European Journal of Operational Research 256 (2017) 1004–1014

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

Geometric mean quantity index numbers with Benefit-of-the-Doubt weights

Tom Van Puyenbroeck, Nicky Rogge*

KU Leuven, Faculty of Economics and Business, Warmoesberg 26, 1000 Brussels, Belgium

ARTICLE INFO

Article history:
Received 5 November 2015
Accepted 18 July 2016
Available online 22 July 2016

Keywords:
Geometric mean index numbers
Multiplicative aggregation
Benefit-of-the-Doubt model
Composite indicator
Multi-factor decomposition

ABSTRACT

Geometric mean index numbers are a multiplicative aggregation of (price or quantity) ratios with their importance exponents/weights derived from one or more observed budget shares. In the specific context of composite indicator construction, we propose to use the budget shares as naturally generated by the linear Benefit-of-the-Doubt model. This approach is directly inspired by the literature on index number theory. Our basic model is easily extended to provide transitive composite indicator orderings in a multilateral setting. Also, a multi-factor decomposition is proposed to explain the intertemporal evolution of a single entity. We illustrate our results with social inclusion data for the EU-countries.

1. Introduction

Composite Indicators (CIs hereafter), such as the UNDP's well-known Human Development Index, have attracted increased attention over the last years, as can be witnessed by the growing literature addressing both methodological aspects and their application to many different policy-relevant themes. Essentially, CIs boil down to aggregates of sub-indicators that typically contain numerical information on theme-relevant dimensions. In line with many applications we will use the convention in this paper that such CIs are constructed to summarize the aggregate performance of countries $i = 1,\ldots, N$. Recurring and interrelated concerns pertaining to the construction of such aggregates are (i) the selection of specific sub-indicators, (ii) the fact that these sub-indicators may have different measurement units in their original form, (iii) the importance one should attach to a sub-indicator in terms of its contribution to the final composite construct, (iv) the functional form of the aggregator function, (v) robustness of the obtained results to some of these modelling options, etc. (for a more elaborate discussion, see Nardo et al., 2005). In this paper we are primarily concerned with (ii), (iii), and (iv), i.e. with commensurability, weighting and aggregation. Specifically, we contribute to the CI-literature by presenting a method that combines a multiplicative aggregator with the idea of so-called "benefit of the doubt"-weighting, whilst taking care of possibly different original measurement units. In addition, and importantly, we aim for a method that constructs CIs with some desirable mathematical properties that help to render their numerical values as well as changes of these values intuitive.

The Benefit-of-the-Doubt (BoD) weighting technique has become an established method in the CI-literature. As is well-known, the BoD-technique is inspired by (the multiplier formulation of) Data Envelopment Analysis (DEA) (Charnes, Cooper, & Rhodes, 1978), an efficiency measurement technique popular in the Management Science and Operation Research literature. Succinctly, and in its basic formulation, this is a data-oriented method for deriving weights, such that an observation's CI value is maximized subject to an upper bound (usually 100 percent) that is attained either by the weighted sum of that country's sub-indicator values itself or by the similarly weighted sum of at least one other observation's sub-indicator values. In the latter scenario, a strong case (and a rather intuitive degree interpretation) can be provided for the idea that the evaluated observation itself is outperformed, since there exists at least one observation in the sample that achieves a higher CI value even when the most favorable weights of the evaluated observation are applied to that other observation's sub-indicator values. In addition, the outperforming (set of) observation(s) can be regarded as best-practice benchmark(s) (for an extensive introduction to the BoD-method, see e.g. Cherchye, Moesen, Rogge, & Van Puyenbroeck, 2007). The increasing popularity of the BoD-approach shows by a rapidly increasing number of applications in various policy contexts (e.g., human development Blancard & Hoarau, 2013; Mariano, Sobreiro, & do Nascimento Rebelatto, 2015),

* Corresponding author.
E-mail addresses: Tom.vanpuyenbroeck@kuleuven.be (T. Van Puyenbroeck), nicky.rogge@kuleuven.be (N. Rogge).

http://dx.doi.org/10.1016/j.ejor.2016.07.038
0377-2217/

environmental performance (Rogge, 2012; Zanella, Camanho, & Dias, 2012), measuring active ageing (e.g., Amado, São José, & Santos, 2016), etc., as well as by different extensions and variants of the basic model in the literature.

Whereas the traditional BoD-weighting model constructs CIs as weighted arithmetic averages, some variants have been proposed that combine both multiplicative aggregation and BoD-weighting (e.g., Blancas, Contreras, & Ramírez-Hurtado, 2012; Giambona & Vassallo, 2014; Tofallis, 2014; Zhou, Ang, & Zhou, 2010). A key feature of these variants is that they directly combine both multiplicative aggregation and BoD-weighting in one computation step (i.e., multiplicative BoD-models). Relative to our proposal below, these variants could indeed be labeled "direct" approaches given that they are based on programming problems looking for optimal weights such that $CI_i = \prod_{r=1}^{s} y_{ri}^{w_r}$ (Blancas et al., 2012; Giambona & Vassallo, 2014; Zhou et al., 2010) or $CI_i = w_{0i} \prod_{r=1}^{s} y_{ri}^{w_r}$ (Tofallis, 2014) are maximized, subject to appropriate constraints (for more see Section 2). Building on index number theory, our alternative approach suggests a two-step procedure in which information derived from a linear BoD-model in a first step is used to impute shadow price based 'budget share' expressions in the construction of the CI as a geometric quantity index (which itself is not maximized) in a second step. We argue and illustrate that our alternative method, while indirect in terms of its construction, has some comparative advantages, notably when taking the perspective that a CI ideally ought to embed some intuitive (axiomatic) characteristics.

