Student learning assessment:
a non-compensative robust composite measure in Italy

F. Vidoli

University of Roma Tre

Efficiency in Education workshop,
The Work Foundation,
London 2014

Aim of the paper

The aim of the present research is to obtain a composite measure of learning for the elementary school students from the Italian national student assessment annual survey, year 2011.

This measure would be part of an extensive project, started in 2014 in Italy, that aims to assess the schools' value added, an indicator of school effectiveness with the purpose to estimate the specific contribution of each school into the knowledge development of the elementary school students.

Aim of the paper

The aim of the present research is to obtain a composite measure of learning for the elementary school students from the Italian national student assessment annual survey, year 2011.

This measure would be part of an extensive project, started in 2014 in Italy, that aims to assess the schools' value added, an indicator of school effectiveness with the purpose to estimate the specific contribution of each school into the knowledge development of the elementary school students.

Value Added

The Value Added is a key measure of school effectiveness that can effectively approximate the "specific contribution" of a school to the increase in knowledge of students.

In statistical terms, it is obtained from the difference between the score obtained by a student at the end of a standardized test and its expected return. An institute adds value to the increase of knowledge if the final yield of the latter exceeds the expected one.

Value Added

The Value Added is a key measure of school effectiveness that can effectively approximate the "specific contribution" of a school to the increase in knowledge of students.

In statistical terms, it is obtained from the difference between the score obtained by a student at the end of a standardized test and its expected return. An institute adds value to the increase of knowledge if the final yield of the latter exceeds the expected one.

Univariate multilevel model - 1^{st} path

Formally, a univariate multilevel (two-levels) model in the education field - *e.g.* with pupils at level 1 and classes at level 2 - can be written as:

$$y_{ij} = \alpha + \sum_{r=1}^{R} \beta_r x_{rij} + \sum_{s=1}^{S} \gamma_s z_{sj} + u_j + e_{ij} \quad (1)$$

Multivariate multilevel model - 2^{nd} path

More generally, in a R^p multivariate setting the mathematical formulation can be written as:

$$\begin{cases} y_{ij}^1 = \alpha^1 + \sum_{r=1}^R \beta_r^1 x_{rij}^1 + \sum_{s=1}^S \gamma_s^1 z_{sj}^1 + u_j^1 + e_{ij}^1 \\ y_{ij}^p = \alpha^p + \sum_{r=1}^R \beta_r^p x_{rij}^p + \sum_{s=1}^S \gamma_s^p z_{sj}^p + u_j^p + e_{ij}^p \end{cases} \quad (2)$$

In order to estimate eq.(2) *Grilli et al*. 2014 suggest to consider the multivariate p-level model as a single-outcome p-level model with subjects at level 1, pupils at level 2 and classes at level 3 $\Rightarrow$ too many level 1 effect.

Multivariate multilevel model - 2^{nd} path

More generally, in a R^p multivariate setting the mathematical formulation can be written as:

$$\begin{cases} y_{ij}^1 = \alpha^1 + \sum_{r=1}^{R} \beta_r^1 x_{rij}^1 + \sum_{s=1}^{S} \gamma_s^1 z_{sj}^1 + u_j^1 + e_{ij}^1 \\ y_{ij}^p = \alpha^p + \sum_{r=1}^{R} \beta_r^p x_{rij}^p + \sum_{s=1}^{S} \gamma_s^p z_{sj}^p + u_j^p + e_{ij}^p \end{cases} \quad (2)$$

In order to estimate eq.(2) *Grilli et al.* 2014 suggest to consider the multivariate p-level model as a single-outcome p-level model with subjects at level 1, pupils at level 2 and classes at level 3 $\Rightarrow$ too many level 1 effect.

Multivariate multilevel model - 3^{rd} path

Given these premises, in this paper, we suggest a third path, by calculating in this first step a composite indicator for the dependent variable and then regress it by a univariate multilevel model; following this way:

- you have access to a single outcome measure school (especially in the case of a plurality of scores);
- you get a more simple and robust model to estimate and more immediate to communicate;
- the univariate by composite indicator multilevel model can be useful to test the robustness of the multivariate one.

