A VEA Benefit-of-the-Doubt model for the HDI

Panagiotis Ravanos[1] and Giannis Karagiannis[2]

[1]PhD candidate, Department of Economics, University of Macedonia, Thessaloniki, GREECE; email: ravanos@uom.edu.gr

[2]Professor, Department of Economics, University of Macedonia, Thessaloniki, GREECE; email: karagian@uom.edu.gr; tel. ++30 2310 891 759

ABSTRACT: The Benefit-of-the-Doubt (BoD) model for constructing composite indicators does not account for prior judgments concerning the relative importance of individual indicators. In this paper we combine the BoD model with Value Efficiency Analysis (VEA) to propose a new formulation that incorporates external preferences and value judgments by means of a "model" decision-making unit which serves as the benchmark for all other units. We explore the potential of the proposed model by using it to re-estimate the United Nations Human Development Index (HDI).

KEYWORDS: Value Efficiency Analysis, Benefit-of-the-Doubt, Most Preferred Solution, Human Development Index.

AKNOWLEDGMENT: We would like to thank two anonymous reviewers for useful comments and suggestions. The first author acknowledges financial support by the Hellenic Foundation for Research and Innovation (HFRI) and the General Secretariat for Research and Technology (GSRT), under the HFRI PhD Fellowship grant (GA. no. *698*).

September 2020

A VEA Benefit-of-the-Doubt model for the HDI

1. Introduction

The Benefit-of-the-Doubt (BoD) as an input-oriented Data Envelopment Analysis (DEA) model with a single constant input, mainly used for constructing composite indicators (CI) (see OECD, 2008) but also applied to several multi criteria decision making problems such as supplier selection, inventory classification, quality perception, preference voting, location selection, etc. (see Karagiannis (2020) and the references therein), has the advantages and disadvantages of DEA models. Among them is the flexibility of weights that may vary across DMUs and indicators in such a way so as to maximize the overall achievement of each evaluated Decision-Making Unit (DMU). This flexibility may in some cases imply that DMUs are rated as efficient by doing well only on a single performance dimension. In such a case, zero weights are assigned to all but one indicator. However, more often, DMUs are evaluated based only on a subset of the considered indicators (most notably those in which they perform relatively better), implying that the rest have no effect on the CI. As this subset of indicators may differ across DMUs, it makes their multilateral comparison difficult or even inappropriate.[1,2]

To avoid this kind of problems when implementing the BoD model several researchers have restricted weight flexibility by imposing different type of restrictions. These may take the form of (*i*) absolute weight bounds (see e.g. Rogge and Self, 2019), (*ii*) relative restrictions in the form of assurance regions (see e.g. Gaaoul and Khalfallah, 2014), (*iii*) ordinal ranking of indicators' importance (see e.g. Cherchye *et al.* 2007a), and (*iv*) pie shares (see e.g. Cherchye *et al.,* 2007b; Gonzalez *et al.,* 2018). Such forms of weights restrictions require experts, stakeholders or social planners to set the absolute or relative bounds or shares, a task that can be proved to be difficult and time-consuming since usually experts may have diverging opinions regarding the relative importance of indicators.

An alternative way to incorporate value judgments in BoD is by means of Value Efficiency Analysis (VEA), developed by Halme *et al.* (1999).[3] This alternative has been suggested in OECD (2008, p. 92), where it is noted that "the benchmark could also be determined by a hypothetical decision maker … who would locate the target in the efficiency frontier with the most preferred combination of individual indicators'', but to the best of our knowledge it has not been implemented so far. In VEA, the views of an expert, a decision maker (DM) or a supervising agency are reflected in the selection of a "model" DMU that determines the most-preferred solution (MPS) from their point of view. This *alternative in gauging preferences* might be proven to be more appealing to Decision Makers, as the latter might be more keen on selecting one DMU to serve as a benchmark, rather than engage in the task of selecting weight restriction bounds (Korhonen *et al.,* 2002).[4] The choice of the "model" DMU restricts the weights that the evaluated DMUs can select by determining a preferred range of indicator mixes.

In this paper we use the VEA-BoD model to re-estimate the United Nations (UN) Human Development Index (HDI) using data for 2015. For these purposes we rely on the notion of uniformity, namely the intension for equal prioritization among the considered indicators, to select the "model" country. Based on this objective and normative argument, countries with a relatively balanced prioritization among health, education and income would be promoted as peers for improving human capabilities in the rest of the countries. The latter may need a shift of focus towards policies aimed at improving their most deprived human development dimensions.

The rest of the paper proceeds as follows: In the next section we briefly discuss the conventional BoD model and we introduce the VEA-BoD model. In the third section, we present the empirical results for the HDI based on the VEA-BoD model. Concluding remarks follow in the last section.

2. The Conventional and the VEA BoD models

The BoD is a model facilitating the (linear) aggregation of a number of quantitative indicators into a single CI when exact knowledge of the weights is not available.[5] The model endogenously selects the best possible weights for each DMU, assuming implicitly that the DMUs attach less (more) importance to those indicators on which they perform relatively weak (strong) compared to the other evaluated units. The model is a special case of the input-oriented constant-returns-to-scale DEA model with a single constant input that takes the value of one for all DMUs (see Karagiannis, 2020). Its multiplier and envelopment forms are given as:

$$\begin{array}{llll} \max\limits_{u_j^k} & \sum\limits_{k=1}^{K} u_j^k y_j^k & \qquad \min\limits_{\mu_h^k} & \sum\limits_{h=1}^{K} \lambda_h^k \;\; (= \theta^k) \\ s.t. & \sum\limits_{j=1}^{J} u_j^k y_j^h \leq 1 \quad h = 1, \ldots, K & \qquad s.t. & \sum\limits_{h=1}^{K} \lambda_h^k y_j^h \geq y_j^k \quad j = 1, \ldots, J \\ & u_j^k \geq 0 \qquad\quad j\; = 1, \ldots, J & & \lambda_h^k \geq 0 \qquad\quad h = 1, \ldots, K \end{array} \tag{1}$$

where y refers to the component indicators, u to their relative weights (multipliers), θ to the efficiency score, λ to the intensity variables, j=1,…,J is used to index indicators and k=1,…,K to index DMUs.

Expert judgments, DM's mandates or public opinion and social planner norms regarding the performance of the evaluated DMUs are not *a priori* incorporated in the conventional form of the BoD model in (1). As a result, each DMU has the benefit-of-the-doubt in the selection of its relative weights in order to maximize the value of its CI. This allows DMUs that dominate all others in a single indicator to be rated as efficient even though they preform relatively weak in terms of all other indicators. The CI values resulting from (1) are thus the most optimistic for each DMU.

VEA takes into account DM's preferences or public opinion about desired norms and managerial or social goals by means of a pseudo-concave value function (i.e., an indifference curve) that becomes tangent to the DEA efficient frontier at a point referred to as the MPS. This point, ultimately selected by a DM or a supervising agency, corresponds to a virtual or real DEA-efficient DMU, which is viewed as the "model" DMU having the most preferred mix of inputs and outputs. The VEA frontier is constructed by extending towards the axes the hyperplanes of the DEA efficient facets intercepting at the MPS, a process that naturally results in efficiency scores that are lower or equal to those of the conventional DEA model. This is depicted in Figure 1 for the case of two indicators. Selecting for example DMU B as the MPS extends facets AB and BC towards the axes, creating the VEA frontier (the blue kinked line). This defines a range of preferred mixes given between rays OA and OC. As a result, the DEA benchmark profiles complying with the desired norms are now limited to facets AB and BC. For all inefficient DMUs with projection points in these two facets, the CI value that results from the VEA-BoD model is equal to that of the conventional BoD model while inefficient DMUs projected elsewhere on the BoD frontier and thus using a mix outside of the preferred range are "penalized" and their CI value is less than that obtained from the conventional BoD model.

Following Halme *et al*. (1999), the VEA formulation of the BoD model in (2) is given, in its multiplier and envelopment form, as follows:

$$
\begin{array}{ll}
\max\limits_{u_j^k} & \sum\limits_{k=1}^{K} u_j^k y_j^k \\
s.t. & \sum\limits_{j=1}^{J} u_j^k y_j^h \leq 1 \quad h = 1,\dots,K, h \neq MPS \\
& \sum\limits_{j=1}^{J} u_j^k y_j^h = 1 \quad h = MPS \\
& u_j^k \geq 0 \qquad j = 1,\dots,J
\end{array}
\qquad
\begin{array}{ll}
\min\limits_{\mu_h^k} & \sum\limits_{h=1}^{K} \lambda_h^k \\
s.t. & \sum\limits_{h=1}^{K} \lambda_h^k y_j^h \geq y_j^k \quad j = 1,\dots,J \\
& \lambda_h^k \geq 0 \quad h = 1,\dots,K, h \neq MPS \\
& \lambda_h^k \; free \quad h = MPS
\end{array}
\qquad (2)
$$

As we can see there are only slight differences between (1) and (2): in the multiplier form, the restriction corresponding to the MPS is turned from an inequality to a strict equality in order the range of acceptable weights to be restricted into those that are optimal for the "model" DMU. This in turn restricts the preferred mix range to be between rays OA and OC in Figure 1, if DMU B is selected to be the MPS. In the envelopment form, the equality restriction corresponds to removing the non-negativity restriction from the intensity variable corresponding to the MPS, thus forcing it to be peer for all evaluated DMUs.

In practical settings, the most crucial step in VEA is the selection of the MPS, as it affects the resulting frontier and, consequently, the derived efficiency scores. Nevertheless, no general rule of thumb exists for selecting the MPS, but several alternatives have been proposed. These involve the choice of either a real (usually DEA-efficient) DMU or a virtual combination of DEA-efficient DMUs. The latter case can be operationalized as long as the combined DMUs share at least one common facet, in which case the resulting VEA frontier expands only those common facets. If the selected MPS units do not share a common facet, then their average will not be DEA efficient and thus its DEA efficient peers would be used as the MPS. The same is true if the DM selects an existing inefficient DMU as the MPS. For example, in Figure 1, the average of DMUs B and C lies on facet BC and limits the preferred mixes between rays OB and OC. On the other hand, the average of DMUs A and E, denoted as AE, is inefficient and its peers (i.e. DMUs B and C) are used as the MPS.

