begin{aligned} \min_{\lambda, \mathbf{s}^-, \mathbf{s}^+} \quad & \tilde{\omega} = \rho_{\mathbf{x}}^- \mathbf{s}^- + \rho_{\mathbf{y}}^+ \mathbf{s}^+ & (9) \\ \text{subject to} \quad & \\ & \mathbf{x}_o \geq \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{x}_j - \mathbf{s}^- \\ & \mathbf{y}_o \leq \sum_{j=1, \neq o}^{n-1} \lambda_j \mathbf{y}_j + \mathbf{s}^+ \\ & \mathbf{e}\lambda = \mathbf{1} \\ & \lambda \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}. \end{aligned}$$

The constraints in the program are modified as inputs need to be increased and outputs reduced so as to reach the production possibility set. Again, $(\rho_{\mathbf{x}}^-, \rho_{\mathbf{y}}^+) \in \mathbb{R}_+^m \times \mathbb{R}_+^s$ are the corresponding weight vectors, and changing their value allows to calculate a wide range of efficiency scores, including the MIP, RAM and BAM models. By default, as in (7), program (9) calculates the measure of inefficiency proportions (MIP) with $(\rho_{\mathbf{x}}^-, \rho_{\mathbf{y}}^+) = (1/\mathbf{x}_o, 1/\mathbf{y}_o)$. However one may set the individual weights to any value, including the unitary values corresponding to the standard additive model. On this occasion an observation is super-efficient if the optimal solution $(\lambda^*, \mathbf{s}^{-*}, \mathbf{s}^{+*})$ yields positive slacks, and the largest their sum, the largest the super-efficiency.

The additive super-efficiency model can be computed using the function `deaadditsuper(X, Y, ...)`, and specifying returns to scale by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale). Inputs and outputs weights are specified in the `rhoX` and `rhoY` parameters. If omitted, the weights corresponding to the MIP model are used.

```
additsuper = deaadditsuper(X, Y, 'rts', 'vrs', 'names', statenames);
deadisp(additsuper, 'names/eff/slack.X/slack.Y');
```

Data Envelopment Analysis (DEA)

DMUs: 48

Inputs: 4 Outputs: 3
 Model: additive-supereff
 Orientation: none
 Returns to scale: vrs (Variable)

DMU	Eff	slackX1	slackX2	slackX3	slackX4	slackY1	slackY2	slackY3
CA	1.5210	0.0000	0.0000	0.0000	0.0000	2362.0479	13108.1558	1058.1630
CT	0.0181	0.0000	0.0000	0.0000	2.5121	0.0000	0.0000	0.0000
DE	1.0445	0.0000	0.0000	14.3546	0.0000	355.3853	0.0000	18.3736
.								
TX	0.2305	0.0000	0.0000	0.0000	0.0000	197.1796	0.0000	388.8985
VT	0.1754	0.0000	0.0000	0.0000	1.6959	76.3999	0.0000	0.0000
WI	0.0168	0.0000	0.0000	0.0000	0.0000	77.4414	0.0000	0.0000

Program (9) solves the super-efficiency model only for the states that are efficient in its standard additive counterpart (7), as inefficient observations have no interpretable solution. In this case, the MATLAB output returns NaN for these observations. That is why previous inefficient states are not reported in the above results (e.g., Alabama (AL), Arkansas (AR), etc.). Only Wisconsin (WI) is shown with a super-efficiency score of $\tilde{\omega}_{\text{VRS}}^* = 0.0168 > 0$. In this case, the super-efficiency slack can be found on the output side, implying that to reach the production frontier from which this state is excluded, only the first output (livestock) must be reduced. This is the reason why its super-efficiency score is rather low when compared to other efficient states such as California (CA), with $\tilde{\omega}_{\text{VRS}}^* = 1.5210$. California presents both a greater slack value in the first output, as well as positive values in the remaining two outputs (crops and other farm-related output). Other states, e.g., Delaware (DE) with $\tilde{\omega}_{\text{VRS}}^* = 1.0445$, present super-efficiency slacks both in the input and output dimensions, but their absolute values are less than those of California, and therefore this state ranks in a lower position. From these results it is worth remarking that a livestock slack is consistently found for the majority of states, showing its relevance when defining the production frontier in the standard additive model.

4. Cross efficiency

As the previous approach, cross efficiency models allow to overcome the low discriminatory power of standard DEA, incapable of ranking the subset of efficient firms. Ultimately, the flexibility of DEA when searching for the most favorable weights may turn against the method itself by hampering its discriminatory power. A problem that is aggravated when the degrees of freedom are limited, with a small number of observations relative to the number of inputs and outputs. Initially developed by [Sexton, Silkman, and Hogan \(1986\)](#), cross efficiency methods rely on peer appraisals instead of self-evaluations. Here we follow [Doyle and Green \(1994\)](#) and implement the cross efficiency model based on the multiplier formulations of the radial input measures under constant returns to scale, (1). The basic cross efficiency concept

uses the obtained weights for observation j , (u_i^{*j}, v_i^{*j}) , and multiplies them by the input and output quantities of the unit under evaluation $(\mathbf{x}_o, \mathbf{y}_o)$, thereby generating a peer-appraisal (cross efficiency) score for every pair (j, o) . This peer-appraisal efficiency score corresponds to:

$$\theta_{j,o} = \frac{\sum_{i=1}^s u_i^{*j} y_i^o}{\sum_{i=1}^s v_i^{*j} x_i^o}.$$

After performing these calculations, one is left with a $n \times n$ matrix of peer-appraisal scores, from which the cross efficiency scores for every observation can be derived. Then cross efficiency may be calculated as $e_o = \frac{1}{n} \sum_{j=1}^n \theta_{j,o}$ if own efficiency scores are included, or $e'_o = \frac{1}{n-1} \sum_{j=1}^n \theta_{j,o}$ if they are excluded. However, because the optimal weights (u_i^{*j}, v_i^{*j}) obtained from the first stage DEA run are not unique, one may obtain different results for e_o in successive implementations. To remedy this situation a secondary goal is defined. The general idea of the benevolent and aggressive approaches is to solve the cross efficiency model with an objective to maximize or minimize the sum of all peer-appraisal scores, subject to the restrictions that: a) the self-appraisal scores remain equal to the results of the standard first stage DEA run, and b) no peer-appraisal score is greater than one. Because optimizing over the sum of ratios results in a non-linear problem, Sexton *et al.* (1986) propose a “linear surrogate”, which uses the sum of all observed j inputs and outputs, $j \neq o$, multiplied by the weights of observation o , (u_i^o, v_i^o) : i.e., aggregate outputs: $\sum_{j \neq o} \sum_i y_i^j u_i^o$, and aggregate inputs: $\sum_{j \neq o} \sum_i x_i^j v_i^o$. Subtracting the former from the latter results in the following linear problem that is solved as a second stage of the method:

$$\max_{v_i^o, u_i^o} E_o = \sum_{j \neq o} \sum_i y_i^j u_i^o - \sum_{j \neq o} \sum_i x_i^j v_i^o \quad (10)$$

subject to

$$\begin{aligned} \sum_i x_i^o v_i^o &= 1, \\ \sum_i y_i^o u_i^o - \theta_{o,o} \sum_i x_i^o v_i^o &= 0, \\ \theta_{j,o} &\leq 1 \quad \forall \quad j \neq o, \\ v_i^o, u_i^o &\geq 0. \end{aligned}$$

Program (10) corresponds to the benevolent approach. Minimizing of the objective function results in the aggressive formulation. One may average the results obtained from both approaches, for a less extreme set of cross efficiencies.¹⁰

The cross efficiency model can be computed in MATLAB using the `deacross(X, Y, ...)` function with the `objective` parameter set to `benevolent` for the benevolent formulation

¹⁰Cook and Zhu (2015) provide a recent review of cross efficiency methods based on multiplicative, game theoretic, and alternative approaches. A promising venue of research brings fuzzy DEA methods into the cross efficiency model, see Costantino, Dotoli, Epicoco, Falagario, and Sciancalepore (2012a) and Dotoli, Epicoco, Falagario, and Sciancalepore (2015).

(maximization) or to `aggressive` for the aggressive formulation (minimization). If the parameter `mean` is set to `inclusive` the cross efficiency score is calculated including the evaluated DMU, whereas if it is set to `exclusive` the evaluated DMU is excluded.

```
load 'DataAgriculture'
X = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y = [LS04, CROPS04, FARMOUT04];
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
cross_ben = deacross(X, Y, 'objective', 'benevolent', 'mean', ...
    'exclusive', 'names', statenames);
deadisp(cross_ben);
```

Data Envelopment Analysis (DEA)

DMUs: 48
Inputs: 4 Outputs: 3
Model: deacross-linear
Orientation: io (Input oriented)
Returns to scale: crs (Constant)

Cross efficiency: Benevolent Approach

```
-----
DMU|      Eff| CrossEff|
-----
AL|  0.9077|  0.7731|
AR|  0.9888|  0.8522|
AZ|  0.9698|  0.8532|
.
.
WI|  0.8452|  0.6721|
WV|  0.6612|  0.4597|
WY|  0.5911|  0.4846|
-----
```

CrossEff = Cross Efficiency (exclusive mean)

```
cross_agg = deacross(X, Y, 'objective', 'aggressive', 'mean', ...
    'inclusive', 'names', statenames);
deadisp(cross_agg);
```

Data Envelopment Analysis (DEA)

DMUs: 48
Inputs: 4 Outputs: 3

Model: deacross-linear
 Orientation: io (Input oriented)
 Returns to scale: crs (Constant)

Cross efficiency: Aggressive Approach

```
-----
DMU|      Eff| CrossEff|
-----
AL|  0.9077|  0.6827|
AR|  0.9888|  0.7430|
AZ|  0.9698|  0.7224|
.
.
WI|  0.8452|  0.5804|
WV|  0.6612|  0.4005|
WY|  0.5911|  0.3806|
-----
```

CrossEff = Cross Efficiency (inclusive mean)

Results for the benevolent and aggressive models show that, when compared to their standard radial input measure (1) counterparts, cross efficiency scores have smaller values. This is consistent with the fact that the optimal solution to the standard DEA corresponds to the most favorable weights (u_i^{*o}, v_i^{*o}) that the program can find, and therefore assessing efficiency based on any other pair of weights (u_i^{*j}, v_i^{*j}) results in smaller individual cross efficiencies; i.e., $\theta_{o,o} > \theta_{j,o}$, and therefore their inclusive or exclusive averages. For illustrative purposes, and taking as example West Virginia (WV), its standard efficiency score (1) is $\theta_{\text{CRS}}^* = 0.6612$, greater than its benevolent cross efficient counterpart, $\text{CrossEff} = 0.4597$. In turn, this latter score is larger than that corresponding to the aggressive approach 0.4005. Similar remarks can be made for the other selected states. Note also that for efficient states, their cross efficiency score would be less than one, and a comprehensive ranking could be obtained using these approaches.

