Book of Short Papers SIS 2020

Editors: Alessio Pollice, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891910776

Contents

Specialized sessions

Solicited Sessions

Contributed papers and Posters

Models and methods - Categorical, Ordinal, Rank Data 1135

Models and methods – Regression...1206

BOD – min range: A Robustness Analysis Method for Composite Indicators

Intervallo BOD – minimo: un metodo per l'analisi di robustezza degli indicatori compositi

Emiliano Seri, Leonardo Salvatore Alaimo, Vittoria Carolina Malpassuti

Abstract Composite indicators are a useful instrument to represent in a synthetic and easy to read way a phenomenon or a specific reality. However, synthesizing indicators using aggregative approaches inevitably involves a loss of information due to the reduction in size of the original matrix. So, in composite indicators construction, the robustness analysis phase is very important to validate the synthetic construct and to attempt to recover some of the lost information. This paper aims to present an interval, useful for the robustness analysis of composite indicators, between the best and the worst performance obtainable from the synthetic construct. The two margins of the range are obtained: for the upper bound, using the Benefit Of The Doubt approach, that is an application for composite indicators construction of the linear programming technique, Data Envelopment Analysis; and, for the lower bound, it has been used the minimum between the considered elementary indicators. The obtained range will comprehend almost every other synthetic measure that can be calculated with other aggregative methods for the considered matrix of indicators, so, our methodological proposal is useful to see the distance from the best and the worst obtainable cases.

Abstract *Gli indicatori compositi sono uno strumento utile per rappresentare in modo sintetico e di facile lettura un fenomeno o una realtà specifica. Tuttavia, la sintesi di indicatori mediante approcci aggregativi comporta inevitabilmente una perdita di informazioni a causa della riduzione delle dimensioni della matrice originale. Pertanto, nella costruzione di indicatori compositi, la fase di analisi della robustezza è molto importante per validare il costrutto sintetico e tentare di recuperare alcune delle informazioni perse. In questo articolo presentiamo un intervallo utile per l'analisi di robustezza degli indicatori compositi, compreso tra la performance*

Emiliano Seri
Sapienza Università di Roma, e-mail: emiliano.seri@uniroma1.it

Leonardo Salvatore Alaimo
Italian National Institute of Statistics - Istat, e-mail: leonardo.alaimo@istat.it

Vittoria Carolina Malpassuti
Sapienza Università di Roma, e-mail: vittoriacarolina.malpassuti@uniroma1.it

migliore e quella peggiore ottenibili dal costrutto sintetico. L'estremo superiore del range è calcolato usando l'approccio del Benefit Of The Doubt, che è un applicazione per la costruzione di indicatori compositi della tecnica di programmazione lineare Data Envelopment Analysis. Per l'estremo inferiore è stato usato il minimo tra gli indicatori elementari considerati. Il range ottenuto comprenderà quasi ogni altra misura sintetica che che può essere calcolata con altri metodi aggregativi per la matrice di indicatori. Di conseguenza, questa proposta metodologica permette di osservare la distanza dal caso migliore e peggiore.

Key words: Composite indicators, Robustness analysis, BOD, Performance interval

1 Introduction

Constructing a composite indicator involves different subjective choices [4]. Subjectivity is an indispensable element in any measurement process; its presence does not make the process arbitrary [1]. But, different choices lead to different results; furthermore, a synthetic measure is useful to give an easy-to-read information about the phenomenon, but implicitly involves a loss of information due to the transition from a multi-dimensional to a uni-dimensional representation. For those reasons, the robustness analysis is a very important step in the composites construction. Its aim is to verify how the synthetic construct reacts to the different choices made to measure it. So, it is an instrument to validate the results and a way to find out some of the information loss during the previous phases of construction [5].

In this paper, we propose a new method for the robustness analysis of composite indicators based on the Benefit of the Doubt (BOD) approach. This method generates an upper bound and a lower bound to find the best and the worst performance that could be obtained aggregating a system of indicators. The space between the two bounds, it is going to include almost all the other constructs made with other mean-based aggregative methods, like: minimum, harmonic mean, geometric mean, arithmetic mean, quadratic mean, cubic mean and maximum. Thus, the proposed method could be useful to compare each of the cited techniques, observing their distance from the best and the worst performances. It is also useful to shrink the range of variation of the construct to observe the position of each unit in a real range of variation, in which each position can be reached by each unit.