Many of the proposed MPS selection alternatives involve the subjective judgments of a DM. Such examples include simply "apple-picking" a DMU (Halme *et al.*, 2014), choosing the DMU performing the best in a particular model dimension (Marshal and Shortle, 2005), and using interactive multiobjective algorithms (see Halme *et al.,* 1999). This inherent subjectivity makes the task of selecting an MPS less transparent and raises concerns as it might compromise the evaluation process in the case of malevolent DMs who wish to curb the results in favor of certain DMUs.[6]

Nevertheless, there are other, relatively objective alternatives, the use of which can make the MPS selection as transparent as possible from the viewpoint of stakeholders or the public. They can also provide compromise solutions in cases where a DM is absent or unable to point at a preferred DMU and in cases of disagreement among a board of DMs.[7] These include: *first,* averaging inputs and outputs over more than one DMUs selected by the same or different DMs and using the resulting artificial DMU as a compromise solution MPS (Korhonen *et al.*, 2002). *Second,* using a participatory approach such as the Analytic Hierarchy Process (AHP, Saaty, 1980) or the Budget Allocation Process (BAP, see OECD, 2008).[8] *Third,* using an established criterion for ranking the DEA-efficient DMUs such as superefficiency (Andersen and Petersen, 1993) scores (see Halme and Korhonen, 2015). *Fourth,* selecting the MPS on the basis of a strong normative argument, which mandates what the preferred performance of DMUs "ought to be" within the particular evaluation context. This alternative is followed in the empirical application of this paper, in which the MPS selection is based on the notion of uniformity. *Fifth,* using an ideally-performing virtual DMU, i.e. one that utilizes the maximum observed indicator values across DMUs. As such an Ideal DMU usually lies beyond the DEA efficient frontier, its DEA-efficient peers should be identified through a superefficiency model and be used as the MPS its place.[9]

3. Re-estimating the Human Development Index

3.1. Variables and modeling choices

In this section we use the VEA-BoD model to re-estimate UN's HDI for the year 2015. The HDI is a CI reflecting country achievements in human development, the underpinnings of which can be found in Sen's capability approach.[10] The capability approach views people as the main recipients (the "ends") of the development process and development itself as a process which expands people's choices, thereby placing the emphasis on "what people get from development, not only what they put into it" (Anand and Sen, 2000, p. 84). The HDI contains three basic and universally valued capabilities, namely to be knowledgeable, to live a healthy life, and to have adequate command over resources in order to enjoy a decent standard of living (Anand and Sen, 2000). Ever since the first Human Development Report in 1990, there has been a quite long literature regarding (*i*) the choice of the capabilities to be included in the index, (*ii*) their relevant proxy variables, (*iii*) the normalization of these variables, (*iv*) the choice of the aggregator function, and (*v*) the selection of aggregation weights. We next consider these steps in sequence.

First, several important aspects of human development such as environmental sustainability, political rights and freedom, income or gender inequality as well as other demographic factors are not included in the current specification of the HDI (see e.g. Desai (1991), Ranis *et al.* (2006) and Klugman *et al.* (2011)) and several attempts (see e.g. Hicks (1997), Sagar and Najam (1998), Neumayer (2001) and Herrero *et al.* (2019)) have been made to incorporate them. For the purpose of this paper we keep the current HDI specification for both the capabilities considered and the variables used to proxy them. That is, we use life expectancy at birth as a proxy for living a healthy life, the arithmetic average of the mean and the expected years of schooling as a proxy for being knowledgeable, and the logarithm of GNI per capita in 2011 $ PPP to a proxy for the standard of living.[11]

Second, the HDI is based on the min-max normalization with the goalposts (minimum and maximum) values for each indicator being those of 1994.[12] This has been criticized as the normalized indicators and the resulting CI depend on the choice of these min and max values (Noorbakhsh, 1998a; Panigrahi and Sivramkrishna, 2002). Several alternatives have been proposed: in particular, Mazumdar (2003) and Chakravarty (2003) used sample minimum and maximum goalpost values, Noorbakhsh (1998a; b) employed the z-score normalization, Herrero *et al.* (2012) relied on the distance-to-the-leader normalization (i.e., divide each variable with its maximum value across countries), while Luque *et al.* (2016) suggested a normalization with two reference points, an aspiration point reflecting the desired level and a reservation point beyond which performance is not acceptable.[13] For the

BoD model, the distance-to-the-leader is the appropriate normalization in order to ensure unit invariance, required in any DEA model. Notice that unit invariance is violated with the min-max normalization (Filippetti and Peyrache, 2011).

Third, the UNDP initially (1990-2009) used an arithmetic aggregation function but as of 2010 it has switched to a geometric aggregation function.[14] The main reason behind this switch is the implied perfect substitutability between the component indicators in the arithmetic aggregation (see Klugman *et al.* (2011) and the references therein). However, as argued by Ravallion (2012), arithmetic aggregation implies perfect substitutability between the considered indicators but not between capabilities due to the logarithmic transformation of the income variable. Both the BoD and the VEA-BoD models assume arithmetic aggregation of the component indicators. Previous attempts to estimate the HDI by means of a multiplicative BoD model, with the logarithm instead of the actual values of the component indicators, have been criticized (Tofallis, 2014) as they do not satisfy unit invariance.[15]

Fourth, choice of the weights for aggregating the component indicators is probably the most debated step. The UNDP used equal weights, which implies that each indicator and its corresponding capability are of equal importance to human development (Klugman *et al.,* 2011). Even though this has been criticized as arbitrary (see e.g. Desai, 1991), there are several studies that support the equal weights scheme either on the basis of the principle of parsimony (Hopkins, 1991) or empirical evidence based on Principal Components Analysis (Owgang, 1994; Noorbakhsh, 1998a, b; Owgang and Abdou, 2003; Nguefack-Tsangue *et al.,* 2011), expert opinion surveys (Chowdhury and Squire, 2006), or statistical criteria from Information Theory (Stapleton and Garrod, 2007). On the other hand, several other studies have called for variable weights: Srinivasan (1994, p. 240) noted that relative weights "need not be the same across individuals, countries, and socioeconomics groups". Along the same line, Fukunda-Parr (2003, p.306) referred that "the relative importance of capabilities can vary with social context-from one community or country to another and from one point of time to another", Klugman *et al.* (2011, p. 261) suggested that in an ideal situation the relative weights "should be traced either to individual preferences, some collective social choice process or to a strong normative argument'', and Noorbakhsh (1998a, p. 593) argued that "an alternative way is to derive these weights from the data".

Following these considerations, several studies favored the use of variable weights that vary either across countries or across both indicators and countries. In the former case, Lind (2010) used revealed preferences to obtain such a set of common across countries but unequal weights,[16] Pinar *et al*. (2013, 2017) employed non-parametric stochastic dominance techniques, Karagiannis and Karagiannis (2020) relied on Shannon entropy, Despotis (2005), Hatefi and Torabi (2010) and Sayed *et al*. (2015) used goal programming, and Tofallis (2013) a regression model with the CI obtained from the conventional BoD model as dependent variable and the component indicators as independent variables. In the latter case, Mahlberg and Obensteiner (2001) used a normalized variant of the BoD model,[17] Despotis (2005) employed the conventional BoD model in (1), Bougnol *et al*. (2010) considered a BoD model with weighted restrictions, and Lozano and Gutierrez (2008) relied on the range-adjusted-measure (RAM) BoD model.

The VEA-BoD model used in this paper allows weights to vary across both indicators and countries but only within a certain range, which is determined by the "model" country. This is in line with Sen (1999, p. 78) who mentioned that weights for each capability can be chosen from a specified range on which there is agreement. We base our choice of the "model" country on the notion of uniformity. Following Mishra and Nathan (2018), uniformity implies that, between two countries with the same average attainment across indicators, the CI should favor the most balanced country, i.e. the country with the minimum dispersion across indicators. Palazzi and Lauri (1998, p.196) also favored such a choice by postulating that "there are explicit or potential endogenous forces working to move the values of the single variables towards a more balanced relation". In Figure 1, by choosing a country such as C, which displays balanced performance, implies that the preferred range of mix lies between rays OB and OD. For any country within this range, the conventional and the VEA-BoD model scores coincide while the farther a country's mix is from those between OB and OD rays, the lower its VEA-BoD score will be compared to its BoD score.

We consider three alternatives for choosing a balanced MPS country: *first*, the country that is ranked first in 2015 UN HDI, namely Norway; *second*, the country with the minimum dispersion across indicators, namely Lithuania; and *third*, an artificial country with all indicators set at 0.5.[18] Norway is also a BoD-efficient country and thus it can serve as MPS by its own. The other two alternatives, namely

Lithuania and the artificial country, are BoD-inefficient but share the same peers, namely Norway and Australia, and thus result in the same VEA-BoD model.

3.2. Empirical Results

The empirical results for the conventional BoD model in (1) and the two VEA-BoD models with Norway and Norway and Australia as MPS are presented in Table 1. The average CI value for the BoD model is 0.861, with five countries receiving CI scores of one, namely Norway, Australia, Singapore, Hong-Kong and Qatar.[19] As it was expected, VEA-BoD results on average into relatively lower scores and less DMUs as being efficient. From the BoD-efficient countries, Hong-Kong drops from the list when Norway is chosen as MPS while Hong-Kong and Qatar drop from the list when Norway and Australia are chosen as MPS. Qatar had an extremely unbalanced mix that implicitly places higher importance on the "command over resources" indicator, for which it ranks 1st compared to education (82nd) and longevity (39th). Hong-Kong, on the other hand, implicitly places a higher importance on the longevity indicator, for which it ranks 1st (see Table 2).