The alternative that we propose in the present paper (Section 3) starts from a trivial, but fundamental observation that actually provides considerable guidance in forwarding such desirable characteristics, viz., that CIs are ultimately constructed for comparing the (aggregate) *relative* performance of different entities among each other and/or over time. We accordingly use 'quantity relatives' (i.e. ratios of sub-indicator values) as the primitive arguments in our aggregator, and thus express each sub-indicator relative to some base performance standard. In turn, this choice limits the family of sensible aggregators to what Aczél (1990) called 'relative merging functions', which are in fact multiplicative functions (see also Fleming & Wallace, 1986). As noted by Aczél, aggregating such relatives is in fact also done with some (price and) quantity indices, and it are the latter –including some of their attractive properties – that provide direct inspiration to the CI-formulation we propose in this paper. This group comprises geometric versions of the Laspeyres and Paasche indices, the Törnqvist indices, the Vartia-Sato index, etc. In (price or) quantity index settings, the basic difference between members of this family relates to the choice of the weights attached to the underlying relatives, but in many cases these quantity-specific weights have a direct connection with the importance of the quantity in total outlays, as measured by the budget share. In our context, we derive such budget shares from the original (linear) BoD-model.

Section 4 provides a first extension of this basic idea. There, we look at the well-known fact that, since optimal weights are observation-specific in a basic BoD-model, there is an intrinsic difficulty with ordering the resulting aggregate (CI) values. We propose three different ways of coming up with a shared weighting structure, and thus three variants of a multilateral multiplicative index. One of these is directly connected with the Multilateral Generalized Törnqvist index, another allows for a richer underlying information basis by building on Doyle and Green (1994) cross-efficiency notion.

Section 5 takes an intertemporal perspective. There we offer a multi-factor decomposition of changes in our geometric CI, disentangling to what extent the observed aggregate performance change is the result of changes in a country's sub-indicator values, base performance standards, and set of optimal weights.

Section 6 concludes and suggests some other possible extensions.

Throughout, we illustrate our findings with European data on social inclusion performance, figuring on the European Commission's dashboard of key social indicators, that are similar to those studied by Giambona and Vassallo (2014). More precisely, (the inverse of)[1] four social inclusion and poverty indicators are used to construct country CIs: (1) y_1: At risk of poverty rate (cut-off point: 60 percent of median equivalised income after social transfers), (2) y_2: People aged 0–59 living in jobless households: share of persons aged 0–59 who are living in households where no-one is working, (3) y_3: Early leavers from education and training, and (4) y_4: Material deprivation rate – Economic strain and durables dimension (see Appendix A for the social inclusion and poverty data for the European Member States in 2006 and 2010). Luxembourg was excluded from the analysis as it is typically considered an outlier (as discussed by Whelan & Maître, 2010, the small population in combination with the distinct nature of the data observations makes that the presence of Luxembourg in the analysis could distort the results). As discussed in detail by Cherchye, Moesen, and Van Puyenbroeck (2004), the subsidiarity principle plays a key role in European social inclusion policy, which provides a *prima facie* reason for using BoD-weights when creating a CI in these circumstances.

2. Literature

In the CI-literature, multiplicative aggregation, and in particular the use of weighted geometric averages of (normalized) indicators, has been advocated as superior to the more widespread 'weighted arithmetic average' CI-specification for several reasons. First, Ebert and Welsch (2004) have demonstrated that, if sub-indicators y_r $(r = 1,\ldots, s)$ are measured on a ratio-scale and are strictly positive, a weighted geometric average $\prod_{r=1}^{s} y_r^{w_r}, w_r \geq 0, \sum w_r = 1$ is meaningful in the sense that, when it is applied to different countries, their resulting ordering is independent of the exact scaling of each of the sub-indicators. That is, if we have any two countries i resp. j with associated sub-indicators y_{ri} resp. y_{rj} $(r = 1,\ldots, s)$ such that $\prod_{r=1}^{s} y_{ri}^{w_r} \geq \prod_{r=1}^{s} y_{rj}^{w_r}$ and one (or more) sub-indicator(s) is rescaled, say to $\rho y_{\ell i}$ resp. $\rho y_{\ell j} (\rho > 0)$, we obtain accordingly rescaled averages $\rho^{w_\ell} \prod_{r=1}^{s} y_{ri}^{w_r} \geq \rho^{w_\ell} \prod_{r=1}^{s} y_{rj}^{w_r}$. This is an obvious (and in fact quite compelling) advantage, as otherwise it could be that the ordering of two countries in terms of a composite comparison could switch if one of the constituent sub-indicators is measured in, say, dollars instead of euros, in tonnes instead of kilograms, etc. Such invariance results are far more limited in the case of a linear weighted average, implying that there is a greater risk with the latter aggregator that the final ordering is dependent on the original measurement units or the specific normalization method used. Second, multiplicative aggregation provides one way to deal with the concern that using a linear weighted average as the functional form of a CI implies perfect substitutability/constant trade-offs between the different sub-indicators. Or, relatedly, it ensures that the marginal returns to an increase in a sub-indicator value are diminishing rather than constant. Finally, a weighted geometric average

[1] In their original definition, all these sub-indicators are such that a higher value indicates a lower level of social inclusion. We therefore use their inverse values in our application. Rather than tailoring all our formulas below to our specific application by switching the role of numerators and denominators –which is the straightforward result of using inverse values–, we opt to present them for the conventional case in which (original or transformed) sub-indicators are 'goods' rather than 'bads'. Or, to put this otherwise, while other ways of treating undesirable indicators have been proposed in the literature (see, e.g., Zanella, Camanho & Dias, 2015 for a comparative assessment), we opt here to take inverses since this fits in naturally with our use of quantity relatives (hence, ratios of a country's values and corresponding benchmark values) as building blocks of our CI.

penalizes inequality among (normalized) sub-indicators for a country, i.e. results in a lower value for the CI, relative to when a linear aggregator would be used for the same data. The two only coincide when $y_{ri} = y_{r'i}, \forall r, r' \in \{1, \ldots, s\}$ (cf. the arithmetic-geometric mean inequality). The 2010 change in the construction of UNDP's Human Development Index is a well-known illustration of recognizing and dealing with these issues (UNDP, 2010; but see Ravallion, 2012 for a more nuanced view).