5. Productivity change: The Malmquist index

The Malmquist index, introduced by Caves, Christensen, and Diewert (1982), measures the change in productivity of the observation under evaluation by comparing its relative performance with respect to reference technologies corresponding to two different time periods. The constant returns to scale production possibility set corresponding to $t = 1, \dots, T$ periods defines as $P_{\text{CRS}}^t = \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \geq X^t \lambda, \mathbf{y} \leq Y^t \lambda, \lambda \geq \mathbf{0}\}$, with the variable returns to scale counterpart denoted accordingly by P_{VRS}^t . The standard Malmquist index relies solely on the concept of radial efficiency and requires calculation of the input – or output – oriented scores of $(\mathbf{x}_o^t, \mathbf{y}_o^t)$ observed in two periods $t = 1, 2$, with respect to the constant returns to scale reference frontier of any period. Taking as the reference the first period, P_{CRS}^1 , we denote both scores by $\theta_{\text{CRS}}^{1,1}$ and $\theta_{\text{CRS}}^{2,1}$, where the first superscript refers to the time period of the observation and the second one to that of the reference technology. While $\theta_{\text{CRS}}^{1,1}$ is the solution to program (1), the intertemporal score $\theta_{\text{CRS}}^{2,1}$ corresponds to the following program that evaluates

period 2 observation $(\mathbf{x}_o^2, \mathbf{y}_o^2)$ with respect to period 1 technology:

$$\begin{aligned} \min_{\theta, \lambda} \quad & \theta & (11) \\ \text{subject to} \quad & & \\ & \theta \mathbf{x}_o^2 \geq X^1 \lambda & \\ & Y^1 \lambda \geq \mathbf{y}_o^2 & \\ & \lambda \geq \mathbf{0}. & \end{aligned}$$

Equivalently, the analogous intertemporal score $\theta_{\text{CRS}}^{1,2}$ using the second period technology as reference corresponds to the same program but reverses the time superscripts of the firm under evaluation from $(\mathbf{x}_o^2, \mathbf{y}_o^2)$ to $(\mathbf{x}_o^1, \mathbf{y}_o^1)$, and those of the reference technology from (X^1, Y^1) to (X^2, Y^2) .

After the contemporary and intertemporal efficiency scores have been calculated it is possible to define the following Malmquist indices: $M_1 = \theta_{\text{CRS}}^{2,1}/\theta_{\text{CRS}}^{1,1}$ and $M_2 = \theta_{\text{CRS}}^{2,2}/\theta_{\text{CRS}}^{1,2}$. For both indices, if $M > 1$ there is productivity increase, while if $M = 1$ productivity remains unchanged and $M < 1$ signals productivity decline. Following [Färe, Grosskopf, Norris, and Zhang \(1994\)](#), productivity change can be decomposed into efficiency change and technical change.¹¹ The former corresponds to the so-called *catch-up* effect; i.e., the change in the technical efficiency of the observation under evaluation with respect to the two periods, which defines for both indices as $MTEC = \theta_{\text{CRS}}^{2,2}/\theta_{\text{CRS}}^{1,1}$. The latter corresponds to the *frontier-shift* effect, i.e., the change in the reference frontier between both periods, which defines for M_1 as $MTC_1 = (\theta_{\text{CRS}}^{2,1}/\theta_{\text{CRS}}^{2,2})$ using period 2 observation as the reference benchmark to evaluate the shift in the frontier. For M_2 it defines a $MTC_2 = (\theta_{\text{CRS}}^{1,1}/\theta_{\text{CRS}}^{1,2})$. As in the previous cases, if $MTEC > 1$ or $MTC > 1$ productivity change is driven respectively by both technical efficiency gains and technical change improvements (technical progress); while $MTEC < 1$ or $MTC < 1$ imply lower productivity due to greater inefficiency and technical regress. Finally, unitary values signal that the technical efficiency and the reference frontier remain unchanged. Therefore the decomposition of year-to-year productivity change defines as $M_1 = MTEC \times MTC_1 = \theta_{\text{CRS}}^{2,2}/\theta_{\text{CRS}}^{1,1} \times \theta_{\text{CRS}}^{2,1}/\theta_{\text{CRS}}^{2,2}$ – and similarly for M_2 , while the geometric mean of both indices and their corresponding decompositions represents a compromise between both values: $M = MTEC \times MTC = \theta_{\text{CRS}}^{2,2}/\theta_{\text{CRS}}^{1,1} \times (\theta_{\text{CRS}}^{2,1}/\theta_{\text{CRS}}^{2,2} \times \theta_{\text{CRS}}^{1,1}/\theta_{\text{CRS}}^{1,2})^{(1/2)}$.

Finally, it is also possible to decompose long term productivity change from an initial period to a final period into consecutive subperiods relying on the transitivity property of index numbers. This allows the analysis of productivity change by subperiods. Therefore, given a sequence of periods, i.e., $t = 1, 2, 3$, it is verified that the Malmquist index between the base and final periods can be expressed in terms of its chain components: $M^{1,3} = M^{1,2} \times M^{2,3}$. The toolbox calculates the sequence of year-to-year Malmquist indices and the cumulative Malmquist index taking the based period as reference.

To compute the Malmquist indices in MATLAB one calls the `dealm(X, Y, ...)` function with the `orient` parameter set to any of the two orientations (`io` or `oo`). The parameters X

¹¹The toolbox calculates a first level decomposition that does not take into account the contribution that scale changes and input and output mixes make to productivity change. Therefore, as productivity change is measured under the CRS assumptions, the input and output orientations are equal. To account for scale effects alternative second level decompositions have been proposed in the literature, see [Zofío \(2007\)](#). An authoritative discussion of meaningful decompositions of total factor productivity indices is presented by [Balk and Zofío \(2018\)](#), along with its accompanying toolbox, [Balk, Barbero, and Zofío \(2020\)](#).

and Y must be 3D arrays, with the third dimension corresponding to different time periods. By default, Malmquist indices are computed as the geometric mean of M_1 and M_2 . However, for simplicity, the indices can be computed taking the first period as the base for technical change, M_1 , by setting the parameter `geomean` to 0. Also, in case of having more than two time periods, the transitive Malmquist index taking always the first period as the reference technology can be computed by setting the `fixbaset` parameter to 1.

```
load 'DataAgriculture'
X00 = [CAPITAL00, LAND00, LABOR00, INTINP00];
Y00 = [LS00, CROPS00, FARMOUTO0];
X04 = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y04 = [LS04, CROPS04, FARMOUTO4];
X = X00;
X(:,:,2) = X04;
Y = Y00;
Y(:,:,2) = Y04;
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
malmquist = deamalm(X, Y, 'orient', 'io', 'names', statenames);
deadisp(malmquist)
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: radial-malmquist
Orientation: io (Input oriented)
Returns to scale: crs (Constant)
```

```
Malmquist:
Geometric mean is computed
Base period is previous period
```

```
-----
DMU|      M|      MTEC|      MTC|
-----|-----|-----|-----
AL|  1.0390|  1.0032|  1.0356|
AR|  1.1543|  1.1122|  1.0378|
AZ|  0.9948|  0.9769|  1.0183|
.
.
WI|  1.0469|  1.0392|  1.0075|
WV|  1.0282|  1.0666|  0.9641|
WY|  0.9170|  0.9115|  1.0061|
-----
```

M = Malmquist. MTEC = Technical Efficiency Change. MTC = Technical Change.

Calculating the Malmquist productivity index for the selected states results in quite different trends and sources of productivity change in the five year period between 2000–2004. While Alabama (AL), Arkansas (AR), Wisconsin (WI) and West Virginia (VW) experience productivity growth, $M > 1$, Arizona (AR) presents productivity stagnation, $M \approx 1$, and Wyoming (WY) exhibits productivity decline, $M < 1$. Moreover, looking at the sources of productivity change, technological progress drives productivity growth, $MTC > 1$, for all states but West Virginia (VW), $MTC < 1$. Efficiency change is in line with productivity change, contributing to its growth in the four states where progress is observed, $MTEC > 1$, and causing productivity decline in the remaining two states, $MTEC < 1$. Indeed, in the case of Wyoming (WY), falling behind the production frontier is the main cause of productivity decline: $MTEC = 0.9115$, while observed technological progress, $MTC > 1.0061$, cannot counterbalance this declining trend.

6. Allocative models: Economic efficiency

Assuming an economic optimizing behavior on the part of the observations, e.g., cost minimization, revenue maximization, or profit maximization, and the corresponding input and output positive prices: $\mathbf{w} \in \mathbb{R}_{++}^m$ and $\mathbf{p} \in \mathbb{R}_{++}^s$, it is possible to measure their economic efficiency and, based on duality theory, decompose it into the technical efficiency and allocative efficiency terms (Farrell 1957).