2 Data envelopment analysis

DEA is a linear programming technique, useful to measure the relative efficiency of decision making units (DMU) on the basis of multiple inputs and outputs [7]. The efficiency of a set of variables can be adapted to construct a synthetic indicator using an input-oriented DEA.[3]. The model assumes N inputs and M outputs for each of the I units. For the i^{th} unit, the inputs are represented by an array x_i and the outputs are represented by an array q_i. A first problem's formulation is the following: for each unit i the ratio of all the outputs over all the inputs is defined as

$$f(\mathbf{u},\mathbf{v}) = \frac{\mathbf{u}^{'}\mathbf{q}_i}{\mathbf{v}^{'}\mathbf{x}_i} \tag{1}$$

where **u** is an array of output weights and **v** is an array of input weights. The model seeks to maximize f, which represents the efficiency of the unit i^{th}, under the constraints that all the efficiency measures must be less than or equal to 1 and that the weights must be positive. This linear program is solved by assigning to each unit the most favourable weight. However, this formulation has infinite solutions of the form $(\delta\mathbf{u},\delta\mathbf{v})$ for $\delta > 0$. To avoid it, the following constraint is introduced:

$$\mathbf{v}^{'}\mathbf{x}_i = 1$$

Thus the problem can be written as:

$$\begin{aligned} &\max_{\mathbf{u},\mathbf{v}} \mathbf{u}^{'}\mathbf{q}_i \\ &\text{s.t. } \mathbf{v}^{'}\mathbf{x}_i = 1 \\ &\quad \mathbf{u}^{'}\mathbf{q}_1 - \mathbf{v}^{'}\mathbf{x}_1 \leq 0 \\ &\quad\quad \vdots \\ &\quad \mathbf{u}^{'}\mathbf{q}_I - \mathbf{v}^{'}\mathbf{x}_I \leq 0 \\ &\quad \mathbf{u},\mathbf{v} \geq 0 \end{aligned} \tag{2}$$

Lastly we call the model DEA with no input if, for the array $\mathbf{x}_i$, we have $\mathbf{x}_i = (1,...,1)^{'}$. The model computes the weights so that the unit under investigation is valued as best as possible. The weights can differ from unit to unit and range between 0 and 1.

3 Benefit of doubt approach

The BOD method is an application of DEA that can be used to composite indicators construction[2][6]. The synthetic measure is expressed as the weighted sum of the elementary indicators relatively to a benchmark; more precisely, is defined as the performance of the single unit divided for the performance of the benchmark:

$$CI_c = \frac{\sum_{q=1}^{Q} I_{cq} w_{cq}}{I_{cq}^{*}} \tag{3}$$

where I_{cq} is the normalized score of the q$^{\text{th}}$ basic indicator ($q = 1,\dots,Q$) for the unit c$^{\text{th}}$ ($c = 1,\dots,C$) and w_{cq} is the corresponding weight. The benchmark I^*_{cq} is defined as:

$$I^*_{cq} = \max_{I_{c\in[1\dots C]}} \sum_{q=1}^{Q} I_{cq} w_{cq} \tag{4}$$

The optimal set of weights (if exist) guarantees that each unit is associated to the best possible position compared to all the others. The optimal weights are obtained by solving equation 3 where the weights are non-negative. The result will be between zero and 1.

4 BOD – min range of performance: a robustness analysis method for composite indicators

In this paper, we propose a new interval of performance for composite indicators using the BOD approach for the upper bound and the minimum for the lower bound. With the obtained range, it is possible to compare the performance of the synthetic construct, calculated using most of the mean-based aggregative methods, for each considered unit, with the best and the worst obtainable performance. The proposed interval can be calculated on a set of indicators standardized using the Min-Max method, that bring each indicator on a range between 0 and 1. Our interval, shrink this range to lead to a real *photograph* of the units for the considered phenomenon and solve the common problem in composite indicators construction, that the results of each unit can have an evaluative function only relatively to those of the others.

We have chosen the BOD approach to make the upper bound because, as explained in Section 3, with this method it is possible to calculate the best reachable performance, associating to each variable the optimal weight for each unit. The lower bound corresponds to the minimum, because it is not reasonable that a unit should want to drop below its worst level.

Let's see now an application on a set of seven variables about the quality of work in the Italian regions in 2017 (Table 1). The selected indicators are:

- share of employed persons with temporary jobs for at least 5 years (V1);
- share of employees with below 2/3 of median hourly earnings (V2);
- share of employed people aged 15-64 years working over 60 hours per week (V3);
- share of employed persons not in regular occupation (V4);
- involuntary part-time (V5);
- share of employed persons who feel satisfied with their work (V6);
- Share of employed persons who feel their work insecure (V7).

