Compind: Composite indicators functions based on frontiers in R

(Compind package version 2.0)

F. Vidoli and E. Fusco

July 2, 2026

Introduction

CI's methods are increasingly recognized as a useful tool in policy analysis and public communication (Nardo *et al.*, 2005) for a variety of policy matters such as public units benchmark, industrial competitiveness, sustainable development, quality of life assessment, globalization and innovation. They provide simple comparisons of units that can be used to illustrate complex and sometimes elusive issues in wide ranging fields, *e.g.* the environmental, economical, social or technological development. These indicators often seem easier to interpret by the general public finding a common trend in many separate indicators and have proven useful in benchmarking country performance.
Along such lines the Joint Research Centre of European Commission asserts that *"no uniformly agreed methodology exists to weight individual indicators before aggregating them into a composite indicator"*[1].

Several steps are involved in creating composite indicators: *investigating the structure* of simple indicators by means of multivariate statistics, handling the problem of *missing data* that can be missing either in a random or in a non-random fashion, bringing the indicators to the same unit by *normalization* and finally selecting an appropriate *weighting and aggregation* model. (for a complete explanation of every step, please see Nardo *et al.*, 2005).
A much wider ranging literature is found for the *aggregation* methods than the one regarding *weight* systems; however, the two aspects are related and interwoven and often lead to the same solutions.
Several weighting techniques exist in literature[2], derived both from statistical methodologies, such as factor analysis, *DEA* and unobserved components models (UCM), or from more specific methods like budget allocation processes (BAP), analytic hierarchy processes (AHP) or conjoint analysis (CA).

[1] `https://composite-indicators.jrc.ec.europa.eu`

[2] For a complete review, please see Nardo *et al.* (2005) and Freudenberg (2003) for major applications and papers.

The applicative difficulties in applying composite indicators (CI) methods derived from the production frontier analysis (*i.e.* Benefit of the Doubt - BoD) have often discouraged the practical adoption of the more complex methods, while having desirable properties.
`Compind` package make comparable and easily calculable composite indicators developed with a plurality of methods and supports researcher into robustness analysis through repeated simulations on subsamples of units or variables.
Given that, the first question is: why a frontier CI package in R? Answer is easy: R is the most comprehensive statistical analysis package available (over 4800 packages), R is free, cross-platform and open source software, but especially R is a programming language (no specific pull-down menu software) allowing to rethinking CI not only as an evaluation tool, but as a part of the main research flow making easy carry on sensitivity analysis through bootstrap replications.

So the subsequent question become: how design a CI package in R? In our opinion, the package would have these basis properties:

- It has to be as simple as possible to use;
- The syntax has to be easy and independent (as possible) from the chosen method;
- Package must cover several steps of the CI calculation (not only the weighting and aggregation step).

Given these premises, `Compind` R package contains a plurality of methods can be divided into:

- Frontier methods;
- Non frontier methods;
- Utilities.

1 Frontier methods

Table 1 shows the BoD-frontier functions implemented in Compind: more specifically, the functions differ due to the possibility of constraining the sets of variation of individual weights (*Weight constraints*), of being robust with respect to outliers or out-of-scale data (*Robust*), of being able to natively include indicators with negative polarity (*Bad output*), to take into account external factors (*Conditional*) and finally to impose a direction defined by the user in the relationship between simple indicators (*Directional*).

BoD function	**Weight constraints**	**Robust**	**Bad output**	**Conditional**	**Directional**
ci_bod					
ci_bod_constr	X				
ci_bod_constr_bad	X		X		
ci_bod_dir					X
ci_rbod		X			
ci_rbod_constr_bad	X	X	X		
ci_rbod_constr_bad_Q	X	X	X	X	
ci_rbod_dir		X			X
ci_rbod_spatial		X		Spatial	

Table 1: Frontier functions by additional capabilities

However, not all combinations have been developed: it is our intention, however, in the next versions to develop them.

1.1 Benefit of the Doubt approach

"The Benefit of the Doubt approach is formally tantamount to the original input-oriented CRS-DEA[3] *model of Charnes* et al. *(1978), with all questionnaire items considered as outputs and a dummy input equal to one for all observations"*, Witte & Rogge (2009).

In particular BoD approach offers several advantages:

1. Weights are *endogenously determined* by the observed performances and benchmark is not based on theoretical bounds, but it's a linear combination of the observed best performances.

2. Principle is *easy to communicate*: since we are not sure about the right weights, we look for "benefit of the doubt" weights (such that your overall relative performance index is as high as possible).

3. BoD CI is *weak monotone*.

[3]Constant Returns to Scale Data Envelopment Analysis.

So, let's draw a sample of 100 units for two simple indicators $i1$ and $i2 \in [0, 1]$ and two "particular" rows: the first one is an outlier, while the second one have a `NA` on the second indicator.

```
i1 <- seq(0.3, 0.5, len = 100) - rnorm (100, 0.2, 0.05)
i2 <- seq(0.3, 1, len = 100)   - rnorm (100, 0.2, 0.05)
dati = data.frame(i1, i2)
random1 = data.frame(i1=0.6, i2=1)
random2 = data.frame(i1=0.5, i2=NA)
Indic = rbind(dati,random1,random2)
```

As pointed out by the OECD Handbook on Constructing Composite Indicators, dataset must not contain missing data; to overcome this issue researcher can make imputation or delete the observations. For this reason, all the `Compind` functions alert users to the presence of missing values within the data (depending on the function the calculation can stop or not).

```
CI1 = ci_bod(Indic)

## Pay attention:  NA values at column:  102 , row 2 .  Composite indicator
has been computed, but results may be misleading, Please refer to OECD
handbook, pg.  26.
```

Given that, in this example, missing row has been deleted and the BoD composite indicator by `ci_bod` function recalculated; Figure 1 show the sample data highlighting the contribution of the outlier on the composite scores of the other units.

```
Indic = Indic[complete.cases(Indic),]
CI1 = ci_bod(Indic)
Indic_CI = data.frame(Indic, CI_est= CI1$ci_bod_est)
ggplot(data = Indic_CI, aes(x = i1, y = i2)) +
       geom_point(aes(colour = CI_est),size=3)
```

It may be readily noted that the BoD composite score depends exclusively on the frontier's distance; in this framework one drawback is directly linked with the *DEA* problem solution: since the weights are unit specific, cross-unit comparisons are not possible and the values of the scoreboard depend on the benchmark performance.
There are also three other drawbacks we will discuss in the following paragraphs: the multiplicity of equilibria, the lack of robustness and perfect non compensability among indicators.