6.1. Cost efficiency

Since we are concerned with overall efficiency in the input space, we assume that observations minimize production costs. This implies that if observations succeed in using the inputs mix (bundle) resulting in the minimum cost of producing a given output level at the existing market prices, they are cost efficient. Let us denote by $C(\mathbf{y}, \mathbf{w})$ the minimum cost of producing the output level \mathbf{y} given the input price vector \mathbf{w} : $C(\mathbf{y}, \mathbf{w}) = \min \left\{ \sum_{i=1}^m w_i x_i \mid \mathbf{x} \geq X, \lambda \mathbf{y}_o \leq Y \lambda, \lambda \geq \mathbf{0} \right\}$, which considers the input possibility set capable of producing \mathbf{y}_o .¹² For the observed output levels we can calculate minimum cost and the associated optimal quantities of inputs consistent with the production technology by solving the following program:

$$\begin{aligned} \min_{\mathbf{x}, \lambda} \quad & C(\mathbf{y}, \mathbf{w}) = \mathbf{w}\mathbf{x} & (12) \\ \text{subject to} \quad & \\ & \mathbf{x} \geq X\lambda \\ & Y\lambda \geq \mathbf{y}_o \\ & \lambda \geq \mathbf{0}. \end{aligned}$$

Once minimum cost is calculated, *cost efficiency* defines as the ratio of minimum cost to observed cost: $CE = C(\mathbf{y}, \mathbf{w}) / \mathbf{w}\mathbf{x}_o$. Thanks to duality results – Shephard (1953), CE can be decomposed into the technical efficiency measure associated to (1), θ_{CRS}^* , and the residual difference corresponding to the *allocative cost efficiency*: $AE = CE / \theta_{\text{CRS}}^*$. Therefore

¹²For a recent discussion on the properties of the technology when decomposing economic efficiency into technical and allocative efficiencies see Aparicio, Pastor, and Zofio (2015).

allocative efficiency defines as the ratio between minimum cost and production cost at the technically efficient projection of the observation: $AE = C(\mathbf{y}, \mathbf{w}) / \mathbf{w}\hat{\mathbf{x}}_o$ – with $\hat{\mathbf{x}}_o = \theta_{\text{CRS}}^* \mathbf{x}$ and $CE = TE \times AE$. Consequently, if $CE < 1$ and the observation is technically efficient, $\theta_{\text{CRS}}^* = 1$, all cost inefficiency is allocative, while if the projected benchmark uses the right proportions of the optimal input quantities, which we denote by \mathbf{x}^* , then $\hat{\mathbf{x}}_o = \mathbf{x}^*$, and the observation is allocative efficient and $AE = 1$.

The cost efficiency model can be computed in MATLAB using the `deaalloc(X, Y, ...)` function. The parameter `Xprice` with input prices, as a matrix or as a row vector, must be included.¹³

```
load 'DataAgriculture'
X = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y = [LS04, CROPS04, FARMOUT04];
X = X ./ 1000;
Y = Y ./ 1000;
W = [WCAPITAL04, WLAND04, WLABOR04, WINTINP04];
P = [PLS04, PCROPS04, PFARMOUT04];
statenames = STATE_NAME04;
cost = deaalloc(X, Y, 'Xprice', W, 'names', statenames);
deadisp(cost, 'names/eff.T/eff.A/eff.C');
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: allocative-cost
Orientation: io (Input oriented)
Returns to scale: crs (Constant)
```

```
-----
DMU| TechEff| AllocEff| CostEff|
-----
AL| 0.9077| 0.9422| 0.8552|
AR| 0.9888| 0.9291| 0.9187|
AZ| 0.9698| 0.9242| 0.8963|
.
.
WI| 0.8452| 0.7800| 0.6593|
WV| 0.6612| 0.6050| 0.4001|
WY| 0.5911| 0.7665| 0.4531|
-----
```

¹³If input prices are the same for all DMUs the toolbox generates a matrix with the same values. If input prices differ between DMUs a matrix with individual price information can be supplied. In a previous version of the toolbox the economic models in this section were illustrated using an example from Cooper *et al.* (2007, p. 261).

6.2. Revenue efficiency

In an equivalent way, from an output dimension one may be interested in *revenue efficiency*. Now if observations are able to produce the output mix (bundle) resulting in maximum revenue given their input levels at the existing market prices, they are revenue efficient. Let us denote by $R(\mathbf{x}, \mathbf{p})$ the maximum feasible revenue using input levels \mathbf{x} and given the output prices \mathbf{p} : $R(\mathbf{x}, \mathbf{p}) = \max \left\{ \sum_{i=1}^s p_i y_i \mid \mathbf{x}_o \geq X\lambda, \mathbf{y} \leq Y\lambda, \lambda \geq \mathbf{0} \right\}$; i.e., considering the output possibility set producible with \mathbf{x}_o .

In this case, we calculate maximum revenue along with the optimal output quantities by solving the following program:

$$\begin{aligned} \max_{\mathbf{y}, \lambda} \quad & R(\mathbf{x}_o, \mathbf{p}) = \mathbf{p}\mathbf{y} & (13) \\ \text{subject to} \quad & \\ & \mathbf{x}_o \geq X\lambda \\ & Y\lambda \geq \mathbf{y} \\ & \lambda \geq \mathbf{0}. \end{aligned}$$

Revenue efficiency defines as the ratio of observed revenue to maximum revenue: $RE = \mathbf{p}\mathbf{y}_o / R(\mathbf{x}, \mathbf{p})$. Again, duality results from an output perspective allow us to decompose RE into the output technical efficiency measure associated to the inverse of (3), $TE = 1/\phi_{\text{CRS}}^*$, and the residual difference corresponding to the *allocative revenue efficiency*: $AE = RE \times \phi_{\text{CRS}}^*$. In this occasion allocative efficiency defines as the ratio between revenue at the technically efficient projection of the observation and maximum revenue: $AE = \mathbf{p}\hat{\mathbf{y}}_o / R(\mathbf{x}, \mathbf{p})$ – with $\hat{\mathbf{y}}_o = \phi_{\text{CRS}}^* \mathbf{y}_o$, and therefore $RE = TE \times AE$. Now, if $RE < 1$ and the observation is technically efficient, $\phi_{\text{CRS}}^* = 1$, all revenue inefficiency is allocative and the observation is unable to produce the right output mix. Contrarily, if the projected benchmark produces the right mix of optimal output quantities given their prices, which we denote by \mathbf{y}^* , then $\hat{\mathbf{y}}_o = \mathbf{y}^*$, and the observation is allocative efficient: $AE = 1$.

The revenue efficiency model can be computed in MATLAB using the `deaalloc(X, Y, ...)` function. The parameter `Yprice` with the output prices, as a matrix or as a row vector, must be included.

```
revenue = deaalloc(X, Y, 'Yprice', P, 'names', statenames);
deadisp(revenue, 'names/eff.T/eff.A/eff.R');
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: allocative-revenue
Orientation: oo (Output oriented)
Returns to scale: crs (Constant)
```

```
-----
```

DMU	TechEff	AllocEff	RevEff
AL	0.9077	0.8922	0.8098
AR	0.9888	0.9611	0.9504
AZ	0.9698	0.9718	0.9425
.			
.			
WI	0.8452	0.8474	0.7163
WV	0.6612	0.7859	0.5197
WY	0.5911	0.9536	0.5637

The cost and revenue efficiency models show that none of the selected states minimize the cost of production or maximize the revenue associated to agricultural activities. None of them present a score equal to one, and indeed they are far away from it except Arkansas (AR), with $CE = 0.9187$ and $RE = 0.9504$, and Arizona (AZ), with $CE = 0.8963$ and $RE = 0.9425$. These two states present relatively high values of technical efficiency and allocative efficiency, but are inefficient in both dimensions, either from a cost or revenue perspective. The states that are cost efficient (not reported here) are Delaware (DE), Florida (FL), Georgia (GA), Iowa (IA), Idaho (ID) and Illinois (IL) – and similarly from a revenue efficiency perspective. Therefore all of them are both technical and allocative efficient. On the contrary West Virginia (WV) and Wyoming (WY) have rather low cost and revenue efficiency scores. In general, both sources of inefficiency are balanced, except in the case of Wyoming (WY), whose technical inefficiency is mainly responsible for all revenue inefficiency: $TE = 0.5911$ vs. $AE = 0.9536$.

6.3. Profit efficiency: The directional approach

Profit efficiency allows studying the economic behavior of observations as profit maximizers. In this case, if observations are able to maximize the difference between revenue and cost given market prices by producing and using the right quantities of outputs and inputs respectively, they are profit efficient. The profit function defines as $\Pi(\mathbf{w}, \mathbf{p}) = \max \left\{ \sum_{i=1}^s p_i y_i - \sum_{i=1}^m w_i x_i \mid \mathbf{x} \geq X\lambda, \mathbf{y} \leq Y\lambda, \mathbf{e}\lambda = \mathbf{1}, \lambda \geq \mathbf{0} \right\}$. Calculating maximum profit along with the optimal output and input quantities and requires solving:

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{y}, \lambda} \quad & \Pi(\mathbf{w}, \mathbf{p}) = \mathbf{p}\mathbf{y} - \mathbf{w}\mathbf{x} & (14) \\ \text{subject to} \quad & \\ & \mathbf{x} \geq X\lambda = \mathbf{x} \\ & \mathbf{y} \leq Y\lambda = \mathbf{y} \\ & \mathbf{e}\lambda = \mathbf{1} \\ & \lambda \geq \mathbf{0}. \end{aligned}$$

Profit efficiency defines as the difference between maximum profit and observed profit. As duality results concerning its decomposition into technical and allocative profit efficiencies rely

on the directional model (5), we can use the directional vector to define a normalized profit efficiency measure that is homogeneous of degree one in prices – invariant to proportional price changes; i.e., $PE = (\Pi(\mathbf{w}, \mathbf{p}) - (\mathbf{p}\mathbf{y}_o - \mathbf{w}\mathbf{x}_o)) / (\mathbf{p}\mathbf{g}_y^+ + \mathbf{w}\mathbf{g}_x^-)$ – see Chambers, Chung, and Färe (1998) and Zofío *et al.* (2013) for more recent proposals. Subsequently, profit efficiency can be decomposed into the directional technical efficiency measure associated to (5) under variable returns to scale, $TE = \beta_{\text{VRS}}^*$, and the residual difference corresponding to the *allocative profit efficiency* term: $AE = PE - \beta_{\text{VRS}}^*$. Therefore, allocative efficiency defines as the difference between maximum profit and profit at the technically efficient projection of the observation: $AE = (\Pi(\mathbf{w}, \mathbf{p}) - (\mathbf{p}\hat{\mathbf{y}}_o - \mathbf{w}\hat{\mathbf{x}}_o)) / (\mathbf{p}\mathbf{g}_y^+ + \mathbf{w}\mathbf{g}_x^-)$, with $(\hat{\mathbf{y}}_o, \hat{\mathbf{x}}_o) = (\mathbf{y}_o + \beta_{\text{VRS}}^* \mathbf{g}_y^+, \mathbf{x}_o - \beta_{\text{VRS}}^* \mathbf{g}_x^-)$, and $PE = TE + AE$. Now, if $PE > 0$ and the observation is technically efficient, $\beta_{\text{VRS}}^* = 0$, all profit inefficiency is allocative, and the observation is unable to produce with the optimal combination of outputs and inputs. Contrarily, if the benchmark observation supplies and demands the profit maximizing quantities of optimal outputs and inputs, which we denote by \mathbf{y}^* and \mathbf{x}^* , then $(\hat{\mathbf{y}}_o, \hat{\mathbf{x}}_o) = (\mathbf{y}^*, \mathbf{x}^*)$, and the observation is allocative efficient: $AE = 0$.