Reg.	$V1$	$V2$	$V3$	$V4$	$V5$	$V6$	$V7$	Min	BOD
PIEMONTE	11.6	7.1	23.1	10.8	5.4	7.5	4.6	0.57	1.00
VALLE D'AOSTA	16.9	5.0	21.2	10.4	4.4	7.7	6.3	0.61	1.00
LIGURIA	17.4	5.8	23.4	12.1	6.0	7.3	6.1	0.29	0.91
LOMBARDIA	10.7	4.9	21.5	10.3	5.4	7.4	5.1	0.43	1.00
TRENTINO	19.8	4.3	18.4	9.6	3.4	7.8	4.2	0.64	1.00
VENETO	11.9	5.3	23.6	8.9	3.4	7.5	5.6	0.57	1.00
FRIULI	14.9	5.1	24.6	10.6	4.3	7.5	6.4	0.53	0.98
EMILIA ROMAGNA	16.7	5.0	25.3	10.0	4.9	7.5	6.4	0.48	0.97
TOSCANA	14.8	6.1	25.9	10.9	6.6	7.4	6.4	0.43	0.91
UMBRIA	12.1	6.4	31.7	12.9	6.2	7.4	7.3	0.00	0.95
MARCHE	14.1	6.3	27.5	10.3	5.1	7.4	6.8	0.32	0.91
LAZIO	21.2	8.6	28.4	15.6	7.5	7.3	6.7	0.25	0.70
ABRUZZO	17.7	9.5	30.0	15.9	5.6	7.2	8.4	0.13	0.75
MOLISE	20.9	9.6	25.5	15.6	8.0	7.5	5.9	0.38	0.73
CAMPANIA	21.0	16.4	23.7	20.1	8.2	7.1	8.4	0.00	0.68
PUGLIA	21.8	16.5	23.3	16.7	8.3	7.3	9.2	0.07	0.69
BASILICATA	23.4	10.1	27.9	14.4	6.7	7.2	9.4	0.04	0.60
CALABRIA	31.3	17.7	26.6	22.3	10.6	7.2	9.6	0.00	0.38
SICILIA	35.7	17.4	22.2	19.8	10.8	7.1	9.4	0.00	0.71
SARDEGNA	11.3	12.5	20.7	15.2	8.9	7.4	8.9	0.13	1.00

Table 1 Application of BOD – min range of performance: quality of work indicators; lower bound (Min); upper bound (BOD); Italian regions; year 2017.

First, we standardized the basic indicators and changed the polarity to those with present *negative polarity*[1]. The normalization method must be the relative indices with respect to the variation range, commonly called Min-Max:

$$r_{ij} = \frac{x_{ij} - \min_i \left\{x_{ij}\right\}}{\max_i \left\{x_{ij}\right\} - \min_i \left\{x_{ij}\right\}} \tag{5}$$

Where x_{ij} is the value of i^{th} unit for the j^{th} indicator. We calculated the upper bound with the BOD approach (see equation 3); the lower bound was set equal to the minimum of each basic indicator normalized. The bounds obtained are reported in Table 1. We aggregated the standardized indicators r_{ij} using harmonic mean, geometric mean, arithmetic mean, quadratic mean, cubic mean; we report all the results in Figure 1[2].

[1] Polarity is the sign of the relation between the indicator itself and the phenomenon. All the indicators must have positive polarity, i.e. an increase in the normalized indicators corresponds to an increase in the composite index. To invert polarity, we use a linear transformation [1].

[2] Note that in geometric mean, 0 values have been changed with 0,001 in order to have explanatory results of the considered phenomena.

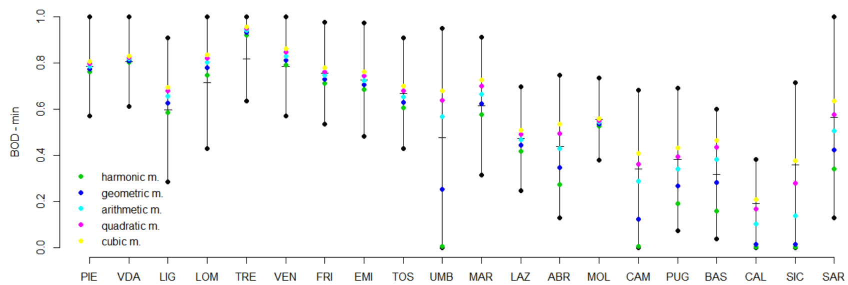

Fig. 1 BOD–min range of performance

5 Conclusions

Looking at Figure 1, it is clear that using the proposed method the range of variation of each unit has been widely narrowed or anyway made more truthful of the reality. It is also easier to compare the results obtained from different aggregation methods and to highlight where each of them is collocated in respect of the obtainable minimum and maximum.

References

1. Alaimo, L.S.: Complexity of Social Phenomena: Measurements, Analysis, Representations and Synthesis. Unpublished Doctoral Dissertation, University of Rome" La Sapienza", Rome, Italy (2020)
2. Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: An Introduction to 'Benefit of the Doubt'Composite Indicators. Social Indicators Research **82**(1), 111–145 (2007)
3. Coelli, T.: A Guide to DEAP version 2.1: A Data Envelopment Analysis (Computer) Program. Centre for Efficiency and Productivity Analysis, University of New England, Australia **96**(08) (1996)
4. Commission, J.R.C.E., et al.: Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD publishing (2008)
5. Maggino, F.: Complexity in Society: From Indicators Construction to Their Synthesis, vol. 70. Springer (2017)
6. Rogge, N., Cherchye, L., Moesen, W., Van Puyenbroeck, T.: 'Benefit of the Doubt'Composite Indicators. In: European Conference on Quality in Survey Statistics (2006)
7. Terzi, S., Pierini, A., et al.: Data Envelopment Analysis (DEA) Assessment of Composite Indicators of Infrastructure Endowment. Rivista di statistica ufficiale **17**(1), 5–18 (2015)