Multivariate multilevel model - 3^{rd} path

Given these premises, in this paper, we suggest a third path, by calculating in this first step a composite indicator for the dependent variable and then regress it by a univariate multilevel model; following this way:

- you have access to a single outcome measure school (especially in the case of a plurality of scores);
- you get a more simple and robust model to estimate and more immediate to communicate;
- the univariate by composite indicator multilevel model can be useful to test the robustness of the multivariate one.

Multivariate multilevel model - 3^{rd} path

Given these premises, in this paper, we suggest a third path, by calculating in this first step a composite indicator for the dependent variable and then regress it by a univariate multilevel model; following this way:

- you have access to a single outcome measure school (especially in the case of a plurality of scores);
- you get a more simple and robust model to estimate and more immediate to communicate;
- the univariate by composite indicator multilevel model can be useful to test the robustness of the multivariate one.

Multivariate multilevel model - 3^{rd} path

Given these premises, in this paper, we suggest a third path, by calculating in this first step a composite indicator for the dependent variable and then regress it by a univariate multilevel model; following this way:

- you have access to a single outcome measure school (especially in the case of a plurality of scores);
- you get a more simple and robust model to estimate and more immediate to communicate;
- the univariate by composite indicator multilevel model can be useful to test the robustness of the multivariate one.

Composite indicators

"*Weighting and aggregation*" step seems in general the most controversial phase in the construction of a composite indicator.

There is no universally accepted methodology and choice greatly affects the final result, particularly when there isn't a specific priority between indicators.

Composite indicators

"*Weighting and aggregation*" step seems in general the most controversial phase in the construction of a composite indicator.

There is no universally accepted methodology and choice greatly affects the final result, particularly when there isn't a specific priority between indicators.

Composite indicators

In the CI framework, there are no functional relationships between single indicators covering different aspects of a specific economical or social phenomenon and it can not be assumed nomic causality (*Born,1949*) [can not be assumed certain or probabilistic general function covering relationship among instances].

Moreover, in a nonparametric perspective, it is not even useful to introduce constraints, parametric functional forms or penalties linked to a specific theoretical model ⇒ results would be clearly linked to the theoretical model that has generated it.

Composite indicators

In the CI framework, there are no functional relationships between single indicators covering different aspects of a specific economical or social phenomenon and it can not be assumed nomic causality (*Born,1949*) [can not be assumed certain or probabilistic general function covering relationship among instances].

Moreover, in a nonparametric perspective, it is not even useful to introduce constraints, parametric functional forms or penalties linked to a specific theoretical model $\Rightarrow$ results would be clearly linked to the theoretical model that has generated it.

Benefit of the doubt (BoD) approach

BoD methodology is an application of DEA.

DEA-based composite indicators have inter alia been used to assess European labour market policy (*Storrie and Bjurek, 2000*), European social inclusion policy (*Cherchye, Moesen and Van Puyenbroeck, 2004*), internal market policy (*Cherchye et al., 2005*) and Human Development Index (*Mahlberg and Obersteiner, 2001; Despotis, 2005*).

Benefit of the doubt (BoD) approach

BoD methodology is an application of DEA.

DEA-based composite indicators have inter alia been used to assess European labour market policy (*Storrie and Bjurek, 2000*), European social inclusion policy (*Cherchye, Moesen and Van Puyenbroeck, 2004*), internal market policy (*Cherchye et al., 2005*) and Human Development Index (*Mahlberg and Obersteiner, 2001; Despotis, 2005*).

Benefit of the doubt (BoD) approach

"This adjusted model is formally tantamount to the original input-oriented CCR-DEA model of Charnes et al. (1978), with all simple indicators considered as outputs and a dummy input equal to one for all observations", *de Witte and Rogge* (2009).

The Farrel-Debreu (output) efficiency score is:

$$\lambda(\mathbb{1}, \mathbf{y}) = \sup\{\lambda > 0 | H(\mathbb{1}, \lambda\mathbf{y}) > 0\} \tag{3}$$

where

$$H(\mathbb{1}, \mathbf{y}) = Prob(X \equiv \mathbb{1}, \mathbf{Y} \geq \mathbf{y}) \tag{4}$$

Benefit of the doubt (BoD) approach

"*This adjusted model is formally tantamount to the original input-oriented CCR-DEA model of Charnes et al. (1978), with all simple indicators considered as outputs and a dummy input equal to one for all observations*", de Witte and Rogge (2009).