The profit inefficiency model can be computed in MATLAB using the `deaalloc(X, Y, ...)` function. Both the parameters `Xprice` and `Yprice` with inputs and outputs prices must be included. The input and output directions are specified in the `Gx` and `Gy` parameters. If omitted, `X` and `Y` will be used for `Gx` and `Gy` respectively.

```
profit = deaalloc(X, Y, 'Xprice', W, 'Yprice', P, 'names', statenames);
deadisp(profit, 'names/eff.T/eff.A/eff.P');
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
```

```
Inputs: 4      Outputs: 3
```

```
Model: allocative-profit
```

```
Orientation: ddf (Directional distance function)
```

```
Returns to scale: vrs (Variable)
```

```
-----
DMU| TechEff| AllocEff| ProfEff|
-----|-----|-----|-----
AL|  0.0161|  1.3321|  1.3481|
AR|  0.0055|  0.5049|  0.5104|
AZ|  0.0141|  2.6562|  2.6703|
.
.
WI|  0.0000|  0.3881|  0.3881|
WV|  0.1932|  6.2741|  6.4673|
WY|  0.2570|  6.1169|  6.3739|
-----
```

6.4. Profit efficiency: The weighted additive approach

An alternative decomposition of *profit efficiency* based on the weighted additive measure and duality results is possible (Cooper, Pastor, Aparicio, and Borras 2011). In contrast to the

radial and directional approaches, the technical efficiency component of the decomposition accounts for all inefficiencies (i.e., it includes individual input and output slacks).

Profit efficiency defines in this case as the difference between maximum profit and observed profit, normalized by the minimum among the ratios of market prices to weights; i.e., $PE = (\Pi(\mathbf{w}, \mathbf{p}) - (\mathbf{p}\mathbf{y}_o - \mathbf{w}\mathbf{x}_o)) / \min\{(p_1/\rho_1^+), \dots, (p_m/\rho_s^+), (w_1/\rho_1^-), \dots, (\rho_1/\rho_m^-)\}$. This normalization ensures once again that the profit efficiency is homogeneous of degree one in prices, rendering it units invariant.

Once the weighted additive measure (7) is calculated, profit efficiency can be decomposed into technical efficiency, $TE = \omega_{\text{VRS}}^*$, and the residual difference corresponding to the *allocative profit efficiency* term: $AE = PE - \omega_{\text{VRS}}^*$. Consequently, allocative efficiency defines as the difference between maximum profit and profit at the technically efficient projection of the observation: $AE = (\Pi(\mathbf{w}, \mathbf{p}) - (\mathbf{p}\hat{\mathbf{y}}_o - \mathbf{w}\hat{\mathbf{x}}_o)) / \min\{(p_1/\rho_1^+), \dots, (p_m/\rho_s^+), (w_1/\rho_1^-), \dots, (\rho_1/\rho_m^-)\}$, with $(\hat{\mathbf{y}}_o, \hat{\mathbf{x}}_o) = (\mathbf{y}_o + \rho_{\mathbf{y}}^+ \mathbf{s}^+, \mathbf{x}_o - \rho_{\mathbf{x}}^- \mathbf{s}^-)$, and once again $PE = TE + AE$. Now, if $PE > 0$ and the observation is technically efficient, $\omega_{\text{VRS}}^* = 0$, all profit inefficiency is allocative, and the observation is unable to produce with the optimal combination of outputs and inputs. Contrarily, if the observation supplies and demands the profit maximizing quantities of outputs and inputs, $(\hat{\mathbf{y}}_o, \hat{\mathbf{x}}_o) = (\mathbf{y}^*, \mathbf{x}^*)$, it is allocative efficient: $AE = 0$.

In this section we have focused on the decomposition of profit efficiency, while the cost and revenue counterparts presented in the previous sections can be easily obtained by setting outputs and inputs weights equal to zero, and using cost and revenue as support functions, respectively.

The weighted additive profit efficiency model is computed with the `deaadditprofit(X, Y, ...)` function. Both the parameters `Xprice` and `Yprice` with inputs and outputs prices must be included. The returns to scale assumption, `rts` parameter, is set to `vrs` (variable returns to scale). Inputs and outputs weights are specified in the `rhoX` and `rhoY` parameters. As before, the default weights correspond to the measure of inefficiency proportions (MIP) model if not included.

```
addprofit = deaadditprofit(X, Y, 'rts', 'vrs', 'Xprice', W, 'Yprice', P, ...
  'names', statenames);
deadisp(addprofit, 'names/eff.T/eff.A/eff.P');
```

Data Envelopment Analysis (DEA)

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: additive-profit
Orientation: none
Returns to scale: vrs (Variable)
```

```
-----
DMU| TechEff|  AllocEff|  ProfEff|
-----
AL|  1.2677| 1311.5991| 1312.8669|
AR|  0.6874|  811.3089|  811.9963|
AZ|  1.1851| 1809.7652| 1810.9504|
```

```

.
.
WI| 0.0000| 847.1218| 847.1218|
WV| 4.7765| 1217.5219| 1222.2983|
WY| 4.9848| 2001.9177| 2006.9024|
-----

```

Looking at the profit efficiency results, either directional or additive, the state that maximizes profit by not incurring in technical and allocative inefficiencies is California (CA). Consequently this state defines the benchmark against which the remaining states compare themselves. In the profit efficiency models the larger the value of the score, the greater the inefficiency. In the results for the directional distance function model West Virginia (WV) and Wyoming (WY) rank again amongst the worse states, with efficiency scores equal to $PE = 6.4673$ and $PE = 6.3739$, respectively. Arkansas (AR) and Wisconsin (WI) present values that are relatively low, $PE = 0.5104$ and $PE = 0.3881$. As this latter state is efficient technically, $TE = 0$, all inefficiency is allocative, $PE = AE = 0.3881$, which is due to a wrong choice of input and output mixes; i.e., farms are not demanding and supplying the optimal amounts of inputs and outputs given their prices. The results are similar in qualitative terms if the weighted additive approach is used to measure profit efficiency, although the values are substantially different as a result of the change in the objective function. Nevertheless, we can observe that, once again, the change of model does not have a neutral effect on the efficiency rankings. For example, based on the results for the directional distance function, Alabama (AL) is ranked ahead of West Virginia (WV): $PE = 1.3481$ vs. $PE = 6.4673$, but this is reversed in the weighted additive model: $PE = 1312.8669$ vs. $PE = 1222.2983$. Therefore, considering an efficiency measure that takes slacks into account by projecting observations to the strongly efficient frontier, rather than to the weakly efficient frontier, where slacks may still exist, may have relevant implications.

7. Undesirable outputs

Along with desirable and market oriented products, observations may produce undesirable or detrimental outputs as byproducts, such as pollutants or hazardous wastes from an environmental perspective. As a result, efficiency and productivity measures that do not take into account the asymmetry between both types of production: desirable and undesirable, will result in biased assessments of performance and erroneous calculations; e.g., when assessing environmental performance and making recommendations to improve technical efficiency, one seeks increments in desirable production *but* reductions in undesirable outputs. To incorporate undesirable outputs into the efficiency and productivity change models, we rely on the directional measures (5) that treat both sets of outputs differently. This requires a redefinition of the production technology where the initial vector of $i = 1, 2, \dots, s$ outputs $\mathbf{y} \in \mathbb{R}_{++}^s$ is partitioned into desirable and undesirable production, i.e., $\mathbf{y} = (\mathbf{y}^d, \mathbf{y}^u)$, with $\mathbf{y}^d \in \mathbb{R}_{++}^q$ and $\mathbf{y}^u \in \mathbb{R}_{++}^r$, respectively. This translates into the corresponding reference technology $P_{\text{CRS}} = \left\{ (\mathbf{x}, \mathbf{y}^d, \mathbf{y}^u) \mid \mathbf{x} \geq X\lambda, \mathbf{y}^d \leq Y\lambda, \mathbf{y}^u = Y\lambda, \lambda \geq \mathbf{0} \right\}$, characterizing undesirable outputs as weakly disposable (Cooper *et al.* 2007, Chapter 12). As the directional efficiency measure resulting in increases in desirable outputs and reductions in undesirable outputs relative to the same amount of inputs is used in the following section to define the Malmquist-Luenberger productivity measure, we rely on Aparicio, Pastor, and Zofio (2013)

to calculate it, thereby preventing the inconsistencies of the original approach introduced by Chung, Färe, and Grosskopf (1997). In this case, the directional efficiency measure projecting observation $(\mathbf{x}_o, \mathbf{y}_o^d, \mathbf{y}_o^u)$ along the pre-assigned direction corresponding to the output vector $\mathbf{g}_y = (\mathbf{y}^d, \mathbf{y}^u) \neq \mathbf{0}_{m+s}$, corresponds to the solution of the following program:

$$\begin{aligned} & \max_{\beta, \lambda} && \beta && (15) \\ & \text{subject to} && && \\ & && X\lambda \leq \mathbf{x}_o && \\ & && Y^d\lambda \geq \mathbf{y}_o^d + \beta\mathbf{y}_o^d && \\ & && Y^u\lambda \leq \mathbf{y}_o^u - \beta\mathbf{y}_o^u && \\ & && \max\{\mathbf{y}_i^u\} \geq \mathbf{y}_o^u - \beta\mathbf{y}_o^u && \\ & && \lambda \geq \mathbf{0}. && \end{aligned}$$

Again, the optimal solution corresponds to β_{CRS}^* , and if $\beta_{\text{CRS}}^* = 0$, with $\lambda_o = 1, \lambda_j = 0$ ($j \neq o$), the observation is directional efficient. Otherwise, $\beta_{\text{CRS}}^* > 0$ signals inefficiency and $(\lambda X, \lambda Y^d, \lambda Y^u)$ outperforms $(\mathbf{x}_o, \mathbf{y}_o^d, \mathbf{y}_o^u)$. It is possible also to calculate any additional slacks associated to further increases of individual outputs as well as reductions in individual inputs and undesirable outputs.

The undesirable outputs model is computed in MATLAB using the `deaund(X, Y, Yu, ...)` function, where `Y` is now reserved for the matrix of desirable outputs – i.e., from a computational perspective we drop the superscript d –, and `Yu` is the matrix of undesirable outputs.