The HDI frequency distributions of the three estimated models are portrayed in Figure 2. Based on Banker tests (see e.g. Banker *et al.*, 2010) we can confirm that both VEA-BoD distributions differ, in a statistically significant way, from that of the conventional BoD. The same is not however true when we are comparing the two VEA-BoD distributions to each other, for which there are no statistically significant differences. This is also evident from the average rank shift, given as $R = (\frac{1}{N})\sum_{j=1}^{N}\left|rank_1^j - rank_2^j\right|$ (Saisana *et al.,* 2005), which is roughly 1.4 positions when we are comparing the two VEA-BoD scores while it is around 9 positions when we comparing the BoD and the VEA-BoD scores (see Table 3). In addition, relatively large rank shifts (more than ten positions) occur for 70 and 68 countries respectively when we are comparing the BoD with the two VEA-BoD scores, whereas it is limited to only three countries when comparing the two VEA-BoD scores. This rank variability is all but uniform across countries: country-specific Mean Absolute Deviation in ranks (Çillingirtürk and Koçak, 2018) in Figure 2(b) shows that countries in the middle rank positions, as identified by the BoD model, exhibit relatively higher rank variability compared to top or bottom ranked countries. In order to verify the latter, we constructed rolling country subsamples of size 40. More specifically, the first subsample consisted of the top-40 ranked countries by the BoD model. From that, we constructed the second subsample by dropping the country ranked 1st and including the country ranked 41th. Each following subsample was constructed likewise, and the last one consisted of the bottom-40 ranked countries by the BoD model. The average rank shift between pairs of models for each subsample is plotted in Figure 2(c), where we see that the average rank variability between the BoD and the two VEA-BoD scores is considerably higher in subsamples including mostly middle-ranked countries, whereas this pattern is absent when we are comparing the two VEA-BoD scores, which on average displays minor rank differences.

The above results suggest that the VEA-BoD model has a moderate impact on HDI scores compared to the conventional BoD model but a significant impact on country rankings, which is magnified for middle-ranked countries. This finding may however be sensitive to the choice of MPS, for which so far we have based on the notion of uniformity, i.e., relatively balanced achievements. We next examine the sensitivity of our results to MPS choices that go beyond balanced achievements. In the absence of a general consensus, potential candidates for MPS might be all countries found to be BoD-efficient: namely, Norway, Australia, Singapore, Hong-Kong and Qatar. Summary results of the VEA models using each or pairs of the above countries as MPS are given in Table 4 and their frequency distributions are portrayed in Figure 3, where are plotted against the BoD distribution.

Consider first the cases where each of the BoD-efficient countries is chosen as the MPS. The differences between the BoD and the VEA-BoD scores depends on the extent of the preferred mixes implied by each MPS, which in turn is closely related to the number of times an efficient country is used as a peer. For example, in the case of Hong-Kong, which serves as a peer for 145 of the 193 BoD-inefficient countries (see Table 2), the differences between the conventional BoD HDI and the VEA-BoD HDI using Hong-Kong as the MPS are minimal (see Figure 3) and in fact, statistically insignificant (see Table 5). The same is essentially true when Australia (which serves as a peer 49 times) or Singapore (which served as a peer 54 times) are considered as the MPS.[20] On the other hand, MPS choices such as Norway and Qatar result in statistically significant differences between the conventional BoD and the VEA-BoD models (see Figure 3 and Table 5). Norway's mix displays, as we have mentioned, a relatively high balance that is absent from the majority of countries while Qatar's mix placed considerably higher importance on the "command over resources" indicator.[21]

In Figure 4, we plot the values of Mean Absolute Deviation for the VEA-BoD models using each BoD-efficient country as the MPS. The average value of 4.59 (dashed line) indicates that varying the MPS can induce relatively moderate shifts in ranking. Nevertheless, rank variability appears to be higher for countries in the middle of the rankings, whereas top and bottom ranked countries appear to be relatively less affected by the chosen MPS.

We also consider cases with (reasonable) pairs and triads of BoD-efficient countries as MPS (see Table 4 and Figure 3) and several interesting findings emerge from these results:[22] *first,* VEA-BoD models with two or three countries forming the MPS resemble more or less the behavior of the most extreme in terms of mix country. See for example the VEA-BoD models based on Norway alone and on Norway and Hong-Kong as the MPS. *Second*, VEA-BoD models with two or three countries forming the MPS do not differ in a statistically significant sense with VEA-BoD models with the most extreme (in terms of mix) of the two countries as the MPS but they statistically differ from VEA-BoD models with any other BoD-efficient country as the MPS (see Table 6). *Third*, pairing countries with similar preferred ranges of mixes to form the MPS (e.g. Hong-Kong and Singapore) seem to result in negligible differences compared to VEA-BoD models with each one as the MPS (see Figure 3 and Table 6).

This demonstrated sensitivity of the models' estimates to the selected MPS might pose difficulties to select among alternative evaluation results those that will be ultimately presented to stakeholders or the public and used for policy-designing purposes. As this situation is similar to the initial selection of the MPS, a first option for indecisive practitioners or DMs would be to use the evaluation results stemming from an objective and transparent MPS selection alternative among those presented in the previous section, such as AHP or BAP. A second option would be to select the evaluation results that fit the most the DMs' perceptions of "good" and "bad" performing DMUs in the sample. Lastly, selection can also be based on the variability between the BOD and VEA-BoD estimates. For example, DMs wishing for the least (most) rank variability between HDI estimates of the two models would select the evaluation results based on Australia (Norway and Qatar) as the MPS.

Last but not least, we examine the sensitivity of our results with respect to different modeling choices regarding the education indicator. Several authors (e.g., Mahlberg and Obensteiner, 2001; Lozano and Gutiérrez, 2008; Sayed *et al.*, 2015) suggested using the mean and the expected years of schooling as separate indicators while Herrero *et al*. (2012) proposed using only the expected years of schooling.[23] The comparative results concerning these two alternative formulations of the education variable are presented in Table 7. There seem to be no significant differences with our benchmark formulation of using the average of the mean and the expected years of schooling. The most notable difference is that now the VEA-BoD models are based on different countries for the MPS, namely Norway, Australia and Singapore and Norway and Australia. Nevertheless, this change affects only slightly the HDI values.

4. Concluding Remarks

In this paper, we use the VEA formulation of the BoD model, which integrates DM or expert opinion to the conventional BoD model through the selection of a "model" DMU that serves as benchmark for all evaluated units. The "model" DMU defines a preferred range of mixes and for DMUs operating within (outside of) this preferred range, VEA-BoD scores are equal to (lower than) the BoD scores. The proposed model is sensitive to the selection of the MPS. Models with MPS BoD-efficient units featuring a wider range of mixes (as indicated by the times they are used as peers) or with mixes closer to the majority of the evaluated DMUs result in VEA-BoD scores that differ less from the BoD scores. In addition, VEA-BoD scores tend to differ more (less) from each other if their MPS units have highly dissimilar (similar) mixes, while VEA-BoD models with more than one DMU as the MPS resemble closely the pattern of the most extreme of those DMUs. In our empirical application regarding the HDI, the VEA-BoD model causes moderate changes regarding the scores but significant changes in country rankings compared to the conventional BoD model, especially for middle-ranked countries which displayed on average higher rank variability compared to top and bottom performing countries.

The proposed model can be applied to a wide range of social and economic indicators and it could be useful to both the evaluated entities as well as DMs, since it allows pursuing the best-possible aggregation weights but to the extent that these weights comply with managerial goals. Nevertheless, the proposed model is not without limitations, as its current form inherits certain deficiencies of the conventional BoD and DEA models. More specifically, it is sensitive to the presence of outliers- which could also affect the MPS selection- and it fails to account for the effect of background 'contextual' variables which are not under the direct control of DMUs but can create favorable operating conditions for some of them and unfavorable for others. Hence, the present work could be further extended through a robust order-m framework (see Cazals *et al.,* 2002) to mitigate the impact of outlying observations and through a conditional DEA framework (see Daraio and Simar, 2005) in order to account for the effect of contextual variables. In addition, the present model can be readily extended to cases where DMUs select the worst possible aggregation weights by means of the inverted BoD model. In such a case, managerial goals regarding the *least preferred* mixes would be considered.

Table 1: CI estimates and efficient countries, BoD and VEA models.

Model	BoD	VEA (a)	VEA (b/c)
composite indicator estimates			
maximum	1.000	1.000	1.000
minimum	0.611	0.574	0.574
average	0.861	0.834	0.833
median	0.880	0.856	0.856
standard deviation	0.094	0.106	0.106
Q1	0.793	0.748	0.747
Q3	0.926	0.913	0.913
efficient countries			
# of efficient countries (CI=1)	5	4	3
efficient country names	Norway, Australia, Singapore, Hong-Kong, Qatar	Norway, Australia, Singapore, Qatar	Norway, Australia, Singapore

Table 2: HDI and indicator values and ranks, BoD-efficient DMUs

DMU	Longevity	education index	income	HDI 2015	times as peer[b]
Norway	0.971 (17)	0.905 (6)	0.945 (6)	0.949 (1)	23
Australia	0.981 (9)[a]	1.000 (1)	0.906 (20)	0.939 (2)	49
Singapore	0.989 (4)	0.803 (37)	0.957 (2)	0.925 (6)	54
Hong-Kong (HK)	1.000 (1)	0.811 (30)	0.926 (10)	0.917 (12)	145
Qatar	0.931 (39)	0.689 (82)	1.000 (1)	0.856 (35)	30

Notes: (a). Numbers in parentheses denote the country's rank position, (b). The last column denotes the time that each efficient country serves as a peer for inefficient ones in model (2) calculations.