The Farrel-Debreu (output) efficiency score is:

$$\lambda(\mathbb{1}, \mathbf{y}) = \sup \{\lambda > 0 | H(\mathbb{1}, \lambda \mathbf{y}) > 0\} \tag{3}$$

where

$$H(\mathbb{1}, \mathbf{y}) = Prob(X \equiv \mathbb{1}, \mathbf{Y} \geq \mathbf{y}) \tag{4}$$

Benefit of the doubt (BoD) approach

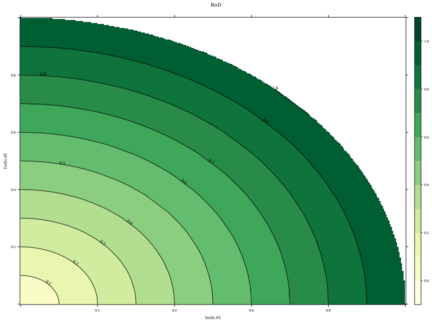


Figure: BoD distribution

Benefit of the doubt (BoD) approach

This approach offers several advantages:

BoD properties

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*.

Benefit of the doubt (BoD) approach

This approach offers several advantages:

BoD properties

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*.

Benefit of the doubt (BoD) approach

This approach offers several advantages:

BoD properties

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*.

Benefit of the doubt (BoD) approach

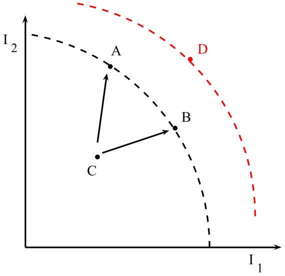
Drawbacks

The main drawbacks are directly related to the BoD framework:

1. Lack of robustness;
2. Indicators' compensability.

Drawback #1: Lack of robustness

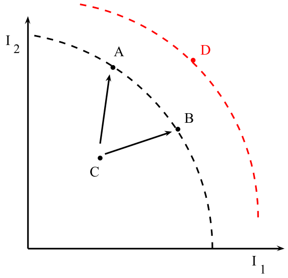
One of the main drawbacks of DEA/FDH nonparametric estimators is their sensitivity to extreme values and outliers.



Drawback #1: Lack of robustness

Cazals *et al.* (2002) proposed a more robust nonparametric estimator of the frontier. It is based on the concept of the expected minimum input function of order-m.

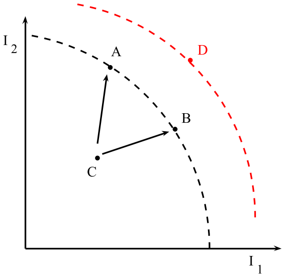
Extending these ideas to the full multivariate case, Daraio and Simar (2005) defined the concept of the expected order-m input efficiency score.



Drawback #1: Lack of robustness

Cazals *et al.* (2002) proposed a more robust nonparametric estimator of the frontier. It is based on the concept of the expected minimum input function of order-m.

Extending these ideas to the full multivariate case, Daraio and Simar (2005) defined the concept of the expected order-m input efficiency score.



Drawback #1: Lack of robustness

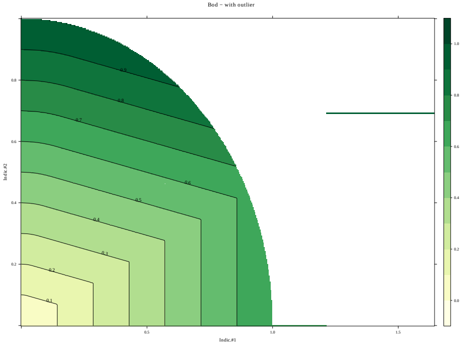


Figure: BoD distribution - outlier

Drawback #1: Lack of robustness

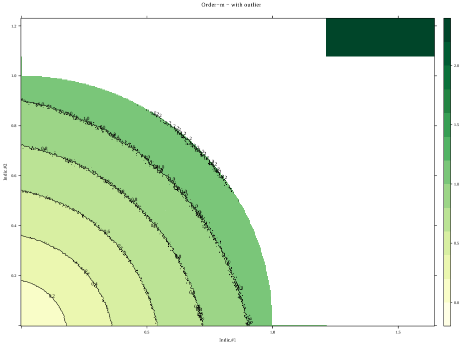
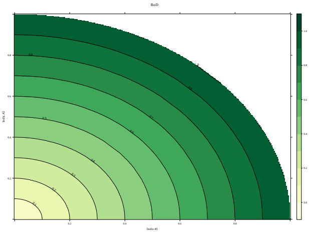


Figure: Robust BoD distribution - outlier

Drawback 2: Compensability

Another feature associated to preferential independence is compensability among the different simple indicators that in standard composite indicators approach is always assumed; this implies complete substitutability among the different indicators.