```
load 'DataAgriculture'
X = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y = [LS04, CROPS04, FARMOU04, ];
Yu = PESTEXPNCASES04;
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
undesirable = deaund(X0, Y0, Yu0, 'names', statenames);
deadisp(undesirable);
```

Data Envelopment Analysis (DEA)

```
DMUs: 48
Inputs: 4      Outputs: 3      Undesirable: 1
Model: directional-undesirable
Orientation: ddf (Directional distance function)
Returns to scale: crs (Constant)
```

```
-----
DMU|      Beta|      slackYu|
-----
AL|  0.0743|  0.0000|
AR|  0.0000|  0.0000|
```

```

AZ| 0.0311| 1042.0088|
.
.
WI| 0.0837| 0.0000|
WV| 0.4770| 0.0000|
WY| 0.1469| 0.0000|
-----

```

The environmental or eco-efficiency model based on the directional distance function takes into consideration the possibility of reducing undesirable outputs when assessing productive performance; in this case the possibility of reducing the number of pesticide exposures among the population as a results of the agricultural activity. It is therefore relevant to compare the results of this model to those obtained for the standard directional model (5). For example, focusing on Arkansas (AR), which is almost efficient in the variable returns to scale model ignoring undesirable outputs $\beta_{\text{VRS}}^* = 0.0055$, it becomes fully efficient when they are taken into consideration (even under the chosen constant returns to scale model, which implies that it is also efficient if variable returns to scale were considered). Moreover, this state defines the weakly efficient production frontier, since no slacks are observed in that dimension. This is not the case however for the rest of the selected states, which are environmentally inefficient. Also, it is worth remarking that Arizona (AZ), besides reducing inputs and increasing outputs in the magnitude represented by the efficiency score times its observed quantities, could further reduce the number of pesticide exposures in 1,042 cases, as measured by its associated slack, showing that additional gains in efficiency could obtained in that individual dimension.

8. Productivity change: The Malmquist-Luenberger index

Finally, it is possible to define the undesirable outputs productivity counterpart to the Malmquist index. Chung *et al.* (1997) introduced the Malmquist-Luenberger (*ML*) productivity index measuring the change in productivity of the observation under evaluation by comparing its relative efficiency with respect to reference technologies corresponding to two different time periods. As its radial counterpart, the standard *ML* uses the directional efficiency score only – disregarding second stage slacks – and requires calculation of the mix period efficiency of $(\mathbf{x}_o^t, \mathbf{y}_o^{t,d}, \mathbf{y}_o^{t,u})$ observed in two periods $t = 1, 2$. We denote both scores by $\beta_{\text{CRS}}^{1,1}$ and $\beta_{\text{CRS}}^{2,1}$, where once again the first superscript refers to the time period of the observation and the second one to that of the reference technology. While $\beta_{\text{CRS}}^{1,1}$ is the solution to program (15), the intertemporal score $\beta_{\text{CRS}}^{2,1}$ corresponds to the following program that evaluates period 2 observation $(\mathbf{x}_o^2, \mathbf{y}_o^{2,d}, \mathbf{y}_o^{2,u})$ with respect to period 1 technology:

$$\begin{aligned}
 & \max_{\beta, \lambda} && \beta && (16) \\
 & \text{subject to} && && \\
 & && X^1 \lambda \leq \mathbf{x}_o^2 && \\
 & && Y^{1,d} \lambda \geq \mathbf{y}_o^{2,d} + \beta \mathbf{y}_o^{2,d} && \\
 & && Y^{1,u} \lambda \leq \mathbf{y}_o^{2,u} - \beta \mathbf{y}_o^{2,u} && \\
 & && \max \{\mathbf{y}_i^{t,u}\} \geq \mathbf{y}_o^{2,u} - \beta \mathbf{y}_o^{2,u} && \\
 & && \lambda \geq \mathbf{0}, &&
 \end{aligned}$$

where $\max\{\mathbf{y}_i^{t,u}\}$ is the maximum observed amount of each undesirable output in all time periods (Aparicio *et al.* 2013). Correspondingly, the score $\beta_{\text{CRS}}^{1,2}$, using the second period technology as reference can be calculated solving an equivalent program that evaluates $(\mathbf{x}_o^1, \mathbf{y}_o^{1,d}, \mathbf{y}_o^{1,u})$ with respect to the reference technology $(X^2, Y^{2,d}, Y^{2,u})$.

Once these scores are calculated the ML indices define as follows: $ML_1 = (1 + \beta_{\text{CRS}}^{1,1}) / (1 + \beta_{\text{CRS}}^{2,1})$ and $ML_2 = (1 + \beta_{\text{CRS}}^{1,2}) / (1 + \beta_{\text{CRS}}^{2,2})$. If $ML > 1$ efficiency increases and the evaluated unit is capable of producing more desirable output with less undesirable production, while if $ML = 1$ productivity remains unchanged, and $ML < 1$ signals productivity decline. Following Chung *et al.* (1997) the index can be decomposed into efficiency change and technical change, which have the same interpretation than their previous Malmquist counterparts. The change in technical efficiency defines now as $MLTEC = (1 + \beta_{\text{CRS}}^{1,1}) / (1 + \beta_{\text{CRS}}^{2,2})$, while the *frontier-shift* effects corresponding to technical change are $MLTC_1 = (1 + \beta_{\text{CRS}}^{2,2}) / (1 + \beta_{\text{CRS}}^{2,1})$ and $MLTC_2 = (1 + \beta_{\text{CRS}}^{1,2}) / (1 + \beta_{\text{CRS}}^{1,1})$. Again, if $MLTEC > 1$ or $MLTC > 1$ productivity change responds to both technical efficiency gains and technical change improvements (technical progress); while $MLTEC < 1$ or $MLTC < 1$ imply lower productivity with greater inefficiency and technical regress. Finally, unitary values signal that the technical efficiency and the reference frontier remain unchanged. Therefore the decomposition of productivity change defines as $ML_1 = MLTEC \times MLTC_1 = (1 + \beta_{\text{CRS}}^{1,1}) / (1 + \beta_{\text{CRS}}^{2,2}) \times (1 + \beta_{\text{CRS}}^{2,2}) / (1 + \beta_{\text{CRS}}^{2,1})$ – and similarly for ML_2 . Given a sequence of years, the toolbox calculates the year-to-year variation of ML_1 and the geometric mean of both indices: $ML = MLTEC \times MLTC = (1 + \beta_{\text{CRS}}^{1,1}) / (1 + \beta_{\text{CRS}}^{2,2}) \times ((1 + \beta_{\text{CRS}}^{2,2}) / (1 + \beta_{\text{CRS}}^{2,1}) \times ((1 + \beta_{\text{CRS}}^{1,2}) / (1 + \beta_{\text{CRS}}^{1,1})))^{(1/2)}$.

The Malmquist-Luenberger indices are computed in MATLAB calling the `deamalm1uen(X, Y, Yu, ...)` function. The parameters X, Y and Yu must be 3D arrays, with the third dimension corresponding to different time periods. By default, Malmquist-Luenberger indices are computed on a year-to-year basis taking the previous year as reference period. The aggregate Malmquist-Luenberger index taking the first period as the base period can be computed by setting the `fixbaset` parameter to 1. By default, Malmquist-Luenberger indices are computed as the geometric mean of ML_1 and ML_2 . However, for simplicity, the indices can be computed taking the first period as the base for technical change, ML_1 , by setting the parameter `geomean` to 0.

```
load 'DataAgriculture'
X00 = [CAPITAL00, LAND00, LABOR00, INTINP00];
Y00 = [LS00, CROPS00, FARMOUT00];
Yu00 = [PESTEXPNCASES00];
X04 = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y04 = [LS04, CROPS04, FARMOUT04];
Yu04 = [PESTEXPNCASES04];
X = X00;
X(:, :, 2) = X04;
Y = Y00;
Y(:, :, 2) = Y04;
Yu = Yu00;
Yu(:, :, 2) = Yu04;
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
```

```
ml = deamalm1uen(X, Y, Yu, 'names', statenames);
deadisp(ml);
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3      Undesirable: 1
Model: directional-malmquist-luenberger
Orientation: ddf (Directional distance function)
Returns to scale: crs (Constant)
```

```
Malmquist-Luenberger:
Geometric mean is computed
Base period is previous period
```