1.2 Multiplicity of equilibria: Variance weighted BoD

As pointed out before, BoD formulation can hide the problem of the multiplicity of equilibria thus weights are not uniquely determined (even though the CI is

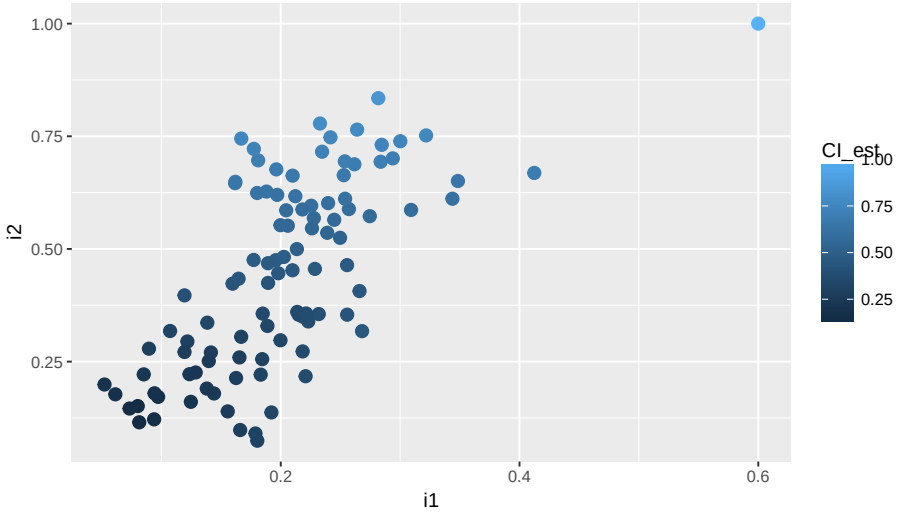


Figure 1: Simple indicators and BoD CI

unique). The weight values for the units are to be chosen from many (infinite) possibilities. It is also worth noting that multiple solutions are likely to depend upon the set of constraints imposed on the weights of the maximization problem: the wider the range of the variation of weights, the lower the possibility of obtaining a unique solution.

The optimization process could lead to many zero weights (see table 2) if no restrictions on the weights are imposed.

	Weights	Freq
1	0 - 1	75
2	1.667 - 0	26

Table 2: BoD weights

There is a wide choice for incorporating "*value judgements*" in a DEA classical model and in general in efficiency analysis (please see Allen *et al.*, 1997, Estellita-Lins *et al.*, 2007 and Thanassoulis *et al.*, 2004); three basic approaches are the most used:

- Direct restrictions on the weights;
- Adjustment of the observed input-output levels;
- Restrictions on the virtual inputs and outputs.

In recent years many additional weighting schema have been proposed (*i.e.* Rogge, 2012); Mazziotta & Vidoli (2009), for example, proposed the inclusion of

additional "Assurance regions", type I (AR I) constraints in order to highlight indicators with a higher sample variance than the others.

The basic thesis involves weighting simple indicators by their own sample variance; thus, indicators with a high variability will strongly affect the composite indicator. There are however consequences to this approach: our measurement has to be read as a "gap indicator" among the unit characteristics. The preliminary hypothesis is that every single indicator $I_q, q = 1, \ldots, Q$ is a probabilistic variable, following a Normal Gaussian distribution[4]:

$$I_q \sim N(\mu_{I_q}, \sigma_{I_q}), \forall q = 1, \ldots, Q \tag{1}$$

In this way, the variance of each indicator can be computed in a standard probabilistic setting and the unbiased variance confidence interval is:

$$P(\frac{n-1}{\chi^2_{n-1,1-\alpha/2}}\overline{S}^2 < \sigma^2 < \frac{n-1}{\chi^2_{n-1,\alpha/2}}) = 1 - \alpha \tag{2}$$

which, for the sake of compactness, can be written:

$$P(low_{I_q} < \sigma^2 < high_{I_q}) = 1 - \alpha \tag{3}$$

Even when the underlying distribution is not Normal, the procedure can be still used to obtain the approximate confidence bounds for the variance estimated. If the distribution is not too far from the Normal one, we have tested the robustness of our procedure. We can use low_{I_q} and $high_{I_q}$ for each indicator to reconstruct the marginal rates of substitution among indicators:

$$\frac{low_{I_i}}{high_{I_j}} \leq \frac{w_{I_i}}{w_{I_j}} \leq \frac{low_{I_j}}{high_{I_i}}, \forall i, j = 1, \ldots, Q \tag{4}$$

When the confidence interval inferior limit of the variance is contrasted with the maximum of another, one assumes a *"benefit of doubt"* attitude in that an exact relationship among weights is not imposed, thereby establishing a range in which every unit obtains the maximum relative weight.

In `Compind` package the implementation of this model thought the `ci_bod_var_w` function is easy and quite similar to the BoD model; Figure 2 shows how the variance weighted CI is, for construction, lower than the BoD one.

```
CI_w1 = ci_bod_var_w(Indic)
Indic_CI2 = data.frame(Indic_CI, CI_w_est= CI_w1$ci_bod_var_w_est)
ggplot(data = Indic_CI2, aes(x = CI_est, y = CI_w_est)) +
       geom_point(size=3)+
       geom_abline(intercept = 0, slope = 1, linetype="dashed")+
       xlab("BoD estimated CI") +
       ylab("Variance weighted BoD estimated CI")
```

[4]To bypass this assumption, future developments of this methodology may involve the analysis of the kernel density estimate of the simple indicators and their own sample variance.