Table 3: Average rank shifts, large rank shifts, and statistical tests of equality between pairs of BoD, VEA(a) and VEA(b/c) models

Pair	Average rank shift	Large rank shifts (>10 positions)	Banker's F1	Banker's F2
BoD-VEA(a)	9.202	70	1.214**	1.438***
BoD-VEA(b/c)	8.952	68	1.229**	1.467***
VEA(a)-VEA(b/c)	1.388	3	1.012	1.020

Note: The F1 (F2) test assumes an exponential (half-normal) distribution of the inefficiency scores, following Banker et al (2010). Both tests compare the initial DEA to the respective VEA distribution and the alternative hypothesis for all cases was that the respective VEA model exhibits higher inefficiency scores. Three, two and one stars denote statistical significance at 1%, 5% and 10% respectively.

Table 4: Composite indicator estimates, model (3), alternative MPS selections

MPS selection	maximum	minimum	average	median	standard deviation	Q1	Q3	efficient countries (CI=1)
Norway	1.000	0.574	0.834	0.856	0.106	0.748	0.913	4
Australia	1.000	0.596	0.852	0.875	0.100	0.780	0.923	4
Singapore	1.000	0.605	0.852	0.872	0.096	0.786	0.921	5
Hong-Kong (HK)	1.000	0.611	0.856	0.875	0.097	0.790	0.924	3
Qatar	1.000	0.589	0.836	0.850	0.099	0.759	0.905	3
Norway-Australia	1.000	0.574	0.833	0.856	0.106	0.747	0.913	3
Norway-Singapore	1.000	0.574	0.833	0.856	0.106	0.747	0.913	4
Norway-Qatar	1.000	0.565	0.823	0.840	0.109	0.739	0.905	3
Australia-Singapore	1.000	0.589	0.843	0.864	0.103	0.763	0.916	4
Australia-HK	1.000	0.596	0.848	0.871	0.102	0.772	0.917	3
Singapore-HK	1.000	0.605	0.848	0.866	0.099	0.776	0.918	3
Singapore-Qatar	1.000	0.589	0.835	0.850	0.099	0.757	0.905	3
Norway-Australia-Singapore	1.000	0.574	0.831	0.856	0.107	0.747	0.913	3
Norway-Singapore-Qatar	1.000	0.565	0.822	0.840	0.109	0.729	0.905	3
Australia-Singapore-HK	1.000	0.589	0.841	0.862	0.104	0.762	0.914	3

Table 5: Average rank shifts, large rank shifts, and statistical tests of equality, model (2) vs. model (3) for alternative MPS selections

MPS selection	Average rank shift	Large rank shifts (>10 positions)	Banker's F1	Banker's F2
Norway	9.202	70	1.214**	1.438***
Australia	3.362	4	1.074	1.159
Singapore	4.532	13	1.068	1.118
Hong-Kong (HK)	4.144	14	1.043	1.091
Qatar	10.537	84	1.195**	1.341**
Norway-Australia	8.952	68	1.229**	1.467***
Norway-Singapore	8.872	66	1.222**	1.452***
Norway-Qatar	12.559	88	1.309***	1.630***
Australia-Singapore	6.128	30	1.148*	1.300**
Australia-HK	4.963	14	1.103	1.222*
Singapore-HK	6.133	24	1.101	1.190
Singapore-Qatar	10.101	80	1.204**	1.360**
Norway-Australia-Singapore	8.654	65	1.241**	1.495***
Norway-Singapore-Qatar	12.271	87	1.316***	1.644***
Australia-Singapore-HK	6.899	34	1.163*	1.334**

Note: The F1 (F2) test assumes an exponential (half-normal) distribution of the inefficiency scores, following Banker *et al.* (2010). Both tests compare the initial DEA to the respective VEA distribution and the alternative hypothesis for all cases was that the respective VEA model exhibits higher inefficiency scores. Three, two and one stars denote statistical significance at 1%, 5% and 10% respectively.

Table 6: Average rank shifts and statistical tests of equality, model (3) for alternative MPS selections

	average rank shift														
MPS selection	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.Norway		8.617	6.170	11.452	4.016	1.388	0.830	3.835	5.596	10.122	8.346	4.314	2.559	3.665	6.771
2.Australia			4.660	4.505	10.463	7.888	8.213	12.197	4.702	2.814	5.229	9.973	7.580	11.888	5.186
3.Singapore				6.261	7.037	5.739	5.766	9.186	3.202	5.654	3.133	6.388	5.090	8.846	3.984
4.Hong-Kong (HK)					12.777	10.649	11.037	14.745	6.633	2.777	4.309	12.170	9.968	14.426	5.691
5.Qatar						4.702	4.335	3.883	7.548	12.053	9.117	1.074	5.383	3.755	8.511
6.Norway-Australia							1.239	4.723	4.686	9.309	7.447	4.681	1.266	4.457	5.777
7.Norway-Singapore								4.303	5.074	9.686	7.910	4.080	1.803	3.931	6.229
8.Norway-Qatar									8.920	13.670	11.160	4.404	5.511	0.681	9.883
9.Australia-Singapore										5.250	3.676	6.878	3.878	8.559	1.516
10.Australia-HK											3.894	11.447	8.649	13.351	4.298
11.Singapore-HK												8.383	6.649	10.777	2.723
12.Singpore-Qatar													4.755	4.032	7.798
13.Norway-Australia-Singapore														5.117	4.936
14.Norway-Singapore-Qatar															9.511
15.Australia-Singapore-HK															
	Banker's F1 test														
1.Norway		0.885	0.879	0.859	0.984	1.012	1.006	1.078	0.945	0.908	0.907	0.992	1.022	1.083	0.958
2.Australia	0.806		0.994	0.971	1.112	1.143*	1.137	1.218**	1.068	1.026	1.025	1.121	1.155*	1.224**	1.083
3.Singapore	0.778	0.965		0.977	1.119	1.150*	1.144*	1.225**	1.075	1.033	1.031	1.128	1.162*	1.232**	1.089
4.Hong-Kong (HK)	0.758	0.941	0.975		1.146*	1.178*	1.171*	1.255**	1.100	1.057	1.056	1.155*	1.190**	1.261**	1.115
5.Qatar	0.932	1.157	1.199	1.229*		1.028	1.022	1.095	0.960	0.923	0.921	1.008	1.039	1.101	0.974
6.Norway-Australia	1.020	1.265*	1.312**	1.345**	1.094		0.994	1.065	0.934	0.898	0.896	0.980	1.010	1.071	0.947
7.Norway-Singapore	1.010	1.253*	1.299**	1.331**	1.083	0.990		1.071	0.939	0.903	0.901	0.986	1.016	1.077	0.952
8.Norway-Qatar	1.133	1.406***	1.458***	1.494***	1.215*	1.111	1.122		0.877	0.843	0.841	0.920	0.949	1.005	0.889
9.Australia-Singapore	0.904	1.121	1.162	1.192	0.969	0.886	0.895	0.797		0.961	0.959	1.049	1.082	1.146*	1.014
10.Australia-HK	0.849	1.054	1.092	1.120	0.911	0.833	0.841	0.749	0.940		0.998	1.092	1.126	1.193**	1.055
11.Singapore-HK	0.827	1.026	1.064	1.091	0.887	0.811	0.819	0.730	0.916	0.974		1.094	1.127	1.195**	1.056
12.Singpore-Qatar	0.946	1.173	1.216*	1.247*	1.014	0.927	0.936	0.834	1.046	1.113	1.143		1.031	1.092	0.966
13.Norway-Australia-Singapore	1.039	1.289**	1.337**	1.370**	1.115	1.019	1.029	0.917	1.150	1.224*	1.256*	1.099		1.060	0.937
14.Norway-Singapore-Qatar	1.143	1.418***	1.470***	1.507***	1.226*	1.121	1.132	1.009	1.265*	1.346**	1.381**	1.209*	1.100		0.884
15.Australia-Singapore-HK	0.927	1.151	1.193	1.223*	0.995	0.909	0.918	0.818	1.026	1.092	1.121	0.981	0.892	0.811	
	Banker's F2 test														

Note: The F1 (F2) test assumes an exponential (half-normal) distribution of the inefficiency scores, following Banker et al (2010). The tests in this table compare the respective row and column VEA model distributions and the alternative hypothesis for all cases is that the row-wise VEA model exhibits higher inefficiency scores. Three, two and one stars denote statistical significance at 1%, 5% and 10% respectively

Table 7: Robustness tests: Alternative formulations for the education indicator

	formulation		
	average of two education variables	two separate education variables	only expected years of schooling
	BoD		
Average	0.861	0.864	0.861
# of countries with CI=1	5	8	5
	VEA (a) (MPS: country ranked the highest in the official HDI-ranking of 2015)		
MPS	Norway	Norway	Norway
Average	0.834	0.840	0.836
# of countries with CI=1	4	6	4
	VEA (b) (MPS: country with minimum dispersion)		
MPS	Lithuania*	Ireland*	Ireland*
	(Norway-Australia)	(Norway-Australia-Singapore)	(Norway-Australia-Singapore)
Average	0.833	0.836	0.833
# of countries with CI=1	3	4	3
	VEA (c) (MPS: virtual country with all normalized indicators equal to 0.5*)		
MPS	(Norway-Australia)	(Norway-Australia)	(Norway-Australia)
Average	0.833	0.839	0.835
# of countries with CI=1	3	4	3

Note: An asterisk denotes an inefficient country based on the BoD model. The countries in parentheses below it are its efficient peers and are used in its place as the MPS.