As mentioned in the introductory section, in the CI-literature several variants have recently been proposed which combine both multiplicative aggregation and BoD-weighting (e.g., Blancas et al., 2012; Giambona & Vassallo, 2014; Tofallis, 2014; Zhou et al., 2010). A key feature of all these variants on the original linear BoD-model is that they propose a direct combination of multiplicative aggregators with BoD-weights (i.e., multiplicative BoD-models). That is, these models directly solve the following *multiplicative* problem (see, e.g., Giambona & Vassallo, 2014; Zhou et al., 2010):

$$CI_i = \max_{w_{1i},\ldots,w_{si}} \prod_{r=1}^{s} y_{ri}^{w_{ri}}$$

s.t.

$$\prod_{r=1}^{s} y_{rj}^{w_{ri}} \le e \quad (N \text{ constraints, one for each country } j = 1, \ldots N)$$

$$w_{ri} \ge 0 \quad (s \text{ constraints, one for each weight } r = 1, \ldots, s)$$

where e is Napier's constant. This model is also solved N times, i.e. for each country separately. This and similar models (e.g. the common weight model of Blancas et al., 2012) are in practice converted to an equivalent linear programming model by applying a logarithmic transformation.

As nicely clarified by Tofallis (2014), one should however be wary that the multiplicative BoD-model as above, contrary to what is the case in the linear BoD-model, needs the addition of an overall (country-specific) scaling factor to the objective function in order to ensure unit invariance. Imposing scale invariance furthermore requires that each of the exponents (weights) is at least unity (see Charnes, Cooper, Seiford, & Stutz, 1983 for similar remarks pertaining to unit invariant multiplicative DEA models). Thus, the scale invariant counterpart of the multiplicative BoD-model is denoted by Tofallis (2014) as follows:

$$CI_i = \max_{w_{0i},w_{1i},\ldots,w_{si}} w_{0i} \prod_{r=1}^{s} y_{ri}^{w_{ri}}$$

s.t.

$$\prod_{r=1}^{s} y_{rj}^{w_{ri}} \le 1 \quad (j = 1, \ldots N)$$

$$w_{ri} \ge 1 \quad (r = 1, \ldots, s)$$

$$w_{0i} > 0$$

It is straightforward that the CI associated with the scale invariant counterpart of the multiplicative BoD-model, while taking care of commensurability, is no longer a geometric weighted average of sub-indicator values, due to the presence of the (observation-specific) scaling factor, and given the weight restrictions $w_{ri} \ge 1$. In fact, while the latter have the benefit of automatically avoiding any concern with zero optimal weights, they do imply a violation of an intuitive linear homogeneity property: while one would expect that a 1 percent increase in all sub-indicators would increase the CI with 1 percent, in this case the increase would be at least N percent (ceteris paribus, i.e. abstraction made of the further possibilities that the optimal weights, including the scaling factor, themselves change and/or the CI reaches its upper bound due to this increase).

One alternative way to ensure commensurability in these direct (CI-maximizing) models would be to work with appropriately normalized "pure number" data, which effectively means introducing quantity relatives of the kind that we discuss below in (2), (3), etc. (for more on the use of ratios in DEA/BoD, see Despić, 2013; Despić, Despić, & Paradi, 2007; Emrouznejad & Cabanda, 2010; Hollingsworth & Smith, 2003). Yet this triggers further concerns that are not straightforward to address. For instance, it would require dealing with possible infeasible solutions if one insists on the otherwise intuitive restriction that $\Sigma_r w_r^* = 1$ (which, by construction, is guaranteed in our indirect approach). Also, recalling that these direct models are equivalent to linear programs after taking logs, either one would then need to ensure that all countries' quantity relatives after such a logarithmic transformation are non-negative, or one must deal head-on with the real possibility of negative data (ratios below unity in the multiplicative formulation) when solving such models.[2]

3. A geometric composite index number with BoD value shares

We start by recalling a simple bilateral setting in which an economic quantity index is constructed for $r = 1,\ldots, s$ commodities. We have data on quantities y_{ri}, y_{rj} and their respective prices p_{ri}, p_{rj} for two countries i, j (which could also refer to two time periods for one country, i.e. the intertemporal setting in which price and quantity indices are also often used). A geometric mean quantity index can then be defined as:

$$Q(p_i, y_i, p_j, y_j) = \prod_{r=1}^{s} \left(\frac{y_{ri}}{y_{rj}} \right)^{\alpha_r} \tag{1}$$

where α_r defines how much the r-th commodity contributes to the aggregate index, and are chosen such that $\Sigma\alpha_r = 1$. These weights can be determined without reference to prices: the Jevons index $Q^{JV}(y_j, y_i)$ sets $\alpha_r = 1/s$ for all r, while a Cobb–Douglas index $Q^{CD}(y_j, y_i)$ uses any real constants $0 \le \alpha_r \le 1$. In other cases, these importance weights are quite logically derived from the shares these quantities take in total outlays, i.e. one uses budget shares $\omega_{ri} = (p_{ri}y_{ri}) / \sum_{r=1}^{s} (p_{ri}y_{ri})$ and/or the similarly defined ω_{rj}. A geometric Laspeyres index uses base shares ω_{rj}, a geometric Paasche index builds on ω_{ri}, and, conceptually identical to the well-known Fisher index, the Törnqvist quantity index $Q^T(p_j, y_j, p_i, y_i)$ uses the arithmetic average $(\omega_{ri}+\omega_{rj})/2$. Still other expressions based on budget shares are possible.