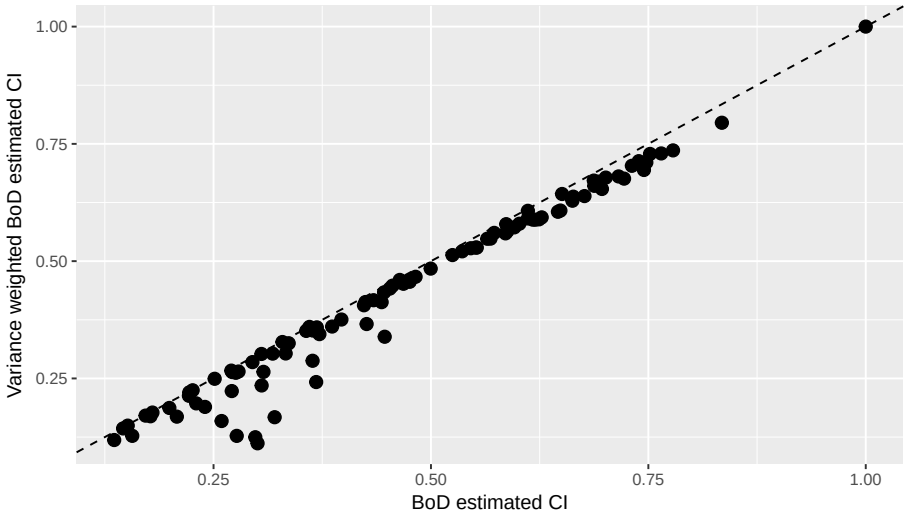


Figure 2: BoD and Variance weighted BoD estimated CI

1.3 Robust BoD

As mentioned in paragraph 1.1, one of the main drawbacks of *DEA/FDH* non-parametric estimators is their sensitivity to extreme values and outliers.

To introduce Robust BoD we first expose the simplified idea (based on the *Order-m* idea, Daraio & Simar, 2005).

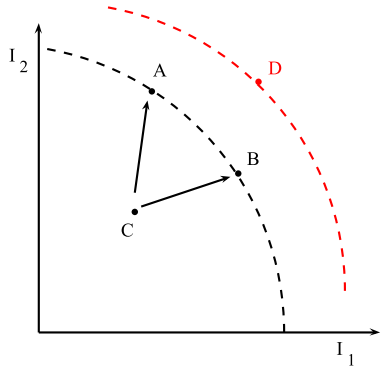


Figure 3: Outlier effects in a frontier framework

We extend the Daraio & Simar (2005) idea into CI's framework by repeatedly and with replacement drawing m observations from the original sample of n observations, choosing only from those observations which are obtaining higher basic indicators (I_1, I_2) - *red lines* - than the evaluated observation C.

In other words and practically speaking:

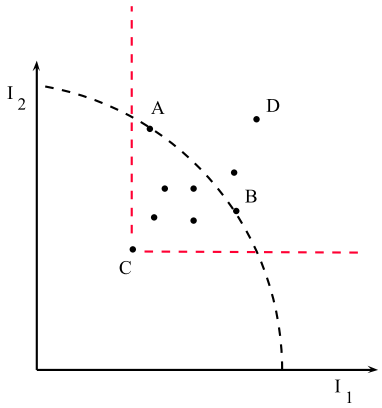


Figure 4: Support of the generic unit C

- we draw m observation only from those observations which are obtaining higher basic indicators than the evaluated observation C;
- we label this set as SET_{bm};
- we estimate BoD scores relative to this sub-sample SET_{bm} for B times;
- having obtained the B scores, we compute the arithmetic average.

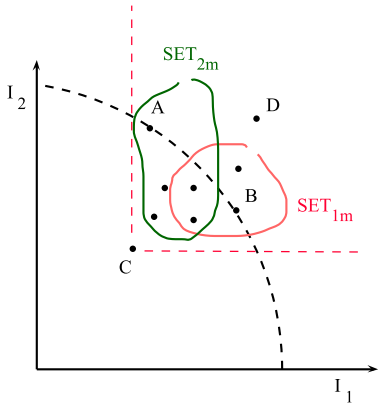


Figure 5: Order-m calculation criteria

This is certainly a less extreme benchmark for the unit C than the *"absolute"* maximum achievable level of output.
Unit C is compared to a set of m peers (potential competitors) having higher basic indicators than its level and we take as a benchmark, the expectation of the maximum achievable CI in place of the absolute maximum CI.

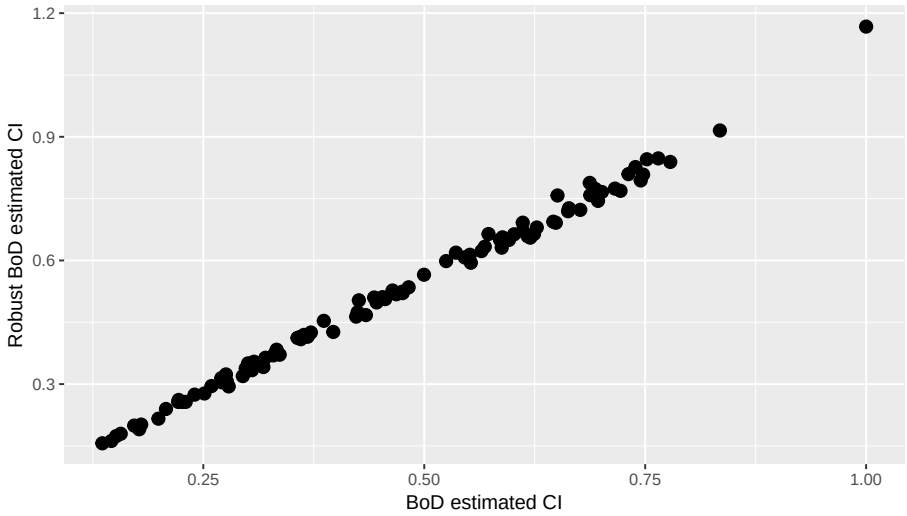


Figure 6: BoD and Robust BoD estimated CI

`Compind` package lets to calculate Robust BoD via `ci_rbod` function; two other options, respect to the `ci_bod` function, are available: `M` to fix the number of peers for the generic unit i in each sample and `B` to indicate the number of bootstrap replicates.

```
CI_r1 = ci_rbod(Indic, B=100)
Indic_CI3 = data.frame(Indic_CI2, CI_r_est= CI_r1$ci_rbod_est)
ggplot(data = Indic_CI3, aes(x = CI_est, y = CI_r_est)) +
       geom_point(size=3)+
       xlab("BoD estimated CI") +
       ylab("Robust BoD estimated CI")
```

Figure 6 allows to detect the outlier (with robust score greater than 1) and, above all, to obtain a score distribution (see Figure 7) not affected by outliers.

```
per_plot = melt(data.frame(Indic_CI3$CI_est,Indic_CI3$CI_r_est))
ggplot(per_plot, aes(x=value, fill=as.factor(variable))) +
       geom_density(alpha=.5)+
       labs(x = "Composite indicator", y="Kernel density")+
       theme(legend.position="bottom")+
       scale_fill_manual(values=c("#999999", "#E69F00"),
                         name="CI estimated value",
                         labels=c("BoD", "Robust BoD"))
```