Figure 1: composite indicator construction based on BoD and VEA-BoD models

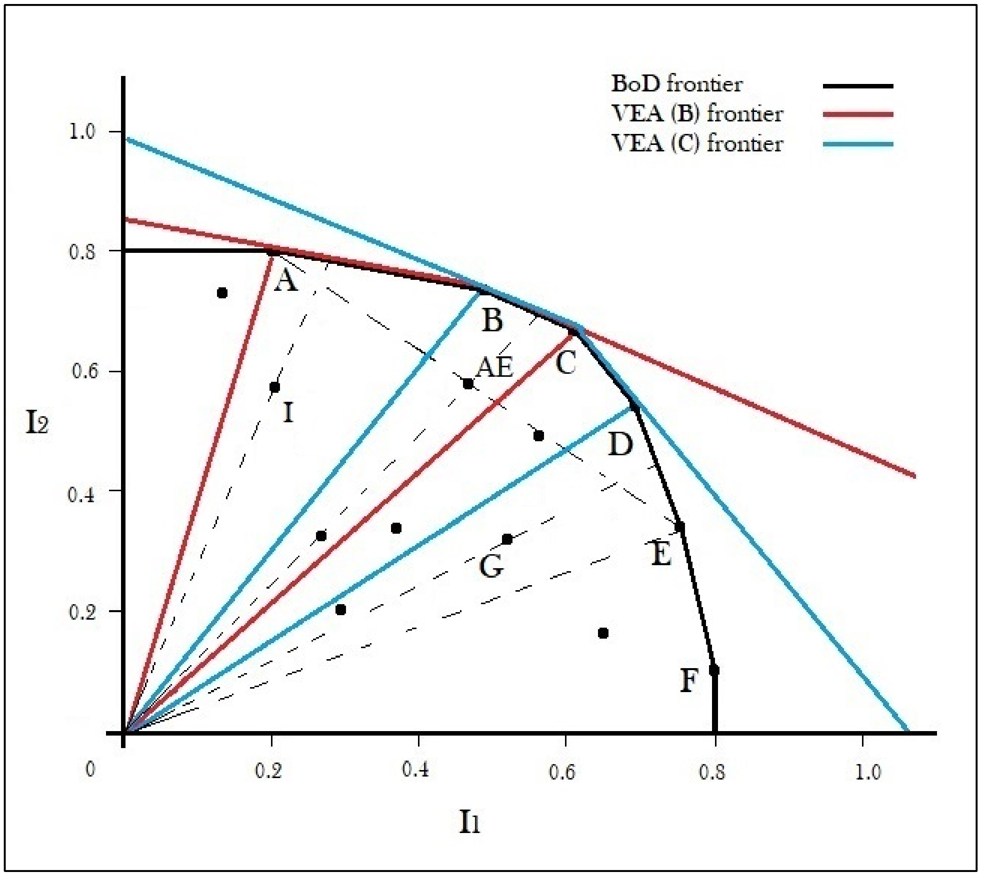
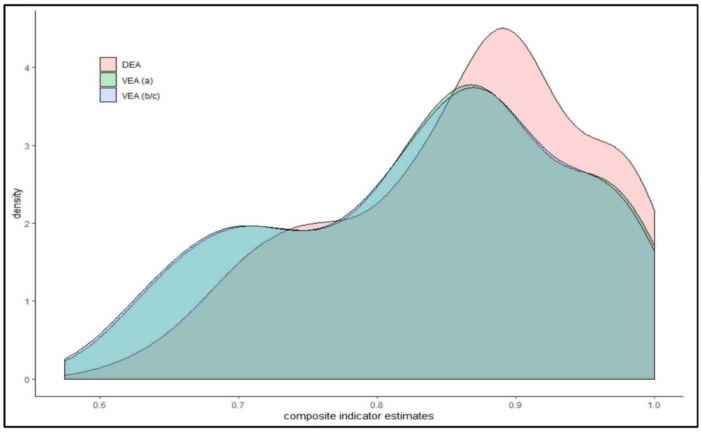
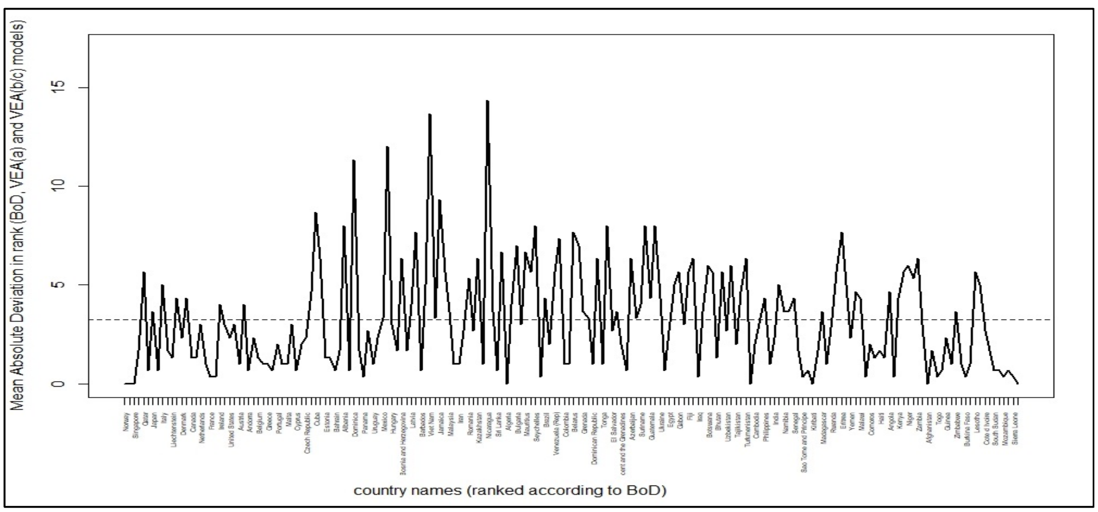


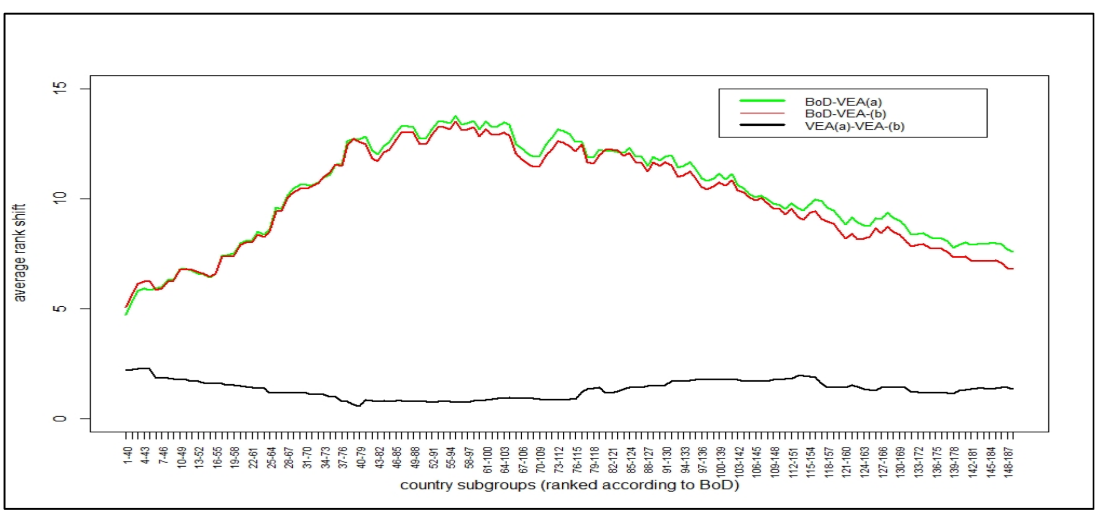
Figure 2: Comparison of distributions and rankings, BoD, VEA (a) and VEA (b/c) HDI.



Panel (a): kernel density distributions, BoD, VEA (a) and VEA (b/c) HDI.

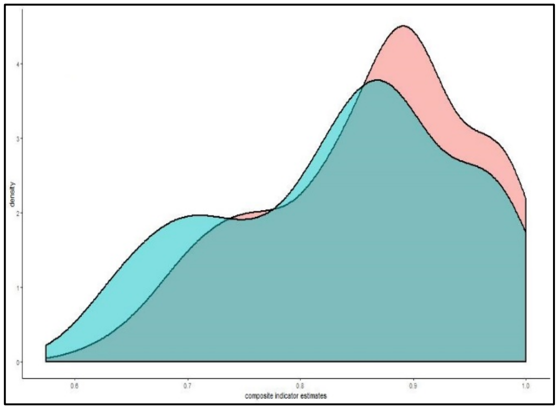


Panel (b): Country specific Mean Absolute Deviation between BoD, VEA (a) and VEA (b/c) HDI.

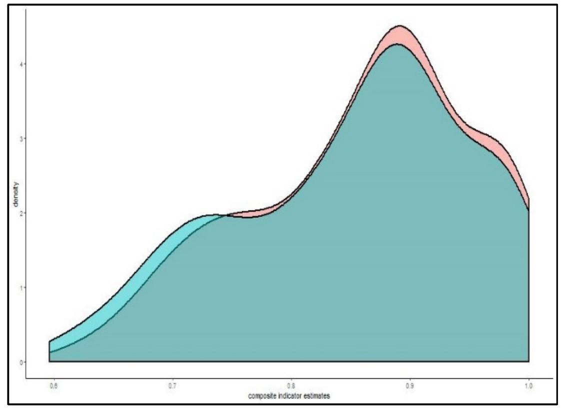


Panel (c): Average rank shift between pairs of models, moving country subsamples (n=40)

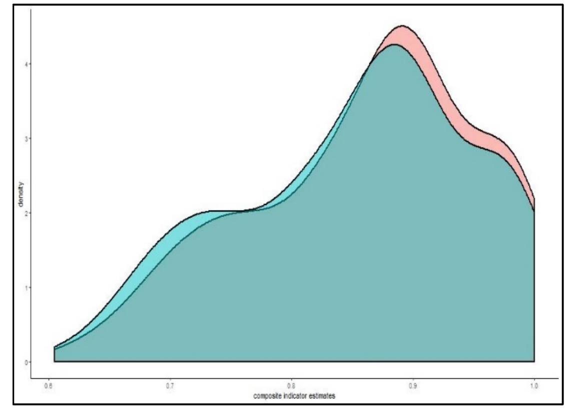
Figure 3: Kernel density distributions, model (2) vs. model (3), alternative MPS selections.