Balk (2008, p. 72) lists a set of axiomatic properties of this family: under normal circumstances these indices are increasing (decreasing) in y_{ri} (y_{rj}) (A1); they are linearly homogenous in comparison quantities, i.e. $Q(p_i, \lambda y_i, p_j, y_j) = \lambda Q(p_i, y_i, p_j, y_j)$ (A2); have the identity property $Q(p_i, y_i, p_j, y_i) = 1$ (A3); are homogenous of degree 0 in quantities: $Q(p_i, \lambda y_i, p_j, \lambda y_j) = Q(p_i, y_i, p_j, y_j)$ (A4); and are invariant to changes in units of measurement (multiplying each y_{ri}, y_{rj} with a rescaling factor ρ_r and using its inverse to accordingly rescale p_{ri}, p_{rj} leave Q unaffected) (A5). In turn, these properties imply other derived characteristics such as the mean value property $\min_{r=1}^{s}\{y_{ri}/y_{rB}\} \le Q(.) \le \max_{r=1}^{s}\{y_{ri}/y_{rB}\}$. These features ensure some logical behavior that one would like to see carried over to composite indices. E.g., one would like to have that a CI-value is 100 percent if a countries' sub-indicator scores coincide with a corresponding set of base performance sub-indicator values,

[2] In our proposal, the choice of base performance values is essentially arbitrary and could thus rest, as in our application, on choosing a straightforward natural point of reference such as a sample average. When using a direct multiplicative BoD-model, the concern for excluding negative data a priori narrows down the range of possible base performance observations, e.g. by choosing $y_{rB} = \min_{i=1}^{N}(y_{ri}), \forall r = 1, \ldots, s$.

is 200 percent if each sub-indicator value is exactly twice as much as its base performance counterpart, etc.

With a view towards the typical CI-setting, we must first pick a specific set of base performance indicators y_{rB}. This is largely an arbitrary choice. Here we take it that the intuitive reading of a CI is fostered by using e.g. the sample average of each sub-indicator as a base performance measure. For instance, since in our social inclusion application we have not only national data but also (population-weighted average) sub-indicator values for EU-27 countries as a whole, we will use the latter as base performance values.

Secondly, we need to attach appropriate weights to each of these sub-indicator relatives. As is well-known, the typical CI framework differs from the economic quantity index framework in that market prices (and, hence, budget shares) are lacking (Cherchye, Lovell, Moesen, & Van Puyenbroeck, 2007). Yet precisely in this respect, the endogenously derived weights of a traditional (linear) BoD-model are instrumental. Specifically, for each country i we solve model (P1):

$$\max_{w_{1i},\ldots,w_{si}} \sum_{r=1}^{s} w_{ri} y_{ri}$$

s.t.

$$\sum_{r=1}^{s} w_{ri} y_{rj} \leq 1 \quad (N \text{ constraints, one for each country } j = 1, \ldots N)$$

$$w_{ri} \geq 0 \quad (s \text{ constraints, one for each weight } r = 1, \ldots, s)$$

The objective function of (P1) normally constitutes the basic linear BoD composite indicator. In this case, however, the sum of shadow prices times their associated sub-indicator values provides only an intermediate result. That is, we are interested in a country's BoD *sub-indicator shares* as these complete the analogy (given shadow prices rather than market prices) with using budget shares as importance weights α_r in the geometric quantity index (1). These BoD sub-indicator shares are given by:

$$\omega_{ri}^{*} = \frac{w_{ri}^{*} y_{ri}}{\sum_{r=1}^{s} w_{ri}^{*} y_{ri}}, \tag{2}$$

Where, evidently, we have $\Sigma_r\ \omega_r^* = 1$. In sum, the basic version of the geometric CI we consider is:

$$CI_i^i(y_i, y_B, \omega_i^*) = \prod_{r=1}^{s} \left(\frac{y_{ri}}{y_{rB}}\right)^{\omega_{ri}^*}, \tag{3}$$

with the 'budget shares' de facto provided by the i-th country's BoD sub-indicator shares as derived from (P1) and given in (2).

For a given ω^*, this CI inherits all the aforementioned properties of its generic counterpart (1). Note that this also holds for the units invariance property, as the latter is incorporated by (P1) (see e.g. Theorem 2.2 in Cooper, Seiford, & Tone, 2000), implying that any ω_r^* is unaffected following a rescaling of the r-th sub-indicators.

We further note that (P1) is the least constrained formulation in terms of the values that optimal shadow prices can take, and is in particular such that zero weights can be attributed to some sub-indicators. This may conflict not only with the very idea of creating a true composite indicator, but also with (broadly defined) expert judgment about the importance of specific dimensions. As is explained in more detail in Cherchye et al. (2007), this concern can be overcome using expert information on sub-indicator importance as derived from a 'budget allocation' process (in which experts have to allocate 100 points over the different dimensions). This information can subsequently be used to append (P1) with additional restrictions on the sub-indicator shares (of the form $\omega_r^L \leq \omega_r^* \leq \omega_r^U$) (on weight restrictions, see Allen, Athanassopoulos, Dyson, & Thanassoulis, 1997; Dyson & Thanassoulis, 1988).