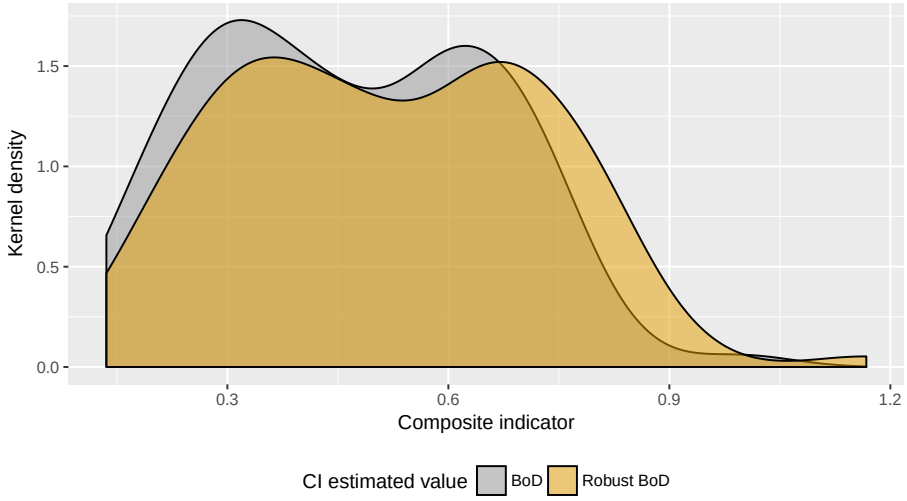


Figure 7: BoD and Robust BoD CI kernel density

1.4 Directional BoD

Most of aggregation methods assume, in weighting phase, the compensability among simple indicators (Bouyssou & Vansnick, 1986) namely allowing lower values in some indicators to be compensated by higher values in others. This property, even not verified in the practical application, is not appropriate especially if CI has to be interpreted as "importance coefficients" (Munda & Nardo, 2005).
In last years multiple solutions have been proposed to avoid this strong assumption introducing weight constraints, weighting each tensor that links the single point to the frontier (see *e.g.* Tsutsui *et al.*, 2009) or including a penalty according to the different mix of simple indicators (De Muro *et al.*, 2010).
Given that in practical application most often exist a preference structure and with the aim to respect the weakly positive monotonicity property (Casadio Tarabusi & Guarini, 2013), Fusco (2015) suggest to include in the BoD model a "directional" penalty using the directional distance function introduced by Chambers *et al.* (1998).
Even if in literature a crucial question in a directional approach is the correct choice of the direction, this issue is irrelevant with the illustration of this package and for this reason it's left to the research decisions.
To better illustrate the characteristics of the Directional BoD method the European regional transport data, year 2012, for 34 NUTS1 regions has been used[5]; Figure 8 relates the kilometres of roads and railways highlighting as, for most of the regions, the "desired" ratio can be set equal to 2 to 10.

[5]In the ode below function *normalise_ci* has been used; see paragraph 3 for more info.

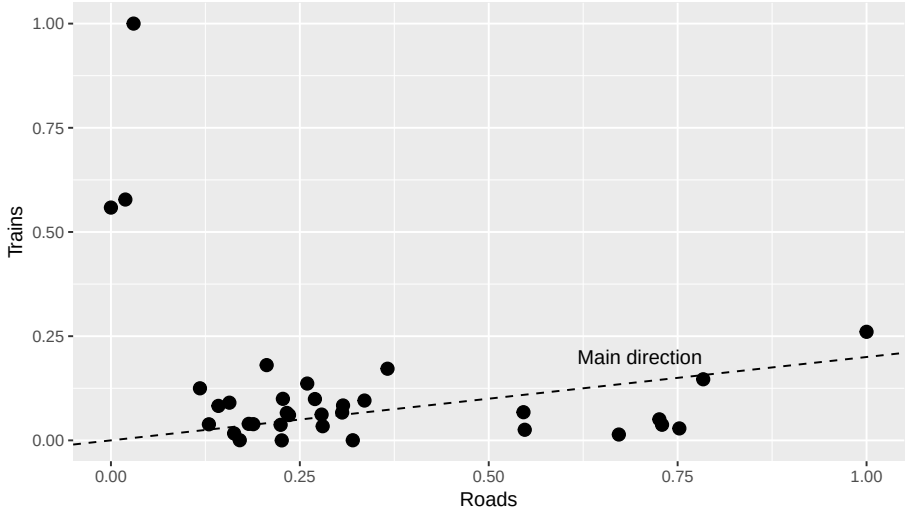


Figure 8: Eu regional transport data, year 2012

```
data(EU_NUTS1)
data_norm = normalise_ci(EU_NUTS1,c(2:3),
                         polarity = c("POS","POS"), method=2)
ggplot(data = data_norm$ci_norm, aes(x = roads, y = trains)) +
       geom_point(size=3)+
       geom_abline(intercept=0, slope=0.2, linetype="dashed")+
       annotate("text", x=0.7, y=0.2, label="Main direction")+
       xlab("Roads") +
       ylab("Trains")
```

Function `ci_bod_dir` allows to calculate Directional BoD given a direction `dir`, expressed as the ratio between the first and the second indicator; Figure 9 highlight as the main differences between BoD CI and Directional BoD CI occur for the units with the lowest values along the chosen direction.

```
CI_bod_est     = ci_bod(data_norm$ci_norm,c(1:2))
CI_bod_dir_est = ci_bod_dir(data_norm$ci_norm,c(1:2),
                            dir = c(1,0.2))
Diff = CI_bod_dir_est$ci_bod_dir_est - CI_bod_est$ci_bod_est
Indic_tot = data.frame(data_norm, Diff)

ggplot(data = Indic_tot,
       aes(x = ci_norm.roads, y = ci_norm.trains)) +
       geom_point(aes(colour = Diff),size=3)+
       theme(legend.position="bottom")+
```