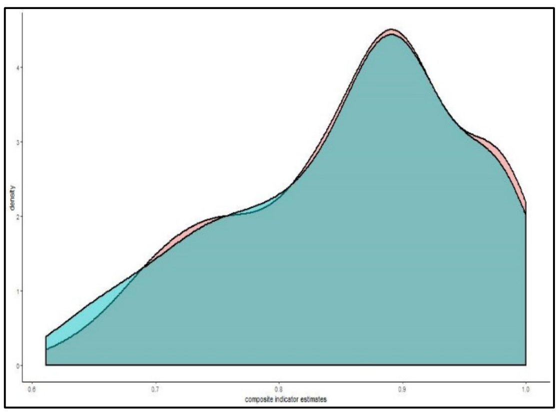
(1) Norway



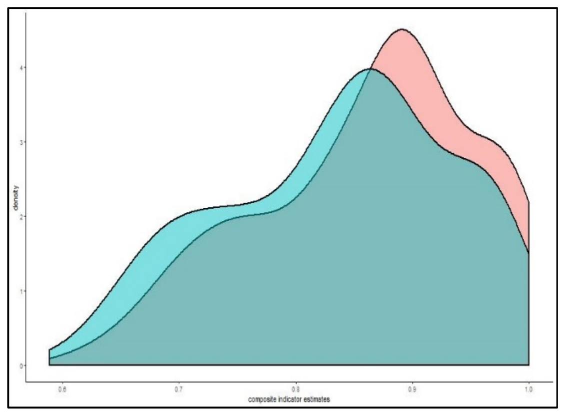
(2) Australia



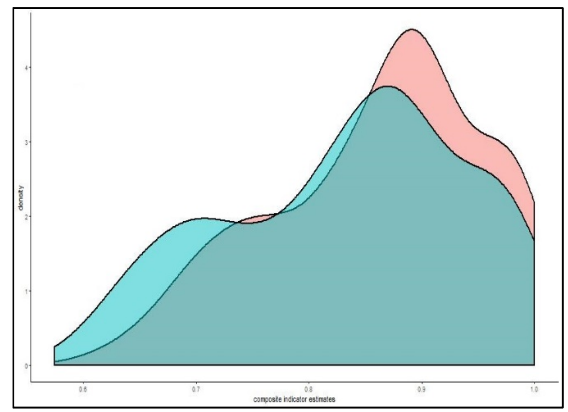
(3) Singapore



(4) Hong-Kong (HK)

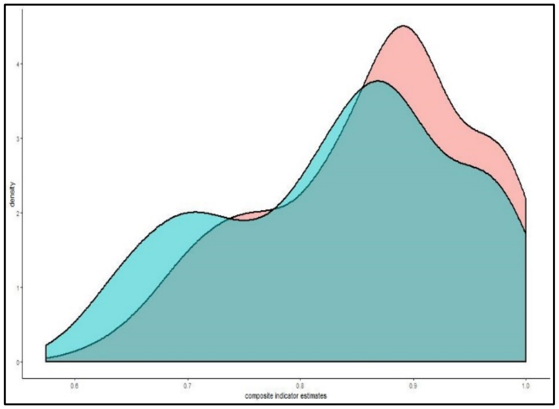


(5) Qatar

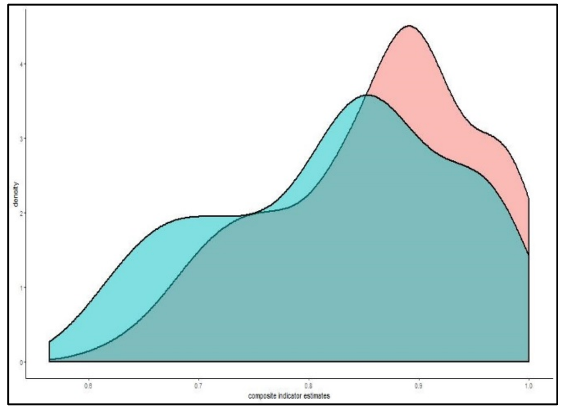


(6) Norway-Australia

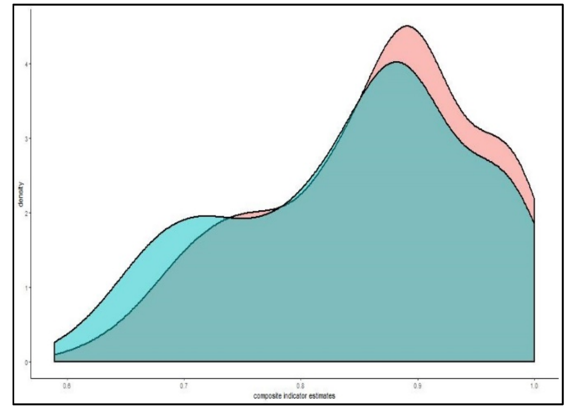
Figure 3 (cont.)



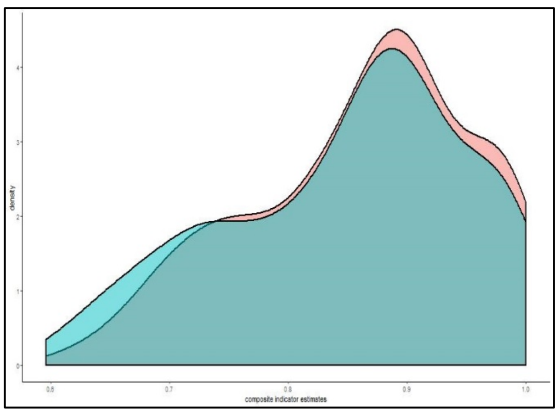
(7) Norway-Singapore



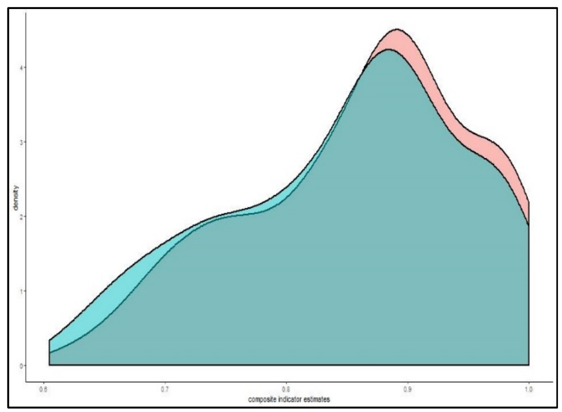
(8) Norway-Qatar



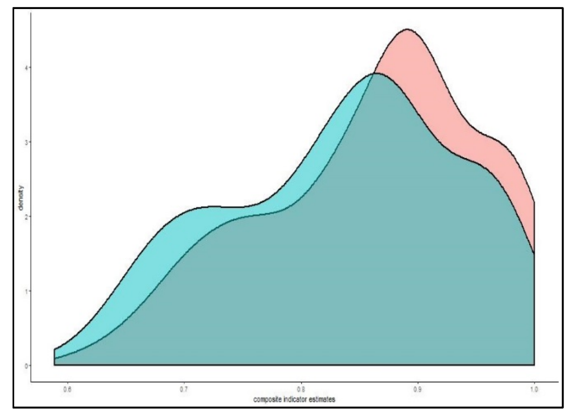
(9) Australia-Singapore



(10) Australia-HK

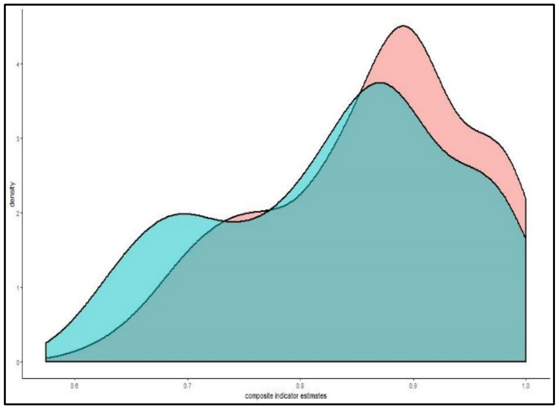


(11) Singapore-HK

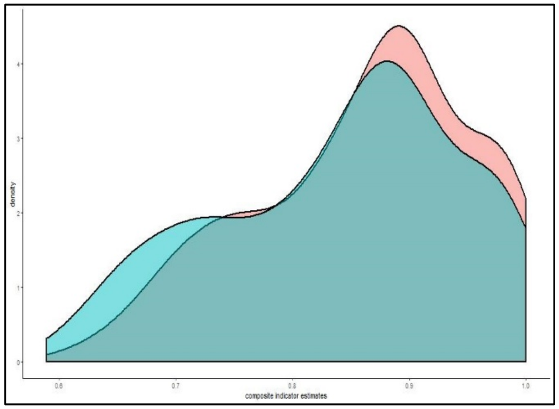


(12) Singapore-Qatar

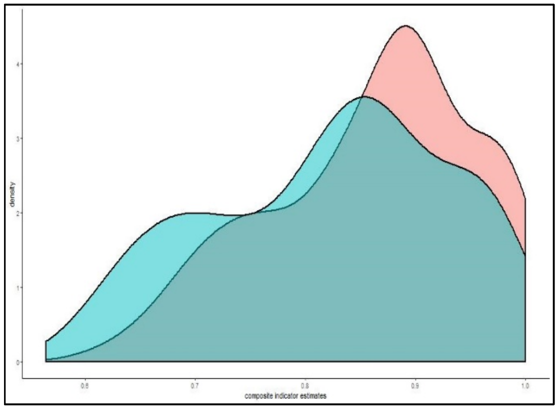
Figure 3 (cont.)