Drawback 2: Compensability



BoD composite score depends exclusively on the frontier's distance and not, contrary to the non-compensative methods, on the relationship *between* simple indicators.

Drawback 2: Compensability

To overcome the "non-compensatory" drawback, in 2012 Vidoli and Mazziotta proposed to incorporate the Mazziotta and Pareto's idea in the Robust BoD model assuming that each indicator may not be replaced by the others or is so only in part.

The method (RBoD-PCV) involves introducing a penalty for units that have not balanced a budget for all components, such as:

$$RBoD\_PCV_i = RBoD_i(1 - cv_i^2), \forall i = 1, ..., N$$

where cv_i^2 represents the coefficient of variation for the unit i between all indicators.

Drawback 2: Compensability

To overcome the "non-compensatory" drawback, in 2012 Vidoli and Mazziotta proposed to incorporate the Mazziotta and Pareto's idea in the Robust BoD model assuming that each indicator may not be replaced by the others or is so only in part.

The method (RBoD-PCV) involves introducing a penalty for units that have not balanced a budget for all components, such as:

$$RBoD\_PCV_i = RBoD_i(1 - cv_i^2), \forall i = 1, ..., N$$

where cv_i^2 represents the coefficient of variation for the unit i between all indicators.

Drawback 2: Compensability

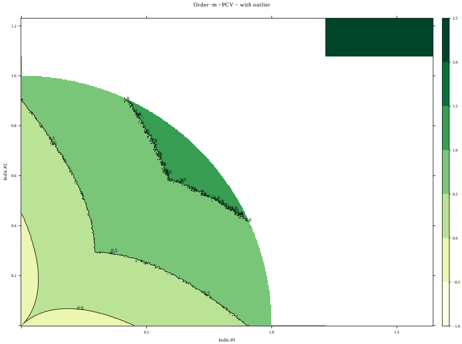
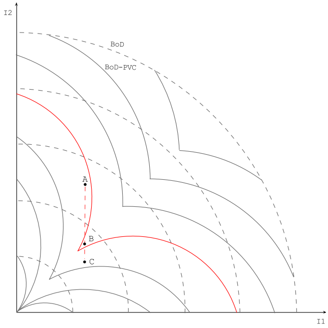


Figure: Robust BoD-PCV distribution - outlier

An important drawback to Robust BoD-PCV

Fusco, 2013

Property: positive monotonicity
$\forall\ c > 0,$
$f(I_1, ..., I_j, ..., I_k) \leq f(I_1, ..., I_{j+c}, ..., I_k).$



Directional idea

Fusco suggests to include in the classical BoD model a directional penalty using a *directional distance function*.

Using the BoD notation, the CI can be calculated as the reciprocal of the directional distance function $D(\mathbb{1}, \mathbf{y}; g_y)$:

$$D(\mathbb{1}, \mathbf{y}; g_y) = \sup\{\beta | (\mathbb{1}, \mathbf{y} + \beta g_y) \in \Psi\}$$

where g_y is the directional vector.

Directional idea

Fusco suggests to include in the classical BoD model a directional penalty using a *directional distance function*.
Using the BoD notation, the CI can be calculated as the reciprocal of the directional distance function $D(\mathbb{1}, \mathbf{y}; g_y)$:

$$D(\mathbb{1}, \mathbf{y}; g_y) = \sup\{\beta | (\mathbb{1}, \mathbf{y} + \beta g_y) \in \Psi\}$$

where g_y is the directional vector.

Directional Benefit of the doubt (DBoD)

The directional BoD adds the following properties to the BoD model:

DBoD properties

4. **Noncompensability property**: the directional vector g_y rewards units along a generic direction (not only along the bisector direction) by penalizing units far from the chosen direction; therefore full compensability can be seen as a special case when $g_y = \mathbb{1}$.
5. **Translation property**: Directional BoD is invariant to the chosen mean normalisation method *i.e.* $D(\mathbb{1}, \mathbf{y} + \alpha g_y; g_y) = D(\mathbb{1}, \mathbf{y}; g_y) - \alpha$ for $\alpha \in \mathbb{R}_+$.