```
-----
DMU|      ML|    MLTEC|    MLTC|
-----
AL|  1.0384|  0.9897|  1.0492|
AR|  1.0389|  1.0000|  1.0389|
AZ|  0.9948|  0.9769|  1.0183|
.
.
WI|  0.9172|  0.9868|  0.9295|
WV|  1.0161|  1.0562|  0.9620|
WY|  0.8869|  1.1616|  0.7635|
-----
```

ML: Malmquist-Luenberger. MLTEC: Technical Efficiency Change.
MLTC: Technical Change.

Just as the undesirable output dimension can be incorporated to the directional efficiency cross-sectional model, the same can be done from a panel data perspective, redefining the notion of productivity to account for undesirable outputs. For this reason it is relevant to compare the productivity trends previously discussed in Section 5 for the Malmquist index (M) with those reported here for the Malmquist-Luenberger indices (ML). In the latter case, the inclusion of the number of pesticide exposures as undesirable output may change the nature of productivity change. A good example is Wisconsin (WI), which shows productivity growth according to the Malmquist index, $M = 1.0469$, but exhibits productivity decline in the case of the Malmquist-Luenberger index $ML = 0.9172$. Moreover, compared with the differences observed for other states, this gap between both indices is remarkable, showing that if one is willing to adopt a definition that accounts for the externalities of production processes, the results can vary dramatically. Also, in the Malmquist case, while Arizona (AR) leads the ranking of productivity growth by far, with Wisconsin (WI) in second place, this is not the case for the ML index, whose value $ML = 1.0389$ is closely followed by that of Alabama (AL), $ML = 1.0384$, with Wisconsin (WI) falling from the second to the fifth place. As for the sources of productivity change, while technological progress, $MTC > 1$, was the main driver of productivity change in the Malmquist case for all but one of the selected states

(West Virginia, WV), in the Malmquist-Luenberger results this is only observed for the first three states, $MLTC > 1$. Similarly, the values for the technical efficiency change component are different depending on the definition of productivity. These results simply show that productivity change is much lower when undesirable outputs are taken into account, and therefore its components also show smaller values. These trends suggest that the effort exerted in reducing pesticide run-offs, so as to internalize the damage that they produce on humans, is lower than that aimed at increasing marketable outputs and reducing costly inputs. Hence the difference in productivity growth in favor of the Malmquist values.

9. Bootstrapping DEA estimators

The efficiency scores and productivity indices calculated in the previous sections can be considered as estimates as their values are subject to uncertainty due to sampling variation. Indeed, because of the uncertainty in many real situations, input and/or output data could be in stochastic form (e.g., historical data) or arising from qualitative judgments (e.g., decision maker's preferences).¹⁴ Simar and Wilson (1998) introduce bootstrap methods that, based on resampling, provide estimation bias, confidence intervals and allow hypotheses testing. This toolbox implements the algorithms presented by these authors following Bogetoft and Otto (2011), who also discuss other non-parametric and asymptotic tests. We implement the tests that allow determining the significance of the models covered by the toolbox: the assumption about returns to scale, i.e., whether constant and variable returns to scale scores are significantly different (Simar and Wilson 2002), and the existence of productivity change over time, i.e., whether the Malmquist index and its components are different from unity (Simar and Wilson 1999).

The algorithm implemented in the toolbox is the following:

1. Selection of B independent bootstrap samples – drawn from the original dataset with replacements.
2. Calculate an initial estimate for the efficiency score of each DMU with respect to each bootstrapped sample and smooth their distribution by perturbing them with a random noise generated from a kernel density function with scale given with bandwidth h ;
3. Correct the original estimates for the mean and variance of the smoothed values;
4. Obtain a second set of bootstrapped samples generating inefficient DMUs, inside the DEA attainable set and conditional on the original input – or output – mix;

¹⁴For recent reviews on alternative methods to overcome the deterministic nature of DEA we refer the reader to Olesen and Petersen (2016) and Hatami-Marbini *et al.* (2011), who review the stochastic and fuzzy approaches to DEA, respectively. Stochastic DEA complements the usual technological axioms characterizing the deterministic production technology with a set of distributional assumptions that provide a statistical foundation for the model. The former authors discuss three recent extensions of deterministic DEA: (i) deviations from the standard frontier are modeled as stochastic variables, (ii) random noise in terms of measurement errors, sample noise, and specification errors is made an integral part of the model, and (iii) the frontier is stochastic as is the underlying production possibility set (PPS). These approaches allow for an estimation of stochastic inefficiency and for statistical inference, while maintaining an axiomatic foundation. The latter authors also review recent advances in fuzzy DEA by presenting a taxonomy of methods. In particular they propose a classification scheme based on the following typology: the tolerance approach, the α -level based approach, the fuzzy ranking approach, and the possibility approach. The interested reader is also referred to empirical implementations of these methods by Kao and Liu (2009) and Costantino *et al.* (2012b).

5. Repeat the process, estimate the efficiency scores for each original DMU with respect to that second set, so as to obtain a set of B bootstrap estimates; and finally,
6. Based on this distribution calculate the threshold values that truncate it according to the pre-determined significance value α , so as to determine the confidence intervals for the efficiency score of each DMU.

In addition, the bootstrapped scores can be used to obtain an estimate of the bias of the true efficiency value, and thereby a bias-corrected estimator.¹⁵

The radial bootstrap model can be computed in MATLAB using the `deaboot(X, Y, ...)` function with the `orient` parameter set to the desired orientation (`io` for input oriented or `oo` for output oriented). The returns to scale assumption can be specified by setting the `rts` parameter to `crs` (constant returns to scale; default) or `vrs` (variable returns to scale). The number of bootstrap replication are set in the `nreps` parameter (which defaults to 200), and the significance level in the `alpha` parameter (0.05).¹⁶

```
load 'DataAgriculture'
X = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y = [LS04, CROPS04, FARMOUTO4];
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
rng(1234567, 'twister'); % Set seed for reproducibility
io_b = deaboot(X, Y, 'orient', 'io', 'nreps', 200, 'alpha', 0.05, ...
    'names', statenames);
deadisp(io_b);
```

```
-----
Data Envelopment Analysis (DEA)
```

```
DMUs: 48
Inputs: 4      Outputs: 3
Model: radial-bootstrap
Orientation: io (Input oriented)
Returns to scale: crs (Constant)
Bootstrap replications: 200
Significance level: 0.05
```

```
-----
DMU|      eff| effboot|  effCI1|  effCI2|
-----
```

¹⁵A similar procedure is implemented to determine the significance of the Malmquist indices – see [Simar and Wilson \(1999\)](#) for the specific algorithm.

¹⁶Calculation of the bootstrapped radially input model with the default options takes around 26 seconds on a PC with an i7-8700K CPU (3.70GHz, 3696 Mhz), 32 GB DDR4 (2666 MH) of RAM, and making use of the parallel computing toolbox connected to 6 workers. Computing time rises to 37 seconds when running the returns to scale test – since both the constant and variables returns to scale need to be calculated, and up to 70 seconds for the bootstrapped Malmquist (including the calculation of several cross-period efficiency scores). The rest of the models in the toolbox are solved within one or two seconds.

AL	0.9077	0.8471	0.7818	0.9042
AR	0.9888	0.9456	0.8988	0.9864
AZ	0.9698	0.9072	0.8117	0.9680
.				
.				
WI	0.8452	0.8070	0.7694	0.8410
WV	0.6612	0.6270	0.5906	0.6593
WY	0.5911	0.5607	0.5319	0.5886

The returns to scale test is computed by calling the `deatestrts(X, Y, ...)` function with the `orient` parameter set to the desired orientation (`io` for input oriented or `oo` for output oriented). The number of bootstrap replication are set in the `nreps` parameter (which defaults to 200), and the significance level in the `alpha` parameter (0.05). Results of the test can be displayed on screen by setting the parameter `disp` to 1.

```
rng(1234567, 'twister'); % Set seed for reproducibility
deatestrts(X, Y, 'orient', 'io', 'nreps', 200, 'alpha', 0.05, 'disp', 1);
```

DEA Test of RTS

H0: Globally CRS

H1: VRS

Bootstrap replications: 200

Significance level: 0.05

S statistic: 0.9813

Critical value: 0.9426

p-value: 0.5800

To compute bootstrapped Malmquist indices in MATLAB one calls the `dealmboot(X, Y, ...)` function with the `orient` parameter set to input orientation (`io`). The parameters `X` and `Y` must be 3D arrays, with the third dimension corresponding to different time periods. The number of bootstrap replication are set in the `nreps` parameter (which defaults to 200), and the significance level in the `alpha` parameter (0.05). Again, by default, Malmquist indices are computed as the geometric mean of M_1 and M_2 . However, for simplicity, the indices can be computed taking the first period as the base for technical change, M_1 , by setting the parameter `geommean` to 0.

```
load 'DataAgriculture'
X00 = [CAPITAL00, LAND00, LABOR00, INTINP00];
Y00 = [LS00, CROPS00, FARMOUT00];
X04 = [CAPITAL04, LAND04, LABOR04, INTINP04];
Y04 = [LS04, CROPS04, FARMOUT04];
X = X00;
X(:, :, 2) = X04;
```

```

Y = Y00;
Y(:, :, 2) = Y04;
X = X ./ 1000;
Y = Y ./ 1000;
statenames = STATE_NAME04;
rng(1234567, 'twister'); % Set seed for reproducibility
malmquist = deamalmboot(X, Y, 'orient', 'io', 'nreps', 200, 'alpha', 0.05);
deadisp(malmquist)

```

```
-----
Data Envelopment Analysis (DEA)

```

```

DMUs: 48
Inputs: 4      Outputs: 3
Model: radial-malmquist-bootstrap
Orientation: io (Input oriented)
Returns to scale: crs (Constant)
Bootstrap replications: 200
Significance level: 0.05

```

```

Malmquist:
Geometric mean is computed
Base period is previous period

```

```

-----
DMU|      M|   Mboot|   McLow|   McUpp|
-----
AL|  1.0390|  1.0232|  0.9566|  1.0679|
AR|  1.1543|  1.0967|  1.0278|  1.1771|
AZ|  0.9948|  1.0062|  0.9411|  1.0740|
.
.
WI|  1.0469|  1.0201|  0.9569|  1.0636|
WV|  1.0282|  0.9906|  0.9066|  1.0416|
WY|  0.9170|  0.9499|  0.8866|  0.9903|
-----

```

M = Malmquist. Mboot = Bootstrapped Malmquist.

McLow = Lower confidence interval. McUpp: Upper confidence interval.

Bootstrapping allows to test whether the standard efficiency scores, the existence of constant returns to scale, and productivity change are statistically significant or not. Focusing first on the bootstrapped scores calculated for the radially oriented input measure in Section 3.1, one sees that at the 5% default significance level, all original scores exceed the upper confidence interval threshold, and therefore none of them are reliable estimates of efficiency change. Taking Alabama (AL) as example, its standard score from program (1), $\theta_{CRS}^* = 0.9077$, is above the confidence interval ($effCI2 = 0.9042$), while the bootstrapped index, $effboot = 0.8471$, falls within the confidence interval defined by the lower and upper bounds. Consequently, bias corrected scores provide a robust and statistically sound alternative. Note however

that the ranking among the selected states does not change for this particular database and default settings, with Arkansas (AR) and Wyoming (WY) ranking in the first and sixth place, respectively. Regarding returns to scale, for the default settings the null hypotheses of constant returns to scale cannot be rejected, suggesting that these property should be considered when solving the different models. Turning now to the standard and bootstrapped Malmquist indices, all of them fall within the confidence interval defined by the lower and upper levels, which also comprise the unitary value, except in the case of Wyoming (WY), whose bounds are $McLow = 0.8866$ and $McUpp = 0.9903$. Therefore, in this particular case, it can be concluded that the bootstrapped value equal to $Mboot = 0.9499$, signaling productivity decline, is indeed different from one, which allows rejecting the hypothesis that productivity remains unchanged.

10. Advanced options, displaying and exporting results

10.1. Advanced optimization options

As recommended by [The MathWorks, Inc. \(2017\)](#), by default all DEA programs are solved using the dual-simplex algorithm, with a constraint tolerance of $1e-7$ and an optimality tolerance of $1e-10$. However, if needed, optimization options can be changed by passing an ‘`optimoptions`’ structure to the `optimopts` optional parameter.¹⁷ As an example, if we want to solve the input oriented DEA program using the interior-point algorithm:

```
opts = optimoptions('linprog', 'display', 'off', 'Algorithm', ...
    'interior-point');
io_ip = dea(X, Y, 'orient', 'io', 'names', statenames, 'secondstep', 0, ...
    'optimopts', opts);
```

10.2. Changing the reference set

Sometimes we need to evaluate some DMUs with respect to a reference set that differs from the observed input and output data. By default, all functions solve all models using the data corresponding to X , Y and Yu – the observed DMUs, which also represent the reference set. However, we can change the data corresponding to the DMU to be evaluated with the `Xeval`, `Yeval` and `Yueval` optional parameters.¹⁸

As an example, using data from Section 3, if we want to evaluate the first DMU with respect to a reference set including all the other DMUs but not itself:

```
Xref = X(2:end, :);
Yref = Y(2:end, :);
Xeval = X(1, :);
Yeval = Y(1, :);
```

¹⁷See the official MATLAB documentation to create an optimization options structure at <https://www.mathworks.com/help/optim/ug/optimoptions.html>.

¹⁸This is the procedure used internally by the toolbox to compute super-efficiency models and the productivity indices.