(13) Norway-Australia-Singapore

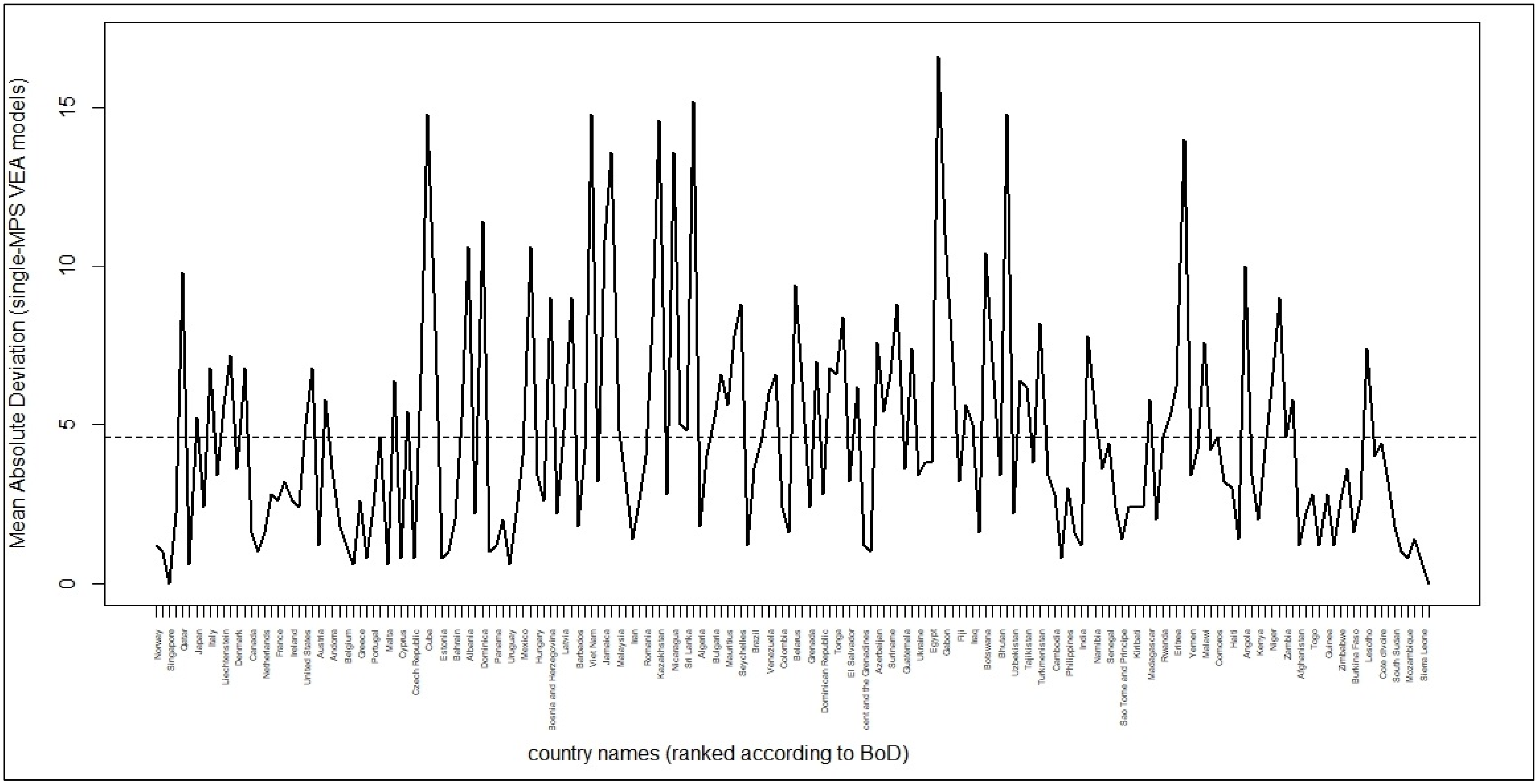


(14) Australia-Singapore-HK



(15) Norway-Singapore-Qatar

Figure 4: country-specific Mean Absolute Deviation between VEA models based on different BoD-efficient DMUs acting as MPS



References.

Anand S, Sen A (2000). The income component of the human development index. *Journal of Human Development and Capabilities* 1:83-106.

Banker RD, Zheng Z, Natarajan R (2010). DEA-based hypothesis tests for comparing two groups of decision-making units. *European Journal of Operational Research* 206:231–238

Blancard S, Hoarau JF (2013). A new sustainable human development indicator for small island developing states: a reappraisal from data envelopment analysis. *Economic Modelling* 30:623-635.

Bougnol ML, Dulá JH, Lins MPE, de Silva ACM (2010). Enhancing standard performance practices with DEA. *Omega* 38:33-45.

Canning D, French D, Moore M (2013). Non-parametric estimation of data dimensionality prior to data compression: the case of the human development index. *Journal of Applied Statistics*, 40:1853-1863.

Cazals C, Florens J-P, Simar L (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics* 106:1-25.

Chakravarty S (2011). A Reconsideration of the tradeoffs in the new Human Development Index. *Journal of Economic Inequality* 9:471-474.

Chakravarty SR (2003). A generalized human development index. *Review of Development Economics* 7:99-114.

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

Cherchye L, Lovell CAK, Moesen W, van Puyenbroeck T (2007a). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review* 51:749–779.

Cherchye L, Moesen W, Rogge N, van Puyenbroeck T (2007b). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research* 82: 111–145.

Chowdhury S, Squire L (2006). Setting weights for aggregate indices: An application to the commitment to development index and human development index. *Journal of Development Studies* 42:761-771.

Çillingirtürk AM, Koçak H (2018). Human Development Index (HDI) Rank-Order Variability. *Social Indicators Research* 137:481–504.

Daraio C, Simar L (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis* 24:93-121.

Desai MJ (1991). Human development: concepts and measurement. *European Economic Review* 35:350-357.

Despotis DK (2005). A reassessment of the Human Development Index via Data Envelopment Analysis. *Journal of Operational Research Society* 56:969-980.

El Gibari S, Gómez T, Ruiz F (2019). Building composite indicators using multicriteria methods: a review. *Journal of Business Economics* 89:1-24.

Filippetti A, Peyrache A (2011). The patterns of technological capabilities of countries: A dual approach using composite indicators and Data Envelopment Analysis. *World Development* 39:1108–1121.

Fukuda-Parr S (2003). The human development paradigm: Operationalizing Sen's ideas on capabilities. *Feminist Economics* 9:301-317.

Gaaloul H, Khalfallah S (2014). Application of the "Benefit-of-the-Doubt" approach for the construction of a digital access indicator: A revaluation of the "Digital Access Index". *Social Indicators Research* 118:45-56.

Gonzalez E, Carcaba A, Ventura, J (2018). Weight constrained DEA measurement of the quality of life in Spanish municipalities in 2011. *Social Indicators Research*, 136:1157-1182.

Greco S, Ishizaka A, Tasiou M, Torrisi G (2019). On the methodological framework of composite indices: a review of the issues of weighting, aggregation, and robustness. *Social Indicators Research* 141:61-94.

Halme M, Joro T, Korhonen P, Salo S, Wallenius T (1999) A value efficiency approach to incorporating preference information in data envelopment analysis. *Management Science* 45:103-115.

Halme M, Korhonen P (2015). Using value efficiency analysis to benchmark nonhomogeneous units. *International Journal of Information Technology & Decision Making* 14:727-747.

Halme M, Korhonen P, Eskelinen J (2014). Non-convex value efficiency analysis and its application to bank branch sales evaluation. *Omega* 48:10-18.

Hatefi SM, Torabi SA (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics* 70:114-120.

Herrero C, Martínez R, Villar A (2012) A newer Human Development Index. *Journal of Human Development and Capabilities* 13:247-268.

Herrero C, Martínez R, Villar, A (2010). Multidimensional social evaluation: An application to the measurement of human development. *Review of Income and Wealth* 56:483-497.

Herrero C, Martínez R, Villar, A (2019). Population structure and the Human Development Index. *Social Indicators Research* 141:731-763.

Hicks DA (1997). The inequality-adjusted human development index: a constructive proposal. *World Development* 25:1283-1298.

Hirai T (2017). *The Human Development Index and its evolution. The creation of the human development approach*. Palgrave Macmillan.

Hopkins M (1991). Human development revisited: a new UNDP report. *World Development* 19:1469-1473.

Hwang CL, Yoon K (1981). *Multiple attribute decision making*, Springer-Verlag, Berlin.

Joro T, Korhonen P (2015) Extension of data envelopment analysis with preference information: Value efficiency. International Series in Operations Research and Management Science 218, Springer.

Karagiannis G (2017) On aggregate composite indicators. *Journal of Operational Research Society* 68:741-746.

Karagiannis G (2020) DEA models without inputs or outputs: A tour de force, in Parmeter C, Sickles, RC (eds.), *Methodological contributions to the advancement of productivity and efficiency analysis*, Springer Proceedings in Business and Economics, Springer, NY.

Karagiannis R, Karagiannis G (2020). Constructing composite indicators with Shannon entropy: The case of Human Development Index. *Socio-Economic Planning Sciences* (in press) DOI: 10.1016/j.seps.2019.03.007.

Kelley AC (1991). The human development index: 'handle with care'. *Population and Development Review* 17:315-324.

Klugman J, Rodríguez F, Choi H-J (2011). The HDI 2010: new controversies, old critiques. *Journal of Economic Inequality* 9:249–288.

Korhonen P, Soismaa M, Siljamaki A (2002). On the use of value efficiency analysis and some further developments. *Journal of Productivity Analysis* 17:49-65.

Krishnakumar J (2018). Trade-Offs in a multidimensional Human Development Index. *Social Indicators Research* 138:991-1022.

Li Q, Maasoumi E, Racine JS (2009). A nonparametric test for equality of distributions with mixed categorical and continuous data, *Journal of Econometrics* 148:186-200.

Lind N (2010). A calibrated index of human development. *Social Indicators Research* 98:301-319.

Lind N (2019). A development of the Human Development Index. *Social Indicators Research* (in press) DOI: 10.1007/s11205-019-02133-9

Lozano S, Gutiérrez E (2008). Data envelopment analysis of the human development index *International Journal of Society Systems Science* 1:132-150.

Luque M, Perez-Moreno S, Rodriguez B (2016). Measuring human development: A multi-criteria approach. *Social Indicators Research* 125:713-733.

Mahlberg B, Obersteiner M. (2001). Remeasuring the HDI by data envelopment analysis. IIASA interim report IR-01-069, Luxemburg: 2001.