Our illustrative application adheres to the latter idea: when computing CI_i^i using 2010 data as shown in Table 1, we imposed that optimal (BoD-estimated) sub-indicator shares should be at least 5 percent (i.e. $0.05 \leq \omega_r^*$). This implies considerable flexibility in the optimal sub-indicator shares, however, not so much as to enable zero weights being attributed to one or more dimensions.[3] Table 1 displays the CI_i^i as based on Eq. (3) together with the optimal sub-indicator shares as in (2). First, as to the optimal sub-indicator shares, BoD-estimations reveal downward binding restrictions for at least one dimension in all EU Member States. Restrictions binding in the downward direction indicate dimensions of relative weakness. For 11 countries, this even implies that three out of the four dimensions get minimal importance, with the remaining dimension granted a (correspondingly maximal) importance coefficient of 0.85, e.g. poverty for Hungary, jobless households for Greece, early leavers from education and training for the Czech Republic, the material deprivation rate for Sweden, etc. Unsurprisingly, these are dimensions for which the pertaining country's original data are showing a good performance relative to other countries. Most countries attribute the highest BoD-derived importance weight to the jobless household dimension, whereas in that sense the at-risk-of-poverty dimension is the second-most important dimension. In the interpretation of the CI_i^i-scores, values higher than one indicate a better aggregate country performance than the EU-27 average in improving social inclusion and reducing poverty. For example, Slovenia's aggregate score of 236.05 percent is obtained as (129 percent)$^{0.05}$(173 percent)$^{0.204}$(280 percent)$^{0.696}$(142 percent)$^{0.05}$. Sweden has by far the best performance in terms of material deprivation rate (y_4) and does good on all the others, which implies that in combination with a tailored importance coefficient set it obtains a very high CI_i^i –value of 5.15. Following this interpretation, the results suggest strong country performances for Sweden, Slovakia, Slovenia, Czech Republic, and Poland. The opposite interpretation holds for CI_i^i-values lower than one; for instance, Romania's score of 91.62 percent is obtained as (78 percent)$^{0.05}$(100 percent)$^{0.85}$(76 percent)$^{0.05}$(27 percent)$^{0.05}$. Countries such as Romania, Bulgaria, and Latvia appear to be among the worst performing countries.

Table 1 also embeds a broad comparison of our results, in terms of ensuing country rankings, with those obtained by Giambona and Vassallo (2014, pp. 279–280), who used a direct (multiplicative) BoD-model (see the ranks between square brackets in Table 1, taken from Giambona & Vassallo, 2014 pp. 279–280) on a similar dataset. Given their fundamentally different construction, it is not unexpected that the results are not fully similar.[4] Still, both methods do agree on specific qualitative findings, e.g., that Slovenia, the Czech Republic, the Netherlands and Sweden are

[3] Ideally, and as stated in the main text, weight bound values should be specified by experts and/or stakeholders. Practical experience teaches us that strong consent, even between experts thoroughly acquainted with the object of study, is unlikely to come about on this matter (on social inclusion within the EU context, see e.g. Cherchye, Moesen, & Van Puyenbroeck, 2004; for an illustration with real data for the Technology Achievement Index see Cherchye et al., 2008). In the current illustrative application we lack such expert information, but still defined the lower weight bound value of 5 percent so as to avoid (quasi-) zero BoD-weights. Stated otherwise, we take it that our social inclusion CI cannot be constructed while disregarding at least one of its constituent sub-indicators, a minimalist position which we take to reflect the underlying idea that all dimensions are considered as providing at least some valuable information to the European Commission's dashboard of key social indicators. As a robustness check we computed the BoD-model as in (P1) with lower weight bound values set equal to 10%. Overall, this implied only minor differences in the majority of resulting budget shares.

[4] Note that this is not a *ceteris paribus* comparison, as Giambona and Vassallo (2014) impose a lower bound value of 20 percent (i.e. 20 percent $\leq \omega_r^*$) leaving less leeway in the definition of the optimal sub-indicator shares (only 20 percent can be assigned freely by the benefit of the doubt principle). In point of fact, as can be seen in our Table 2, narrowing down the range of eligible weights has a detrimental effect on the ranking of specific countries (e.g. Poland, Lithuania).

Table 1
CI_i^i for social inclusion EU Member States (year 2010).

Country	CI_i^i	CI_i^i(Rank)	$\omega_1{}^*$	$\omega_2{}^*$	$\omega_3{}^*$	$\omega_4{}^*$
EU27	**1.0000**	**–**	**0.5514**	**0.3486**	**0.0500**	**0.0500**
Belgium	1.1428	17 [12]	0.7794	0.0500	0.0500	0.1206
Bulgaria	0.7367	26 [26]	0.0500	0.8500	0.0500	0.0500
Czech Republic*	2.6020	2 [1]	0.0500	0.0500	0.8500	0.0500
Denmark	1.4996	11 [8]	0.6680	0.0685	0.0500	0.2135
Germany	1.1395	18 [11]	0.2738	0.5802	0.0960	0.0500
Estonia	1.0277	22 [13]	0.8149	0.0500	0.0500	0.0851
Ireland	1.0970	19 [23]	0.7759	0.0500	0.0500	0.1241
Greece	1.1956	15 [19]	0.0500	0.8500	0.0500	0.0500
Spain	0.9633	23 [21]	0.0500	0.7803	0.0500	0.1197
France	1.2360	13 [9]	0.7670	0.0692	0.0500	0.1139
Italy	1.0608	20 [17]	0.1058	0.7942	0.0500	0.0500
Cyprus	1.7409	7 [10]	0.0500	0.8500	0.0500	0.0500
Latvia	0.8943	25 [25]	0.0500	0.7556	0.1444	0.0500
Lithuania	1.5194	10 [20]	0.0500	0.0500	0.8500	0.0500
Hungary	1.2163	14 [15]	0.8500	0.0500	0.0500	0.0500
Malta	1.1602	16 [22]	0.2014	0.6986	0.0500	0.0500
Netherlands*	2.1249	6 [2]	0.0500	0.5972	0.0500	0.3028
Austria	1.6468	8 [4]	0.0879	0.7964	0.0656	0.0500
Poland	2.1929	5 [14]	0.0500	0.0500	0.8500	0.0500
Portugal	1.3403	12 [18]	0.0500	0.8500	0.0500	0.0500
Romania	0.9162	24 [24]	0.0500	0.8500	0.0500	0.0500
Slovenia*	2.3605	4 [5]	0.0500	0.2039	0.6961	0.0500
Slovakia	2.5404	3 [6]	0.0500	0.0500	0.8500	0.0500
Finland	1.5975	9 [7]	0.0500	0.7905	0.0605	0.0989
Sweden*	5.1463	1 [3]	0.0500	0.0500	0.0500	0.8500
United Kingdom	1.0434	21 [16]	0.7329	0.0500	0.0500	0.1671
			0.2725	**0.4160**	**0.1986**	**0.1128**

Note: EU Member States designates with an asterix are identified by the BoD-model as benchmark performer with a maximum CI-score. By consequence, for these Member States optimal sub-indicator shares may not be unique. See the discussion in the main text.

among the top group performers and Bulgaria, Romania and Latvia among the worst performing countries.