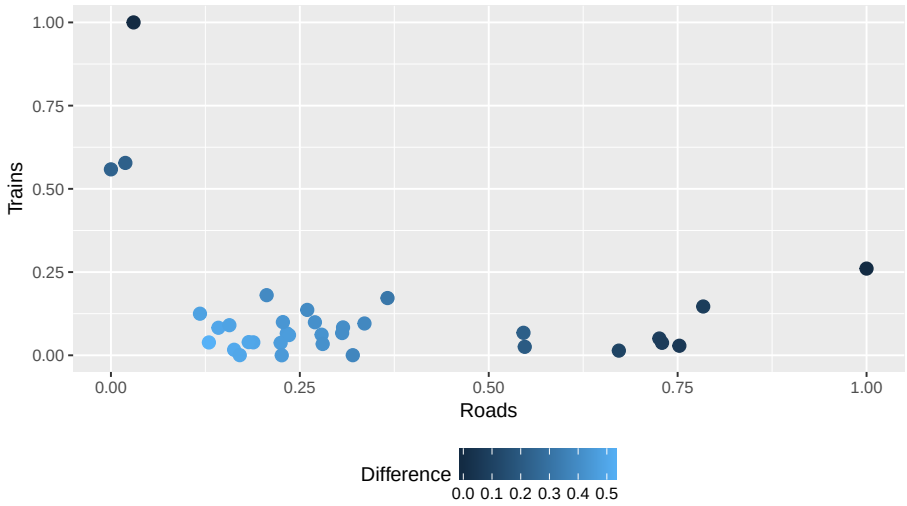


Figure 9: Eu regional transport data - difference between BoD and Directional BoD

```
       scale_colour_continuous(name="Difference")+
       xlab("Roads") +
       ylab("Trains")
```

1.5 Directional Robust BoD

Directional Robust BoD method, proposed in Vidoli *et al.* (2015), is the logical union between the Robust BoD and the directional BoD methods; Figure 10 compares the directional measure with the directional robust one, highlighting how, even in this case, the main differences occur for the units with the lowest values.

```
CI_rbod_dir_est = ci_rbod_dir(data_norm$ci_norm,c(1:2),
                              dir = c(1,0.2))
Indic_tot = data.frame(data_norm,
            CI_dir  = CI_bod_dir_est$ci_bod_dir_est,
            CI_rdir = CI_rbod_dir_est$ci_rbod_dir_est)
ggplot(data = Indic_tot, aes(x = CI_dir, y = CI_rdir)) +
       geom_point(size=3)+
       geom_abline(intercept = 0, slope = 1, linetype="dashed")+
       xlab("Directional BoD estimated CI") +
       ylab("Directional Robust BoD estimated CI")
```

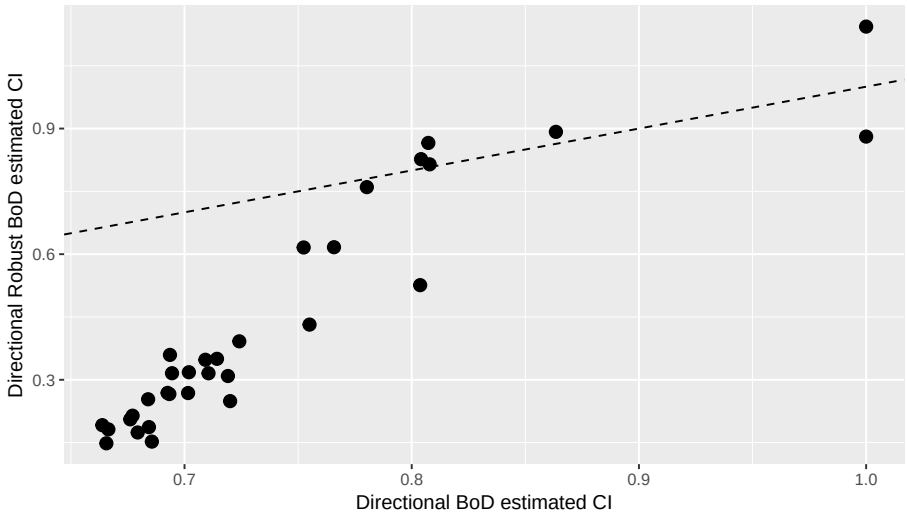


Figure 10: Directional BoD vs Directional Robust BoD estimated CI

2 Non frontier methods

This section provides some functions commonly used in the calculation of composite indicators; `Compind` implements the main methodologies proposed in the OECD manual more closely linked with mathematical procedure avoiding all methods which in some way would provide for a subjective choice of the weights.

2.1 Weighting method based on Factor Analysis

Factor Analysis (FA) aims to describe a set of Q indicators $i1, i2, \ldots, iQ$ in terms of a smaller number of m factors and to highlight the relationship between these variables. Contrary to the Principal Component Analysis, the FA model assumes that the data is based on the underlying factors of the model, and that the data variance can be decomposed into that accounted for by common and unique factors.
On the issue of how factors should be retained in the analysis without losing too much information, methodologists are divided; `Compind` package with the `ci_factor` function offers three possibilities: 1) `method="ONE"` (default) the composite indicator estimated values are equal to first component scores; 2) `method="ALL"` the composite indicator estimated values are equal to component scores multiplied by its proportion variance and 3) `method="CH"` it can be choose the number of the component to take into account.

After choosing five indicators it was applied factorial analysis choosing to weigh the scores on the three components with the associated loadings.

```
data(EU_2020)
data_norm=normalise_ci(EU_2020,c(47:51),
                       polarity = c("POS","POS","POS","POS","POS"),
                       method=2)
CI1 = ci_factor(data_norm$ci_norm,c(1:5),method="CH", dim=3)
summ = summary(as.data.frame(CI1$ci_factor_est))
print(xtable(summ,caption = "Factor Analysis scores based
             on first 3 components",label="tab_factor1"),
      include.rownames=FALSE)
```

V1
Min. :-1.6549
1st Qu.:-0.4842
Median : 0.1630
Mean : 0.0000
3rd Qu.: 0.4416
Max. : 1.1475

Table 3: Factor Analysis scores based on first 3 components

The associated loadings ..

```
round(CI1$loadings_fact,3)
```

[1] 0.698 0.285 0.010

The robustness of the results can be tested even varying the number of components; in this case it was decided to retain only the first factor (`method="ONE"`).

```
CI2 = ci_factor(data_norm$ci_norm,c(1:5),method="ONE")
summ2 = summary(as.data.frame(CI2$ci_factor_est))
print(xtable(summ2,caption = "Factor Analysis scores based
             on first component",label="tab_factor2"),
      include.rownames=FALSE)
```

CI2$ci_factor_est
Min. :-3.1144
1st Qu.:-0.2607
Median : 0.1247
Mean : 0.0000
3rd Qu.: 0.6264
Max. : 1.3446

Table 4: Factor Analysis scores based on first component

It can be noted however very good correlation between the two scores (0.926).