Directional Benefit of the doubt (DBoD)

The directional BoD adds the following properties to the BoD model:

DBoD properties

4. **Noncompensability property**: the directional vector g_y rewards units along a generic direction (not only along the bisector direction) by penalizing units far from the chosen direction; therefore full compensability can be seen as a special case when $g_y = \mathbb{1}$.
5. **Translation property**: Directional BoD is invariant to the chosen mean normalisation method *i.e.* $D(\mathbb{1}, \mathbf{y} + \alpha g_y; g_y) = D(\mathbb{1}, \mathbf{y}; g_y) - \alpha$ for $\alpha \in \mathbb{R}_+$.

Our proposal

Social Indicators Research, 2014 in press

"*Non-compensability in composite indicators: a robust directional frontier method*", with E. Fusco and C. Mazziotta

Our proposal is based on the logical union of two different proposals that aim to bypass two crucial drawbacks: on the one hand the lack of estimates robustness - *Robust BoD* - and, on the other hand, the full compensability between simple indicators - *directional BoD*.

Directional robust BoD

Calculate iteratively the sample subset of size m (for $b = 1, ..., B$ times) and for each b iteration the directional distance for the single unit from the maximum values can be defined as:

$$\widetilde{D}_m^b(\mathbb{1}, \mathbf{y}; g_y) = \sup\{\beta | (\mathbb{1}, \mathbf{y} + \beta g_y) \in \widetilde{\Psi}_m\}, \forall b = 1, \ldots, B$$

Directional robust BoD

Following Cazals, 2002, finally, we can approximate the order-m directional distance estimator - even in a Shepard formulation - by computing the empirical mean over B:

$$\widehat{D}_m(\mathbb{1}, \mathbf{y}; g_y) = 1/\widehat{E}(\widetilde{D}_m^b(\mathbb{1}, \mathbf{y}; g_y))$$

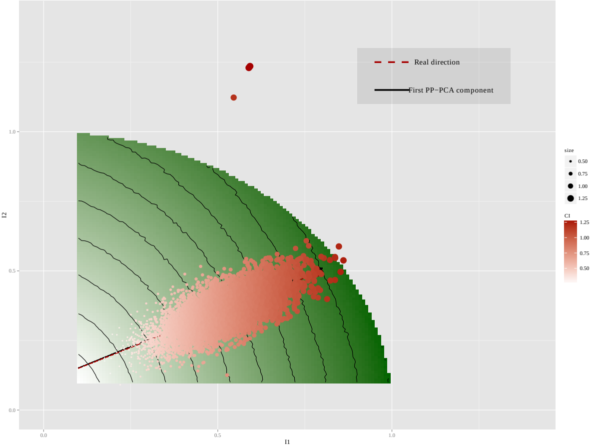
Directional robust BoD

The Robust directional BoD adds the following property to the classical BoD model and to directional BoD:

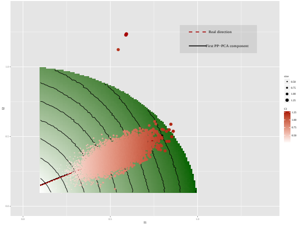
RDBoD properties

6. **External robustness property**: Robust directional BoD allows to remove outliers influence on the CI ranking obtained with directional frontier methods.

Directional (RPca) robust BoD



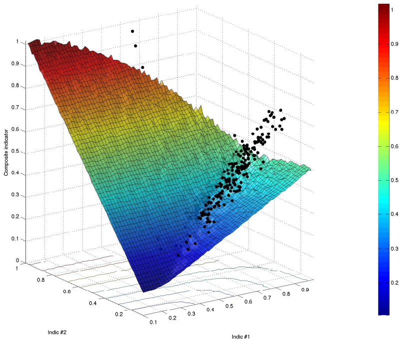
Directional (RPca) robust BoD



Directional (RPca) robust BoD bypass:

- sensitivity to the outliers;
- non-compensability;
- lack of consideration about the rates of substitution between simple indicators.