One final and important remark is that we have sidestepped in our analysis the issue of non-unique shadow prices for observations that get the maximum score in the first step of our procedure, i.e. in (P1). Such alternate optima are an inherent feature of (P1) and of standard DEA ("multiplier") models more generally. They pose few problems in a direct BoD-approach, insofar as the main concern there is with the resulting (and in all alternate cases maximal) value of the objective function. Instead, in our case they define the value of the $\omega_r{}^*$, and, hence, influence a country's CI-value. The maximal extent of this phenomenon is directly related to the number of countries that get the maximum value of 1 in (P1). Table 1 shows that this is the case for 4 out of the 26 countries that – unsurprisingly given their benchmark character in (P1) – all show among the highest CI-scores using our two-step methodology. It tends to be mitigated if, as in our application, additional weight (or budget share) restrictions are appended to (P1); *cfr*. the corner solutions for Sweden and the Czech Republic. More generally, the problem of non-unique multipliers/shadow prices can be overcome by additional selection methods (see e.g. Cooper, Ruiz, & Sirvent, 2007). We return to this point below.

4. Multilateral variants for transitive cross-section comparisons

The CI as defined in (3) clearly exploits the BoD-perspective to the fullest extent: each country is equipped with its proper shadow-price based sub-indicator share values. In that sense, (3) is akin to a geometric Paasche index. The resulting indices are bilateral in nature in the sense that they are tailor-made per country to compare the evaluated country i itself with some base performance observation. Just as in the traditional theory of index numbers for international comparisons, such bilateral comparisons based on country-specific weights are of limited value if one's objective is to consider the performance of all observations simultaneously.

One possibility to come up with full rankings of the complete set of observations would hence consist in arbitrarily picking one specific country that defines both the numerator sub-indicator values and the associated BoD-weights in (3), and repeatedly replacing the base performance values y_{rB} in the denominator with the values as observed in different countries. This way of ranking observations may perhaps provide useful information for that specific country itself, but may be considered unsatisfactory from a truly multilateral perspective.

The crucial element to ensure that an index satisfying axioms (A1)–(A5) is also transitive – $CI(y_i, y_j, \omega)\times CI(y_j, y_k, \omega)=CI(y_i, y_k, \omega)$ for all i, j, k – is to ensure that the exponents α_r in Eq. (1) are common to all possible pairwise comparisons (see e.g. Balk, 2008 pp. 97–98; Von der Lippe, 2007, pp. 234–235). An obvious alternative therefore is to include the base performance sub-indicator data in (P1), so that its BoD sub-indicator shares can be used as common ('Laspeyres') exponents for all countries. Thus, using

$$CI_i^B(y_i, y_B, \omega_B^*) = \prod_{r=1}^{s} \left(\frac{y_{ri}}{y_{rB}}\right)^{\omega_{rB}^*} \tag{3'}$$

with $\omega_B{}^*$ similarly obtained as in (2) on the basis of the BoD-weights $w_B{}^*$ associated with the base performance data y_B, is a first simple way to create a transitive index formula, resting on base-performance observations, that can readily be applied for complete multilateral country comparisons and rankings. In fact, in any pairwise index comparison between countries i, j the base performance values obviously cancel out, so that we obtain direct bilateral indices consistent with (3′):

$$\frac{CI_i^B}{CI_j^B} = \frac{CI_i^B(y_i, y_B, \omega_B^*)}{CI_j^B(y_j, y_B, \omega_B^*)} = \prod_{r=1}^{s} \left(\frac{y_{ri}}{y_{rj}}\right)^{\omega_B^*}$$

Note however that an index such as (3′) is still dependent on a *specific* choice of a set of base performance data; it is possible that the ordering of an i, j pair would be different if another set –and thus other ω_B values– would have been taken. In this respect, Von der Lippe (2007, p. 511) refers to the property of weak transitivity. Strong transitivity, alternatively labeled 'base land invariance', requires absence of such dependence. Referring to Balk (2008, pp. 234–236), this can be achieved if either our CI index does not depend on any (shadow) prices (and their related budget shares) –which seems undesirable– or is a function of *all* the shadow prices/budget shares and sub-indicator values of all countries $i = 1,\ldots, N$.

One approach to create such strongly transitive multilateral indices does not proceed by giving a central role to one ("average") base performance (pseudo-)observation, but rather by averaging results over all observations. Specifically, consider the following version of the Multilateral Generalized Törnqvist quantity index (see Eq. (13) in Hill, 1997, or Von der Lippe, 2007, p. 532), but again adapted to our setting by using BoD-weights:

$$CI_i^{MGT}(y_i, y_B, M(\omega^*)) = \prod_{r=1}^{s}\left(\frac{y_{ri}}{y_{rB}}\right)^{M(\omega_r^*)} \tag{4}$$

Or, in its associated bilateral version:

$$\frac{CI_i^{MGT}}{CI_j^{MGT}} = \frac{CI_i^{MGT}(y_i, y_B, M(\omega^*))}{CI_j^{MGT}(y_j, y_B, M(\omega^*))} = \prod_{r=1}^{s}\left(\frac{y_{ri}}{y_{rj}}\right)^{M(\omega_r^*)}$$

where $M(\omega_r{}^*)$ is the arithmetic mean over all countries in the sample of the r-th BoD sub-indicator share as yielded by solving N optimization problems (P1):

$$M(\omega_r^*) = \frac{1}{N}\sum_{i=1}^{N}\omega_{ri}^* = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{w_{ri}^* y_{ri}}{\sum_{r=1}^{s} w_{ri}^* y_{ri}}\right)$$