```
io1 = dea(Xref, Yref, 'orient', 'io', 'Xeval', Xeval, 'Yeval', Yeval);
disp(io1.eff);
```

0.9077

10.3. Custom display

When calling the `deadisp(out, dispstr)` function after computing a DEA model, appropriate information depending on the estimated model will be displayed on the screen. This setting can be changed to display the desired information by specifying in the `deadisp` function the string `dispstr` (display string) as a second parameter.

The default `dispstr` after using the `dea` function is `names/X/Y/eff/slack.X/slack.Y`. Each of the fields to be displayed in the output table must be separated with a `/` including the names corresponding to the field names of the ‘`deaout`’ structure. The available fields are presented in Table 1.

10.4. Exporting results

Results of the computed DEA models can be easily exported to diverse file formats for posterior analysis and sharing. First the ‘`deaout`’ structure should be converted to a MATLAB ‘`table`’ data type using the `dea2table(out, dispstr)` function using the desired `dispstr`.¹⁹

```
io = dea(X, Y, 'orient', 'io', 'names', statenames);
T = dea2table(io);
```

Then, the table can be exported using the MATLAB function `writetable`.²⁰

```
writetable(T, 'ioresults.csv');
```

11. Conclusions

The **Data Envelopment Analysis Toolbox** covers a wide variety of models calculating efficiency and productivity measures in an organized environment for MATLAB. The models implemented correspond to the classic radially oriented, the directional model and the additive formulation. Both constant and variable returns to scale technical efficiency measures are calculated, which allows the calculation of scale efficiency. The economic performance of firms in terms of technical and allocative criteria is also presented, along with efficiency models including undesirable outputs. Models that overcome the low discriminatory power of the standard DEA models such as the super-efficiency or cross efficiency model are also included. Productivity indices are also implemented, both the standard Malmquist index based on radial efficiency, and the Malmquist-Luenberger defined in terms of the directional distance function. Finally, statistical analyses and hypotheses testing using bootstrapping techniques are also available.

¹⁹If the `dispstr` parameter is omitted, the default one is used.

²⁰See the official MATLAB documentation for this function <http://www.mathworks.com/help/matlab/ref/writetable.html>.

Common fields	
Field	Data
<code>names</code>	DMU names
<code>X</code>	Inputs
<code>Y</code>	Outputs
<code>eff</code>	Efficiency measure
<code>slack.X</code>	Input slacks
<code>slack.Y</code>	Output slacks
<code>lambda</code>	Computed λ 's
<code>Xeff</code>	Efficient X's
<code>Yeff</code>	Efficient Y's
<code>dual.X</code>	Shadow prices of inputs
<code>dual.Y</code>	Shadow prices of outputs
<code>exitflag</code>	Exit flag of the optimization
Scale efficiency models	
<code>eff.crs</code>	CRS efficiency
<code>eff.vrs</code>	VRS efficiency
<code>eff.scale</code>	Scale efficiency
Cross efficiency models	
<code>eff.eff</code>	Efficiency
<code>eff.crosseff</code>	Cross efficiency score
<code>eff.PA</code>	Peer-Appraisal matrix
<code>eff.vab</code>	Inputs weights
<code>eff.uab</code>	Output weights
Malmquist index	
<code>eff.M</code>	Malmquist index
<code>eff.MTEC</code>	Technical efficiency change
<code>eff.MTC</code>	Technical change
Allocative efficiency model	
<code>Xprice</code>	Input prices
<code>Yprice</code>	Output prices
<code>eff.C</code>	Cost efficiency
<code>eff.R</code>	Revenue efficiency
<code>eff.P</code>	Profit efficiency
<code>eff.A</code>	Allocative efficiency
<code>eff.T</code>	Technical efficiency
Undesirable outputs model	
<code>Yu</code>	Undesirable outputs
<code>slack.Yu</code>	Undesirable outputs' slacks
<code>Yueff</code>	Efficient Yu's
Malmquist-Luenberger index	
<code>eff.ML</code>	Malmquist-Luenberger index
<code>eff.MLTEC</code>	Technical efficiency change
<code>eff.MLTC</code>	Technical change

Table 1: Fields of the 'deaout' structure available for the `dispstr` string.

We also show how to organize the data, use the available functions and interpret results. To illustrate all the models in the toolbox we use a reliable database on U.S. agriculture. This positions the new toolbox as a valid self-contained package for these evaluating techniques in MATLAB. Since the code is freely available in an open source repository on GitHub, under the GNU General Public License version 3, users will benefit from the collaboration and review of the community, and can check the code to learn how DEA optimizing programs are translated into suitable code.²¹

Acknowledgments

Javier Barbero and José L. Zofío thank the financial support from the Spanish Ministry for Economy and Competitiveness (Ministerio de Economía, Industria y Competitividad), the State Research Agency (Agencia Estatal de Investigación) and the European Regional Development Fund (Fondo Europeo de Desarrollo Regional) under grants PID2019-105952GB-I00(AEI/FEDER,UE), and ECO2016-79650-P (AEI/FEDER, UE).

References

- Ali A, Seiford LM (1993). “The Mathematical Programming Approach to Efficiency Analysis.” In HO Fried, CAK Lovell, SD Schmidt (eds.), *The Measurement of Productivity Efficiency. Techniques and Applications*, pp. 120–159. Oxford University Press, New York.
- Andersen P, Petersen NC (1993). “A Procedure for Ranking Efficient Units in Data Envelopment Analysis.” *Management Science*, **39**(10), 1261–1264. doi:10.1287/mnsc.39.10.1261.
- Angulo-Meza L, Pereira Estellita Lins M (2002). “Review of Methods for Increasing Discrimination in Data Envelopment Analysis.” *Annals of Operations Research*, **16**(1), 225–242. doi:10.1023/a:1021340616758.
- Aparicio J, Lovell CAK, Pastor JT (2016). *Advances in Efficiency and Productivity*. Springer International Publishing. doi:10.1007/978-3-319-48461-7.
- Aparicio J, Pastor JT, Zofío JL (2013). “On the Inconsistency of the Malmquist-Luenberger Index.” *European Journal of Operational Research*, **229**(3), 738–742. doi:10.1016/j.ejor.2013.03.031.
- Aparicio J, Pastor JT, Zofío JL (2015). “How to Properly Decompose Economic Efficiency Using Technical and Allocative Criteria with Non-Homothetic DEA Technologies.” *European Journal of Operational Research*, **240**(3), 882–891. doi:10.1016/j.ejor.2014.08.012.
- Badunenko O, Mozharovskyi P, Kolomiytseva Y (2020). *npsf: Nonparametric and Stochastic Efficiency and Productivity Analysis*. R package version 0.7.1, URL <https://CRAN.R-project.org/package=npsf/>.

²¹The address of the repository is <https://github.com/javierbarbero/DEAMATLAB>.

- Balk BM, Barbero J, Zoffio JL (2020). “A Toolbox for Calculating and Decomposing Total Factor Productivity Indices.” *Computers & Operations Research*, **115**, 104853. doi:10.1016/j.cor.2019.104853.
- Balk BM, Zoffio JL (2018). “The Many Decompositions of Total Factor Productivity Change.” *Technical Report ERS-2018-003-LIS*, Erasmus Research Institute of Management. Erasmus University. URL <http://hdl.handle.net/1765/104721>.
- Ball VE, Färe R, Grosskopf S, Nehring R (2001). “Productivity of the U.S. Agricultural Sector: The Case of Undesirable Outputs.” In ER Dean, CR Hulten, M Harper (eds.), *New Developments in Productivity Analysis*, pp. 541–586. University of Chicago Press, Chicago.
- Banker RD, Charnes A, Cooper WW (1984). “Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis.” *Management Science*, **30**(9), 1078–1092. doi:10.1287/mnsc.30.9.1078.
- Barr RS (2004). “DEA Software Tools and Technology.” In WW Cooper, LM Seiford, J Zhu (eds.), *Handbook on Data Envelopment Analysis*, pp. 539–566. Springer-Verlag, Boston.
- Blackburn V, Brennan S, Ruggiero J (2014). *Nonparametric Estimation of Educational Production and Costs Using Data Envelopment Analysis*. Springer-Verlag.
- Bogetoft P, Otto L (2011). *Benchmarking with DEA, SFA, and R*. Springer-Verlag, New York.
- Briec W (1997). “A Graph-Type Extension of Farrell Technical Efficiency Measure.” *Journal of Productivity Analysis*, **8**(1), 95–110. doi:10.1023/a:1007728515733.
- Caves DW, Christensen LR, Diewert WE (1982). “The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity.” *Econometrica*, **50**(6), 1393–1414. doi:10.2307/1913388.
- Chambers RG, Chung Y, Färe R (1996). “Benefit and Distance Functions.” *Journal of Economic Theory*, **70**(2), 407–419. doi:10.1006/jeth.1996.0096.
- Chambers RG, Chung Y, Färe R (1998). “Profit, Directional Distance Functions, and Nerlovian Efficiency.” *Journal of Optimization Theory and Applications*, **98**(2), 351–364. doi:10.1023/a:1022637501082.
- Charnes A, Cooper WW, Rhodes E (1978). “Measuring the Efficiency of Decision Making Units.” *European Journal of Operational Research*, **2**(6), 429–444. doi:10.1016/0377-2217(78)90138-8.