2.2 Weighting method based on geometric aggregation

Geometric aggregation (GA) is a simple method less compensatory approach than the additive ones; in other terms, units with low scores in some indicators would prefer a linear rather than a geometric aggregation, that is an increase in an indicator value would have higher marginal utility on the composite indicator if the indicator value is low.
Since in GA compensability degree is not constant, because is higher for composite indexes with high values and vice versa, units with low scores tend to prefer use of linear aggregation, trying to improve their position in ranking.
The implementation in `Compind` package is trivial.

```
data(EU_NUTS1)
CI_geom_estimated = ci_geom_gen(EU_NUTS1,c(2:3),meth = "EQUAL")
summary(CI_geom_estimated$ci_mean_geom_est)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   420.5   914.2  1118.6  1256.9  1455.1  3820.9
```

2.3 Mazziotta-Pareto Index (MPI) method

The MPI is a non-compensative composite index which, starting from a linear aggregation, introduces a penalty for the units with unbalanced values of the indicators (De Muro *et al.*, 2010). It is composed of two parts (a measure of the mean level and a measure of the amount of unbalance) and, differently from other methods, may be used for building both "positive" and "negative" composite indices (`penalty` direction).
MPI method need to normalize simple indicator following two standardizations methods:

- For classic MPI it must use `normalize_ci` function with `method=1`, `z.mean=100` and `z.std=10`;
- For Correct MPI it must use `normalize_ci` function with min-max standardization (`method=2`).

```
data(EU_NUTS1)
data_norm = normalise_ci(EU_NUTS1,c(2:3),
                         c("NEG","POS"),
                         method=1,z.mean=100, z.std=10)
CI_pi_estimated = ci_mpi(data_norm$ci_norm, penalty="NEG")
```

2.4 Adjusted Mazziotta-Pareto Index (AMPI) method

The AMPI method is a non-compensative composite index which introduces a penalty for the units with unbalanced values of the indicators (De Muro *et al.*,

2010). It is composed of two parts (a measure of the mean level and a measure of the amount of unbalance) and, differently from other methods, may be used for building both "positive" and "negative" composite indices (`penalty` direction). Differently from the MPI method, AMPI allows to take into account the time dimension in order to make the estimates over the years comparables.
Normalizing data before use AMPI method is here not needed, because this method use a particular method that is embedded in the code itself.
Data has to be passed in Long format indicating the time variable; switching from wide to long is very simple in R using the *reshape* function (see below).

```
data(EU_2020)

data_test = EU_2020[,c("employ_2010","employ_2011",
                       "finalenergy_2010","finalenergy_2011")]

EU_2020_long<-reshape(data_test,
                      varying=c("employ_2010","employ_2011",
                                "finalenergy_2010",
                                "finalenergy_2011"),
                      direction="long",
                      idvar="geo",
                      sep="_")

CI <- ci_ampi(EU_2020_long,
              indic_col=c(2:3),
              gp=c(50, 100),
              time=EU_2020_long[,1],
              polarity= c("POS", "POS"),
              penalty="POS")
xtable(CI$ci_ampi_est)
```

Results are offered showing the units in row and the estimates for each year in column (see table below).

2.5 Mean-min Function

The Mean-Min Function (MMF), proposed by Casadio Tarabusi & Guarini (2013), can be seen as an intermediate method between arithmetic `mean`, according to which no unbalance is penalized, and `min` function, according to which the penalization is maximum. It depends on two parameters that are respectively related to the intensity of penalization of unbalance ($\alpha, 0 \leq \alpha \leq 1$) and the intensity of complementarity ($\beta, \beta \geq 0$) among indicators.
MMF index can be expressed as:

$$MMF_i = M_{Z_i} - \alpha(\sqrt[2]{(M_{Z_i} - min_j(z_{ij}))} + \beta^2 - \beta) \tag{5}$$

	2010	2011
1	156.05	152.17
2	155.89	152.12
3	138.84	131.66
4	132.74	121.60
5	146.91	140.27
6	163.43	152.53
7	159.32	153.25
8	136.34	139.08
9	130.57	123.59
10	128.99	115.36
11	125.51	119.56
12	143.31	136.40
13	115.57	109.49
14	122.44	119.19
15	161.06	146.64
16	131.64	129.24
17	129.75	130.65
18	147.90	138.34
19	119.77	116.98
20	118.90	118.52
21	166.17	155.68
22	160.49	151.14
23	129.96	125.47
24	147.24	135.89
25	127.16	121.48
26	146.71	134.21
27	130.57	126.28
28	154.67	147.55
29	170.52	162.23
30	155.80	146.74

where Z is the normalized matrix of the data.
The function reduces to the arithmetic `mean` for $\alpha = 0$ (in this case β is irrelevant) and to the minimum function for $\alpha = 1$ and $\beta = 0$. Moreover, with $\alpha = 1$ the function has incomplete compensability; with $\beta = 0$ and $0 \leq \alpha \leq 1$ it has proportional compensability.
Therefore, authors write that: "*by choosing the values of parameters appropriately one can obtain the form of this aggregation function that best suits the specific theoretical approach*".
Once fixed α and β, the implementation in `Compind` package is trivial.

```
data(EU_NUTS1)
CI_mean_min_estimated = ci_mean_min(EU_NUTS1,c(2:3),
                                    alpha=0.5, beta=1)
```

2.6 Wroclaw Taxonomic Method

Wroclaw Taxonomic Method is a technique originally developed at the University of Wroclaw, which has experienced a fairly widespread in Italy, especially for the development of economic and social indicators (see *e.g.* Schifini D'Andrea, 1982; Quirino, 1990; Mazziotta, 1998) and recently by Cwiakala-Malys, 2009.
It's based on a very simple principle: the benchmark is the one that has the least distance from an "ideal" unit, characterized by the best performance for all the indicators considered; following the calculation of (Euclidean) distances of all units by the "ideal" one, it can build a list in which the different units are ordered in proportion with the distance from the optimum situation.
The implementation in `Compind` package is trivial.