In our context, we note that the idea of averaging can be extended to the full matrix of country-wise binary comparisons. Such an extension is intimately related to the notion of cross-efficiency, as developed by Sexton, Silkman, and Hogan (1986) and Doyle and Green (1994) (see also Cook & Zhu, 2015). In terms of (P1), this notion goes broader than connecting country-specific data y_{ri} with their own endogenously derived shadow prices $w_{ri}{}^*$, in that it also looks at cross-weighted sums $\sum_{r=1}^{s} w_{rj}^* y_{ri}$ –i.e. using optimal weights of the other observations– and then averaging over these N different sums to provide a final value for the i-th country. We can similarly define N cross-sub indicator shares for each country i, using the BoD-weights associated with countries j=1, …, N:

$$\omega_{r(j)i} = \frac{w_{rj}^* y_{ri}}{\sum_{r=1}^{s} w_{rj}^* y_{ri}},$$

and take the mean of these shares over the N cross-comparisons to provide an average ("peer-weight based") sub-indicator share for the r-th sub-indicator of country i:

$$\bar{\omega}_{ri}^{cross} = \frac{1}{N}\sum_{j=1}^{n}\left(\frac{w_{rj}^* y_{ri}}{\sum_{r=1}^{s} w_{rj}^* y_{ri}}\right) = \frac{1}{N}\sum_{j=1}^{n}\omega_{r(j)i}.$$

Repeating this for each i and averaging over countries, we arrive at the grand mean

$$\Omega_r = \frac{1}{N}\sum_{i=1}^{n}\bar{\omega}_{ri}^{cross}$$

defining a third possible set of common exponents and hence a third multilateral CI:

$$CI_i^{cross}(y_i, y_B, \Omega) = \prod_{r=1}^{s}\left(\frac{y_{ri}}{y_{rB}}\right)^{\Omega_r}, \tag{5}$$

where the 'cross'-superscript refers to the underlying cross-efficiency idea of building on all N x N combinations of BoD-weights and sub-indicator values. It is straightforward to see that (4) can be regarded as a special case of (5), in the sense that CI^{MGT} only uses the diagonal elements of the $N \times N$-matrix used to arrive at $CI_i{}^{cross}$.

The upper part of Table 2 recaptures the BoD-estimated common set of weights $\omega_{rB}{}^*$, and $M(\omega_r{}^*)$ as given in Table 1 (on the first and bottom lines respectively), and additionally shows the weight set Ω_r. It is immediate that the first set, based on the "average" data of one particular (pseudo-)observation, is rather different from the two (base land invariant) other sets, that are created by averaging over the (own and/or cross-) sub-indicator shares of all countries in the sample. One notes for instance that using the first route the BoD sub-indicator share of the at-risk-of-poverty indicator (ω_1) is the highest, to be followed in importance terms by people aged 0–59 living in jobless households (ω_2), whereas not only the magnitude but also the ordering of these two exponent values changes when reverting to the other two approaches. The values of $CI_i{}^B$ as computed using (3′), $CI_i{}^{MGT}$ using formula (4) and $CI_i{}^{cross}$ using (5), are listed in the bottom part of Table 2 in combination with the associated ranks. In general, with pairwise rank correlations all around 97 percent or higher, the difference in rankings between the three scenarios is negligible. Thus the results for our multilateral indices reveal rather similar findings, with countries such as the Czech Republic, the Netherlands, Slovenia and Austria appearing among the best performers in promoting social inclusion and reducing poverty, and Bulgaria, Romania, Spain, Latvia, and Lithuania figuring at the lower end of the rankings. As should be expected, as compared to Table 1 results, the common weight CI-values are lower for those countries of which the own optimal weighting scheme considerably differs from one of these EU-27 wide variants. Sweden and the Czech Republic provide among the most outspoken illustrations of this effect. At the same time, the results for the Czech Republic show that, given the underlying strong relative performance of its separate constituent sub-indicator values, this does not necessarily need to affect its high ranking (a similar remark holds for, e.g., Slovenia). On the other hand, for a country such as Poland, that has basically 'only' an exceptional performance in the early school leaver dimension, the impact of opting for a EU-27 perspective as regards the importance weights ω_3 has a larger impact on its ranking.

Recalling our remark at the end of the previous section on possible multiple optima, we point out that when base performance data y_B are taken from sample averages, as in our case, they do not constitute extreme points in (P1) by construction and, hence, their associated $\omega_{rB}{}^*$ when using (3′) will be unique. Non-uniqueness of individual countries budget shares evidently does carry over to $M(\omega_r{}^*)$ and Ω_r, and consequently to (4) and (5). Notably in this respect, however, additional selection mechanisms among the multiple optima have been proposed already by Doyle and Green (1994), that can be used to eliminate the problem.

To end this section, we note that the proposals discussed above do not exhaust the possibilities of working with (strongly) transitive weights. Since transitivity requires *common* weights, one notable alternative is to derive such common weights already in the first (LP) stage, using the methodology as introduced by Roll and Golany (1993). This method and variants thereof have already been applied in a 'direct' BoD-setting, e.g. by Bernini, Guizzardi, and Angelini (2013) or Hatefi and Torabi (2010) (or in related fields, see, e.g., Cook & Zhu, 2007; Sinuany-Stern & Friedman, 1998), but could also be used in the indirect fashion as proposed in this paper, i.e. by using these first-stage common weights to derive coefficients of a geometric mean quantity index.