Marshall E, Shortle J (2005). Using DEA and VEA to evaluate quality of life in the Mid-Atlantic States. *Agricultural and Resource Economics Review* 34:185-203.

Mazumdar K (2003). A new approach to human development index. *Review of Social Economy* 61:535-549.

Mishra S, Nathan HSK (2018). A MANUSH or HUMANS characterization of the Human Development Index. *Journal of Human Development and Capabilities* 19:398-415.

Neumayer E (2001). The human development index and sustainability-a constructive proposal. *Ecological Economics* 39:101-114.

Nguefack-Tsangue G, Klasen S, Zucchini, W (2011). On weighting the components of the human development index: a statistical justification. *Journal of Human Development and Capabilities* 12:183-202.

Noorbakhsh F (1998a). A modified Human Development Index. *World Development* 26:517-528.

Noorbakhsh F (1998b). The human development index: some technical issues and alternative indices. *Journal of International Development* 10:589-605.

OECD, *Handbook on constructing composite indicators: Methodology and user guide*, Paris, 2008.

Ogwang T (1994). The choice of principal variables for computing the human development index. *World Development* 22:2011-2014.

Ogwang T, Abdou A (2003). The choice of principal variables for computing some measures of human well-being. *Social Indicators Research*, 64:139-152.

Palazzi P, Lauri A (1998). The human development index: suggested corrections. *BNL Quarterly Review* 51:193-221.

Panigrahi R, Sivramkrishna S (2002). An adjusted Human Development Index: robust country rankings with respect to the choice of fixed maximum and minimum indicator values. *Journal of Human Development and Capabilities* 3:301-311.

Pinar M, Stengos T, Topaloglou N (2013). Measuring human development: a stochastic dominance approach. *Journal of Economic Growth* 18:69-108.

Pinar M, Stengos T, Topaloglou N (2017). Testing for the implicit weights of the dimensions of the Human Development Index using stochastic dominance. *Economics Letters* 161:38-42.

Prados de la Escosura L (2010). Improving human development: A long-run view. *Journal of Economic Surveys* 24:841-894.

Ranis G, Stewart F, Samman E (2006). Human development: beyond the human development index. *Journal of Human Development and Capabilities* 7:323-358.

Ravallion M (2012). Troubling tradeoffs in the Human Development Index. *Journal of Development Economics* 99, 201-209.

Rogge N, Self R (2019) Measuring regional social inclusion performances in the EU: Looking for unity in diversity. *Journal of European Social Policy* 29:325-344.

Saaty TL (1980). The analytic hierarchy process. New York: McGraw-Hill.

Sagar AD, Najam A (1998). The human development index: a critical review. *Ecological Economics* 25:249-264.

Saisana M, Saltelli A, Tarantola S (2005). Uncertainty and sensitivity analysis techniques as tools for quality assessment of composite indicators. *Journal of the Royal Statistical Society Series A* 168:307-323.

Sayed H, Hamed R, Ramadan MA-G, Hosny S (2015). Using meta-goal programming for a new human development indicator with distinguishable country ranks. *Social Indicators Research* 123:1–27.

Sen A (1999). *Development as Freedom*. Oxford University Press, Oxford.

Srinivasan TN (1994). Human development: A new paradigm or reinvention of the wheel? *American Economic Review* 84:238-243.

Stapleton LM, Garrod GD (2007). Keeping things simple: why the Human Development Index should not diverge from its equal weights assumption. *Social Indicators Research* 84:179-188.

Tofallis C (2013). An automatic-democratic approach to weight setting for the new human development index. *Journal of Population Economics* 26:1325-1345.

Tofallis C (2014). On constructing a composite indicator with multiplicative aggregation and the avoidance of zero weights in DEA. *Journal of the Operational Research Society* 65:791–792.

UNDP (1994). *Human Development Report 1994*, New York: Oxford University Press.

UNDP (2016). *Human Development Report 2016: Human Development for Everyone*, New York: Palgrave Macmillan.

Van Puyenbroeck T, Rogge N (2017). Geometric mean quantity index numbers with Benefit-of-the-Doubt weights. *European Journal of Operational Research* 256:1004–1014.

Zhou P, Ang BW, Zhou DQ (2010). Weighting and aggregation in composite indicator construction: a multiplicative optimization approach. *Social Indicators Research* 96:169–181.

Footnotes

---

[1] For example, consider two countries A and B being evaluated on the basis of two indicators, namely I_1 (patents) and I_2 (research grants, in thousand $). If country A outperforms B in terms of patents but is outperformed by B in terms of research grants, the BoD model will base the composite indicator of country A only on the patents indicator and assign a zero weight on the research grants indicator, while the reverse will occur for country B. Comparing the performance of the two countries using these composite indices, would be deemed inappropriate.

[2] The benefit-of-the-doubt weighting might also be criticized for dismissing one of the three basic requirements in social choice theory in response to Arrow's theorem, namely anonymity or the assignment of equal weights to all indicators. Nevertheless, OECD (2008, p. 105) argue that anonymity is not an essential requirement in the construction of a composite indicator, as equal weighting is usually only one of the possible weighting schemes.

[3] For a comprehensive treatment of VEA see Joro and Korhonen (2015).

[4] We should emphasize that the paper's aim is to provide an *alternative* approach to that of weight restrictions in incorporating DM preferences to the conventional BoD model, rather than an approach that performs *better* in restricting the flexibility of weights in conventional BoD, compared to weight restrictions.

[5] The BoD is one of the four approaches proposed by OECD (2008) for constructing composite indicators. However, CI construction is a constantly expanding research field, in which several new methodological advancements exist. Some of these contributions are related to the BoD model, others make use of multicriteria decision-making approaches, such as goal-programming and non-compensatory approaches, while there are also mixed or hybrid approaches combining different methodologies to construct a composite indicator. A review of these approaches is a task out of the scope of the present paper, and the interested reader is referred to Greco *et al.* (2019) and El Gibari *et al.* (2019) for recent reviews.

[6] We emphasize that such subjectivity is also inherent in several stages of the composite indicator construction process, such as the selection of the relevant indicators to be included in the composite and the normalization scheme. It is frequently present in the selection of weight bounds in weight-restricted BoD as well.

Thus, malevolent DMs can also select weight bounds that will curb the BoD efficiency frontier, resulting in an evaluation process that favors certain DMUs.

[7] Some MPS selection alternatives might prove to be as time-consuming as the process of selecting weight restriction bounds. Nevertheless, as Korhonen *et al.* (2002) ague, DMs are more keen on simply pointing at a DMU rather that engaging in the task of selecting weight bounds, meaning that the concept of the MPS is generally easier to understand and to select, compared to absolute or relative weight bounds.

[8] The use of AHP for MPS selection was proposed in Korhonen *et al.* (2002).

[9] We thank an anonymous referee for suggesting this alternative, which is inspired from the multicriteria TOPSIS (Technique for Ordered Similarity to Ideal Solution, see Huang and Yoon, 1981) technique. TOPSIS also involves an Anti-Ideal DMU, namely one utilizing the minimum observed indicator values across DMUs, but such a benchmark choice is not suggested as an MPS as it would be more likely to represent the *least* rather that the *most* preferred solution.

[10] For a recent review of the underpinnings and development of the HDI see Hirai (2017).

[11] There is a long discussion in the literature about the logarithmic transformation of the income variable; see Kelley (1991), Chakravarty (2011), Ravallion (2012), and Herrero *et al.* (2012).

[12] Prior to 1994, the goalposts were set by the sample minimum and maximum values.

[13] For comparative results regarding the first three of these normalizations for the HDI see Karagiannis and Karagiannis (2020).

[14] Sagar and Najam (1998), Prados de la Escosura (2010), Herrero *et al.* (2010), and *Zhou et al.* (2010) have also used geometric aggregation while Noorbakhsh (1998a; b) used the L_2 distance of each country from an ideal country that has the sample maximum value of indicators, Luque *et al.* (2016) and Krishnakumar (2018) set the HDI equal to the minimum of the three indicators (a scheme that allows for no substitutability), and Noorbakhsh (1998a) used the Borda's aggregation rule.

[15] Tofallis (2013) used a multiplicative BoD model that satisfies both unit and scale invariance but which, according to van Puyenbroeck and Rogge (2017), can be no longer considered as a geometric weighted average of indicators, as it violates the linear homogeneity property.

[16] The procedure is repeated in Lind (2019) using world data for the years 1990-2017. The findings differentiate from the 2010 study in that the weight of income is now the lowest.

[17] In this variant of the BoD model, $\sum_{j=1}^{J} u_j^k = 1$ in addition to other restrictions in (1).

[18] The artificial country with all indicators set at 0.5 is a multiple of the "Ideal DMU" country, for which all indicator values are equal to one. Hence, the radial projection of the Ideal DMU on the efficient frontier and, consequently, its DEA-efficient peers, coincide with those of the artificial country (i.e. Norway and Australia). Thus, the use of an "Ideal DMU" country as the MPS will produce the same results with our second and third proposed alternatives.

[19] Notice that, as Karagiannis (2017) has shown, the average accurately reflects the aggregate in the case of the BoD and thus, the numbers in the following Tables and Figures can be seen as aggregate values.

[20] In terms of Figure 1, we may think of Hong-Kong as being DMU A, for which the preferred range of mixes (between the I_2 axis and OB) is very wide and unbalanced. On the other hand, we may think of Singapore and Australia as being DMUs B and D respectively, the preferred mix ranges of which are slightly less wide but relatively more balanced compared to that of DMU A.

[21] In terms of Figure 1, we may think of Norway as being DMU C that displays the most balanced performance but has a relatively narrow preferred mix range, potentially serving as a peer for a few number of inefficient DMUs and of Qatar as being DMU F whose mix favors extremely indicator 2.

[22] With reasonable pairs and triads, we mean pairs or triads of countries that share at least one common facet of the BoD-efficient frontier.

[23] The former choice is also supported by empirical findings indicating that using the average of the two variables results in substantial information loss (Canning *et al.*, 2013).