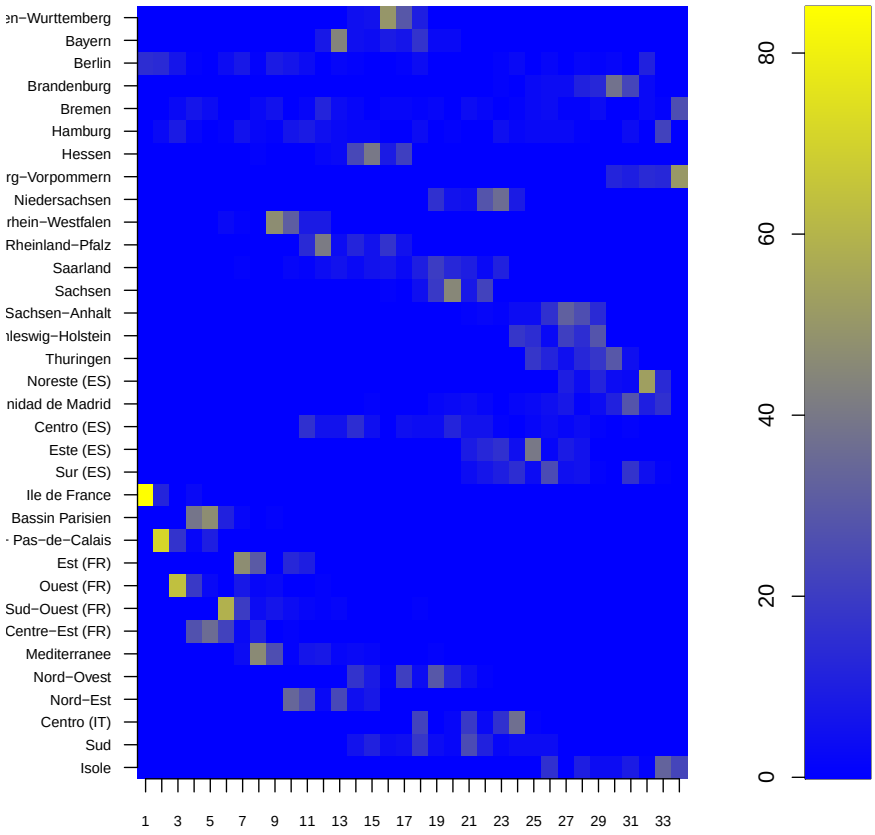
```
data(EU_NUTS1)
CI_wroclaw_estimated = ci_wroclaw(EU_NUTS1,c(2:3))
```

2.7 SMAA - Stochastic multiobjective acceptability analysis

The application of the Stochastic multiobjective acceptability analysis (SMAA) to the composite indicators is relatively recent: for more information, please see Greco *et al.* (2017).

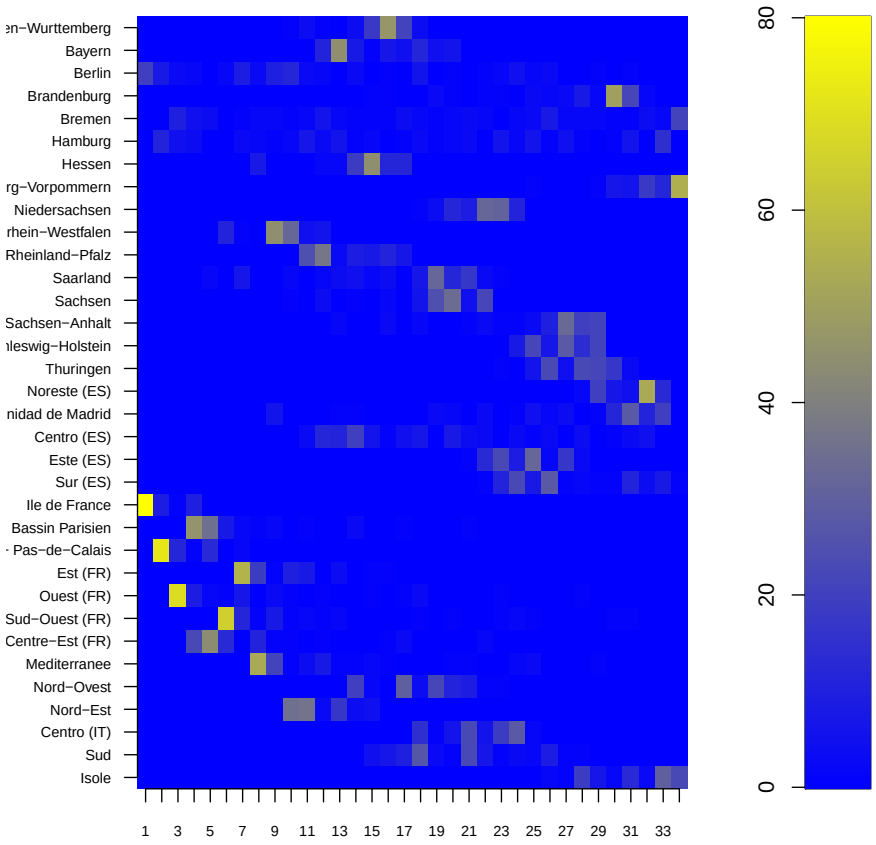
The implementation of the standard SMAA in `Compind` package is trivial.

```
data(EU_NUTS1)
test <- ci_smaa_constr(EU_NUTS1,c(2,3), label= EU_NUTS1[,1], rep=100)
source("http://www.phaget4.org/R/myImagePlot.R")
myImagePlot(test$ci_smaa_constr_rank_freq)
```



Compind package allows also to constraint the range of allowable weights specifying the upper and/or the lower bound.

```
data(EU_NUTS1)
test2 <- ci_smaa_constr(EU_NUTS1,c(2,3), label= EU_NUTS1[,1], rep=100, low_w=c(0,0.2))
source("http://www.phaget4.org/R/myImagePlot.R")
myImagePlot(test2$ci_smaa_constr_rank_freq)
```



3 Utilities: Normalisation and polarity functions

Although presented at the end, the `normalize_ci` is a crucial function that lets to normalise simple indicators according to the polarity of each one.
`Compind` provides three different methods: the standardization or z-scores (`method=1`), the min-max method (`method=2`) and the ranking method (`method=3`); each method provides for the indication of the polarity of the single indicator in order to obtain standardized indicators with the same polarity.