- Chung YH, Färe R, Grosskopf S (1997). “Productivity and Undesirable Outputs: A Directional Distance Function Approach.” *Journal of Environmental Management*, **51**(3), 229–240. doi:10.1006/jema.1997.0146.
- Cook WD, Zhu J (2015). “DEA Cross Efficiency.” In J Zhu (ed.), *Data Envelopment Analysis: A Handbook of Models and Methods*, pp. 23–43. Springer-Verlag.
- Cooper WW, Pastor JT, Aparicio J, Borrás F (2011). “Decomposing Profit Inefficiency in DEA through the Weighted Additive Model.” *European Journal of Operational Research*, **212**(2), 411–416. doi:10.1016/j.ejor.2011.01.054.

- Cooper WW, Seiford LM, Tone K (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*. Princeton University Press, Princeton.
- Costantino N, Dotoli M, Epicoco N, Falagario M, Sciancalepore F (2012a). “A Cross Efficiency Fuzzy Data Envelopment Analysis Technique for Supplier Evaluation under Uncertainty.” In *Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA)*. IEEE. doi:10.1109/etfa.2012.6489600.
- Costantino N, Dotoli M, Epicoco N, Falagario M, Sciancalepore F (2012b). “A Novel Fuzzy Data Envelopment Analysis Methodology for Performance Evaluation in a Two-Stage Supply Chain.” In *Proceedings of the 8th IEEE International Conference on Automation Science and Engineering (CASE 2012)*, pp. 974–979. IEEE. doi:10.1109/coase.2012.6386424.
- Daraio C, Kerstens K, Nepomuceno T, Sickles RC (2019). “Productivity and Efficiency Analysis Software: An Exploratory Bibliographical Survey of the Options.” *Journal of Economic Surveys*, **33**(1), 85–100. doi:10.1111/joes.12270.
- Daraio C, Simar L (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*. Springer-Verlag. doi:10.1007/978-0-387-35231-2.
- Daraio C, Simar L (2016). “Efficiency and Benchmarking with Directional Distances: A Data-Driven Approach.” *Journal of the Operational Research Society*, **67**(7), 928–944. doi:10.1057/jors.2015.111.
- Dotoli M, Epicoco N, Falagario M, Sciancalepore F (2015). “A Cross-Efficiency Fuzzy Data Envelopment Analysis Technique for Performance Evaluation of Decision Making Units under Uncertainty.” *Computers and Industrial Engineering*, **79**, 103–114. doi:10.1016/j.cie.2014.10.026.
- Doyle JR, Green RH (1994). “Efficiency and Cross-Efficiency in DEA: Derivations, Meanings, and Uses.” *Journal of the Operational Research Society*, **45**(5), 567–578. doi:10.1057/jors.1994.84.
- Du J, Liang L, Zhu J (2010). “A Slacks-Based Measure of Super-Efficiency in Data Envelopment Analysis: A Comment.” *European Journal of Operational Research*, **204**(3), 694–697. doi:10.1016/j.ejor.2009.12.007.
- Eaton JW, Bateman D, Hauberg S, Wehbring R (2020). *GNU Octave Version 5.2.0 Manual: A High-Level Interactive Language for Numerical Computations*. URL <https://www.gnu.org/software/octave/>.
- Econometric Software, Inc (2009). *LIMDEP*. Plainview. URL <http://www.limdep.com/>.
- Emrouznejad A, Cabanda E (eds.) (2014). *Managing Service Productivity: Using Frontier Efficiency Methodologies and Multicriteria Decision Making for Improving Service Performance*. Springer-Verlag, Berlin.
- Färe R, Grosskopf S, Margaritis D (2008). “Efficiency and Productivity: Malmquist and More.” In HO Fried, CAK Lovell, SS Schmidt (eds.), *The Measurement of Productive Efficiency and Productivity Change*, pp. 522–621. Oxford University Press.

- Färe R, Grosskopf S, Norris M, Zhang Z (1994). “Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries.” *American Economic Review*, **84**(1), 66–83.
- Farrell M (1957). “The Measurement of Productive Efficiency.” *Journal of the Royal Statistical Society A*, **120**(3), 253–281. doi:10.2307/2343100.
- Fried HO, Lovell CAK, Schmidt SS (2008). “Efficiency and Productivity.” In HO Fried, CAK Lovell, SS Schmidt (eds.), *The Measurement of Productive Efficiency and Productivity Change*, pp. 3–91. Oxford University Press.
- Hatami-Marbini A, Emrouznejad A, Tavana M (2011). “A Taxonomy and Review of the Fuzzy Data Envelopment Analysis Literature: Two Decades in the Making.” *European Journal of Operational Research*, **214**(3), 457–472. doi:10.1016/j.ejor.2011.02.001.
- Hollingsworth B (2004). “Non Parametric Efficiency Measurement.” *The Economic Journal*, **114**(496), F307–F311. doi:10.1111/j.1468-0297.2004.00224.x.
- Hwang SN, Lee HS, Zhu J (eds.) (2016). *Handbook of Operations Analytics Using Data Envelopment Analysis*. Springer-Verlag. doi:10.1007/978-1-4899-7705-2.
- Ji Y, Lee C (2010). “Data Envelopment Analysis.” *Stata Journal*, **10**(2), 267–280. doi:10.1177/1536867x1001000207.
- Kao C, Liu ST (2009). “Stochastic Data Envelopment Analysis in Measuring the Efficiency of Taiwan Commercial Banks.” *European Journal of Operational Research*, **196**(1), 312–322. doi:10.1016/j.ejor.2008.02.023.
- Khezzrimotlagh D, Chen Y (2018). *Decision Making and Performance Evaluation Using Data Envelopment Analysis*. Springer-Verlag. doi:10.1007/978-3-319-76345-3.
- Lovell CAK, Pastor JT (1995). “Units Invariant and Translation Invariant DEA Models.” *Operations Research Letters*, **18**(3), 147–151. doi:10.1016/0167-6377(95)00044-5.
- Oh DH, Suh D (2013). **nonparaeff**: *Nonparametric Methods for Measuring Efficiency and Productivity*. R package version 0.5-8, URL <https://CRAN.R-project.org/package=nonparaeff/>.
- Olesen OB, Petersen NC (2016). “Stochastic Data Envelopment Analysis – A Review.” *European Journal of Operational Research*, **251**(1), 2–21. doi:10.1016/j.ejor.2015.07.058.
- Ozcan YA (2014). *Health Care Benchmarking and Performance Evaluation. An Assessment Using Data Envelopment Analysis (DEA)*. Springer-Verlag.
- Paradi J, Sherman HD, Tam FK (2018). *Data Envelopment Analysis in the Financial Services Industry. A Guide for Practitioners and Analysts*. Springer-Verlag.
- Pastor JT, Lovell CAK, Aparicio J (2011). “Families of Linear Efficiency Programs Based on Debreu’s Loss Function.” *Journal of Productivity Analysis*, **38**(2), 109–120. doi:10.1007/s11123-011-0216-4.
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

- Sexton TR, Silkman RH, Hogan AJ (1986). “Data Envelopment Analysis: Critique and Extensions.” In RH Silkman (ed.), *Measuring Efficiency: An Assessment of Data Envelopment Analysis. New Directions for Program Evaluation 32*, pp. 73–105. Jossey-Bass.
- Sheel H (2000). *EMS: Efficiency Measurement System User’s Manual*. University of Dortmund, Dortmund.
- Shephard RW (1953). *Cost and Production Functions*. Princeton University Press, Princeton, New Jersey.
- Sherman HD, Zhu J (2006). *Service Productivity Management: Improving Service Productivity Using Data Envelopment Analysis (DEA)*. Springer-Verlag, New York. doi:10.1007/0-387-33231-6.
- Simar L, Wilson PW (1998). “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models.” *Management Science*, **44**(1), 49–61.
- Simar L, Wilson PW (1999). “Estimating and Bootstrapping Malmquist Indices.” *European Journal of Operational Research*, **115**(3), 459–471. doi:10.1016/s0377-2217(97)00450-5.
- Simar L, Wilson PW (2002). “Non-Parametric Tests of Returns to Scale.” *European Journal of Operational Research*, **139**(1), 115–132. doi:10.1016/s0377-2217(01)00167-9.
- Simm J, Besstremyannaya G (2020). “**rDEA**: Robust Data Envelopment Analysis (DEA) for R.” R package version 1.2-6, URL <https://CRAN.R-project.org/package=rDEA/>.
- StataCorp (2015). *Stata Statistical Software: Release 14*. StataCorp LP, College Station. URL <http://www.stata.com>.
- The MathWorks, Inc (2017). *MATLAB — The Language of Technical Computing, Version R2017b (8.5)*. Natick. URL <http://www.mathworks.com/products/matlab/>.
- Wilson PW (2008). “**FEAR**: A Software Package for Frontier Efficiency Analysis with R.” *Socio-Economic Planning Sciences*, **42**(4), 247–254. doi:10.1016/j.seps.2007.02.001.
- Zhu J (2014). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. 3rd edition. Springer-Verlag. doi:10.1007/978-1-4757-4246-6.
- Zhu J (ed.) (2015). *Data Envelopment Analysis: A Handbook of Models and Methods*. Springer-Verlag.
- Zofio JL (2007). “Malmquist Productivity Index Decompositions: A Unifying Framework.” *Applied Economics*, **39**(18), 2371–2387. doi:10.1080/00036840600606260.
- Zofio JL, Lovell CAK (2001). “Graph Efficiency and Productivity Measures: An Application to US Agriculture.” *Applied Economics*, **33**(11), 1433–1442. doi:10.1080/00036840010009865.
- Zofio JL, Pastor JT, Aparicio J (2013). “The Directional Profit Efficiency Measure: On Why Profit Pnecfficiency Is Either Technical or Allocative.” *Journal of Productivity Analysis*, **40**(3), 257–266. doi:10.1007/s11123-012-0292-0.

Affiliation:

Inmaculada C. Álvarez, Javier Barbero, José L. Zofío

Department of Economics

Universidad Autónoma de Madrid

28049 Madrid, Spain

E-mail: inmaculada.alvarez@uam.es, jose.zofio@uam.es

URL: <http://www.deatoolbox.com/>