Directional (RPca) robust BoD



Directional (RPca) robust BoD

>||

Directional (RPca) robust BoD properties

- Weights are endogenously determined;
- Weighting scheme is the highest possible;
- Aggregation function is weak monotone;
- Non compensability is not imposed (PCA direction);
- External (frontier) robustness property;
- Internal (direction) robustness property.

Data

In our application we have used data from the annual survey conducted by the INVALSI in 2011 for children attending the 5^{th} grade primary school.

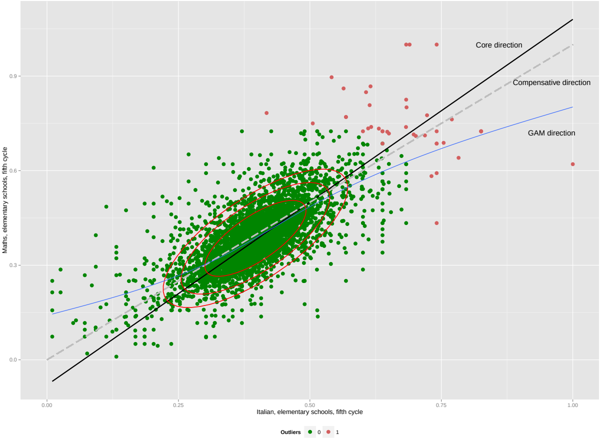
Although survey is census, covering 188.827 children and representing the entire population of children from the last year of junior elementary schools in Italy, for the sake of simplicity, in this first proof, we have used the average evaluation scores in mathematics and Italian by school (5152 schools) only for one year.

Data

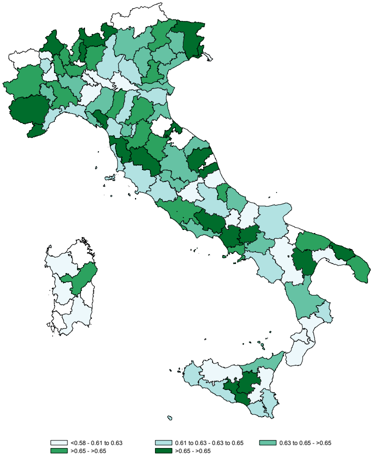
In our application we have used data from the annual survey conducted by the INVALSI in 2011 for children attending the 5^{th} grade primary school.

Although survey is census, covering 188.827 children and representing the entire population of children from the last year of junior elementary schools in Italy, for the sake of simplicity, in this first proof, we have used the average evaluation scores in mathematics and Italian by school (5152 schools) only for one year.

Directional robust BoD CI



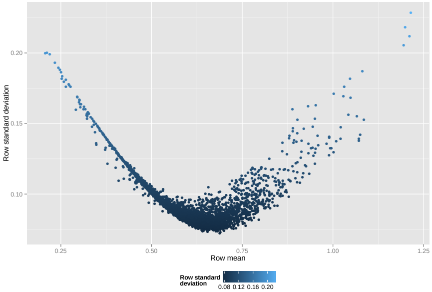
Territoriality of the robust directional indicator



Difference in educational attainment between Northern provinces and Southern ones.

Uncertainty analysis

Finally, we have applied uncertainty analysis with respect to changes in the proposed frontier models - BoD, RBoD, directional BoD and Robust PCA directional BoD.



Mean and standard deviation varying models (BoD, RBoD, directional BoD and Robust PCA directional BoD)

Compind package

Compind is a R package containing functions to enhance several approaches to the composite indicators.

Weighting techniques (1)

ci_bod: Benefit of the Doubt approach (BoD)
ci_bod_dir: Directional Benefit of the Doubt (D-BoD)
ci_bod_var_w: Variance weighted Benefit of the Doubt
ci_rbod: Robust Benefit of the Doubt (RBoD)
ci_rbod_dir: Directional Robust Benefit of the Doubt (D-RBoD)

Compind package

Compind is a R package containing functions to enhance several approaches to the composite indicators.

Weighting techniques (2)

ci_factor: Weighting method based on Factor Analysis
ci_mean_geom: Weighting method based on geometric aggregation
ci_mpi: Mazziotta-Pareto Index (MPI) method
ci_wroclaw: Wroclaw Taxonomic Method

E-mail

Francesco Vidoli *francesco.vidoli@uniroma3.it*
fvidoli@gmail.com

Compind package is available @:
`http://fvidoli.weebly.com/compind.html`
Collaborations in the R package final development and/or improvements are very welcome!