4 Compind web application

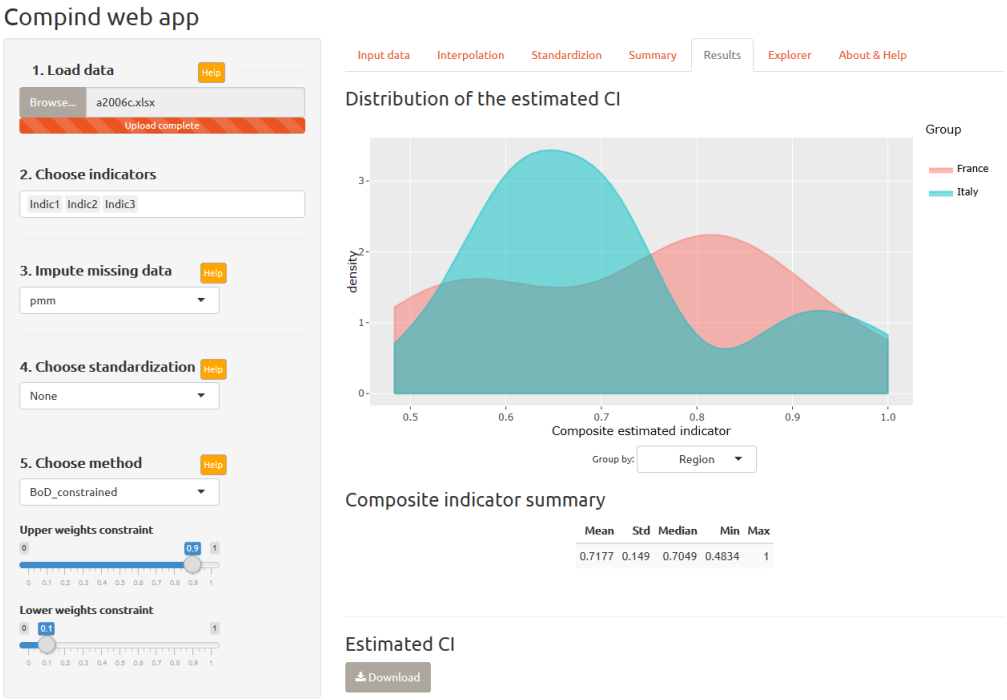


Figure 11: Compind web application

The Compind package has also been made simpler and more immediate through the design of a web interface written in Shiny - `https://fvidoli.shinyapps.io/compind_app/` - that allows to calculate the composite indicators through a guided and intuitive procedure.
Help and relative tutorial are available directly via the interface; not all methods have been implemented yet.

References

Allen, R., Athanassopoulos, A., Dyson, R., & Thanassoulis, E. 1997. Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of operations research*, **73**, 13–34.

Bouyssou, D., & Vansnick, J.-C. 1986. Noncompensatory and generalized noncompensatory preference structures. *Theory and decision*, **21**, 251–266.

Casadio Tarabusi, E., & Guarini, G. 2013. An unbalance adjustment method for development indicators. *Social indicators research*, **112(1)**, 19–45.

Chambers, R. G., Chung, Y., & Fare, R. 1998. Profit, directional distance functions, and nerlovian efficiency. *Journal of optimization theory and applications*, **98(2)**, 351–364.

Charnes, A., Cooper, W., & Rhodes, W. 1978. Measuring the efficiency of decision making units. *European journal of operational research*, **2(4)**, 429 – 444.

Cwiakala-Malys, A. 2009. The application of wroclaw taxonomy in the comparative analysis of public universities. *Operations research and decisions*, **1**, 5–25.

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

De Muro, P., Mazziotta, M., & Pareto, A. 2010. Composite indices of development and poverty: An application to mdgs. *Social indicators research*, **104(1)**, 1–18.

Estellita-Lins, M., da Silva, A. M., & Lovell, C. 2007. Avoiding infeasibility in dea models with weight restrictions. *European journal of operational research*, **181(2)**, 956–966.

Freudenberg, M. 2003. *Composite indicators of country performance: A critical assessment.* Tech. rept. OECD Science, Technology and Industry Working Papers 2003/16, OECD, Directorate for Science, Technology and Industry.

Fusco, E. 2015. Enhancing non-compensatory composite indicators: A directional proposal. *European journal of operational research*, **242**(2), 620–630.

Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. 2017. Stochastic multiattribute acceptability analysis (smaa): an application to the ranking of italian regions. *Regional studies.*

Mazziotta, C. 1998. *Esperienze e nuovi percorsi di ricerca per l'analisi delle economie locali.* Istituto Guglielmo Tagliacarne, Statistica e territorio, Milano: FrancoAngeli. Chap. Definizione di aree e indicatori per la misurazione della dotazione di infrastrutture.

Mazziotta, C., & Vidoli, F. 2009. La costruzione di un indicatore sintetico ponderato. un'applicazione della procedura benefit of the doubt al caso della dotazione infrastrutturale in italia. *Scienze regionali*, **8**(1), 35–69.

Munda, G., & Nardo, M. 2005. *Constructing consistent composite indicators: the issue of weights*. Tech. rept. EUR 21834 EN, European Commission.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. 2005. Handbook on constructing composite indicators: Methodology and user guide. *Oecd statistics working papers 2005/3, oecd, statistics directorate*.

Quirino, P. 1990. Indicatori socio-culturali a livello regionale. *Collana di studi economici, cresa*.

Rogge, N. 2012. Undesirable specialization in the construction of composite policy indicators: The environmental performance index. *Ecological indicators*, **23**, 143 – 154.

Schifini D'Andrea, S. 1982. *Le statistiche dello sviluppo*. AA. VV. Chap. Indagini sul livello di sviluppo regionale in Italia (un'applicazione comparativa).

Thanassoulis, E., Portela, M. C., & Allen, R. 2004. Incorporating value judgments in dea. *Pages 99–138 of:* Cooper, W. W., Seiford, L. M., & Zhu, J. (eds), *Handbook on data envelopment analysis*. International Series in Operations Research and Management Science, vol. 71. Springer US.

Tsutsui, M., Tone, K., & Yoshida, Y. 2009. *Technical efficiency based on cost gradient measure*. Tech. rept. Discussion Paper 09-14, GRIPS Policy Information Center, Tokyo.

Vidoli, F., Fusco, E., & Mazziotta, C. 2015. Non-compensability in composite indicators: a robust directional frontier method. *Social indicators research*, **122(3)**, 635–652.

Witte, K. D., & Rogge, N. 2009. *Accounting for exogenous influences in a benevolent performance evaluation of teachers*. Tech. rept. Working Paper Series ces0913, Katholieke Universiteit Leuven, Centrum voor Economische Studien.