5. An intertemporal perspective

We now return to a bilateral setting, at least in so far that in this section we consider the geometric mean quantity index to

Table 2
CI_i^B, CI_i^{MGT} and CI_i^{cross} for social inclusion EU Member States (year 2010) (Note: Rank diff 1 = Rank CI_i^B – Rank CI_i^{MGT}; Rank diff 2 = Rank CI_i^B – Rank CI_i^{cross}).

			ω_1^*		ω_2^*		ω_3^*		
	EU27-weights		0.5514	0.3486	0.0500	0.0500			
	MGT-weights		0.2725	0.4160	0.1986	0.1128			
	cross-weights		0.2617	0.4048	0.2081	0.1254			
Country	CI_i^B	Rank	CI_i^{MGT}	Rank	CI_i^{cross}	Rank	Rank diff 1	Rank diff 2	
EU27	1.0000	17	1.0000	17	1.0000	16	0	+1	
Belgium	1.0356	14	1.0301	14	1.0458	13	0	+1	
Bulgaria	0.7418	27	0.6924	27	0.6896	27	0	0	
Czech Republic	1.6978	1	1.7151	2	1.7602	2	−1	−1	
Denmark	1.2709	9	1.3478	7	1.3657	7	+2	+2	
Germany	1.1152	11	1.1800	10	1.1901	10	+1	+1	
Estonia	0.9651	19	0.9603	19	0.9765	18	0	+1	
Ireland	0.9064	21	0.8891	21	0.9164	20	+1	+1	
Greece	0.9573	20	1.0160	16	1.0079	15	+4	+5	
Spain	0.8363	23	0.8549	23	0.8389	24	0	−1	
France	1.1696	10	1.1515	11	1.1555	12	−1	−2	
Italy	0.9710	18	0.9874	18	0.9746	19	0	−1	
Cyprus	1.2819	7	1.3357	8	1.3030	9	−1	−2	
Latvia	0.8081	24	0.8013	25	0.7998	25	−1	−1	
Lithuania	0.7916	26	0.8264	24	0.8529	23	+2	+3	
Hungary	1.0150	16	0.8850	22	0.8940	22	−6	−6	
Malta	1.0972	12	1.0563	13	1.0242	14	−1	−2	
Netherlands	1.6821	2	1.7629	1	1.7655	1	+1	+1	
Austria	1.4982	3	1.6028	4	1.6126	4	−1	−1	
Poland	1.0192	15	1.1289	12	1.1693	11	+3	+4	
Portugal	1.0518	13	1.0268	15	0.9796	17	−2	−4	
Romania	0.8055	25	0.7690	26	0.7523	26	−1	−1	
Slovenia	1.4935	4	1.6855	3	1.7285	3	+1	+1	
Slovakia	1.2723	8	1.3059	9	1.3561	8	+1	0	
Finland	1.4045	5	1.5276	5	1.5363	6	0	−1	
Sweden	1.3198	6	1.5267	6	1.6034	5	0	+1	
United Kingdom	0.8882	22	0.8954	20	0.9107	21	+2	+1	

assess the dynamic performance of a specific country. Because we shift to an intertemporal framework, our notation will be extended accordingly. In particular, we distinguish between quantities and sub-indicator shares pertaining to period t versus those for period t+1: $y_{ri,t}$ resp. $y_{ri,t+1}$ and $\omega_{r,t}$ resp. $\omega_{r,t+1}$. Evidently, it may well be the case that the base performance standard also changes over time, as is for example the case in our application. We consequently represent base performance data by the quantities $y^B_{r,t}$ and $y^B_{r,t+1}$.

A natural measure of performance change is then based on the ratio of the (country-specific) geometric mean quantity indices for period t and a subsequent period t+1. Formally:

$$PC^i_i = \frac{CI^i_{i,t+1}}{CI^i_{i,t}} = \frac{\prod_{r=1}^{s}\left(\frac{y_{ri,t+1}}{y^B_{i,t+1}}\right)^{\omega^*_{ri,t+1}}}{\prod_{r=1}^{m}\left(\frac{y_{ri,t}}{y^B_{r,t}}\right)^{\omega^*_{ri,t}}} \tag{6}$$

The performance change metric (6) summarizes whether or not a country has advanced between t and t+1, with PC-values larger (smaller) than 1 reflecting improvement (decline) in country performance. As a summary measure, (6) actually combines the joint effect of three types of changes, viz. those in a country's sub-indicator values, those in the base performance values, and those in the BoD-exponents. We therefore decompose country performance change 'PC_i' in corresponding factors. In particular, by rewriting and rearranging (3) as follows

$$PC^i_i = \prod_{r=1}^{s}\left(\frac{\frac{y_{ri,t+1}}{y^B_{r,t+1}}}{\frac{y_{ri,t}}{y^B_{r,t}}}\right)^{\frac{\omega^*_{ri,t+1}+\omega^*_{ri,t}}{2}} \times \prod_{r=1}^{s}\left(\frac{y_{ri,t+1}}{y^B_{r,t+1}}\,\frac{y_{ri,t}}{y^B_{r,t}}\right)^{\frac{\omega^*_{ri,t+1}-\omega^*_{ri,t}}{2}},$$

we obtain that:

$$\begin{aligned} PC^i_i &= \prod_{r=1}^{s}\left(\frac{y_{ri,t+1}}{y_{ri,t}}\right)^{\frac{\omega^*_{ri,t+1}+\omega^*_{ri,t}}{2}} \times \prod_{r=1}^{s}\left(\frac{y^B_{r,t}}{y^B_{r,t+1}}\right)^{\frac{\omega^*_{ri,t+1}+\omega^*_{ri,t}}{2}} \\ &\times \frac{\prod_{r=1}^{s}\left(\sqrt{\frac{y_{ri,t+1}}{y^B_{r,t+1}}\,\frac{y_{ri,t}}{y^B_{r,t}}}\right)^{\omega^*_{ri,t+1}}}{\prod_{r=1}^{s}\left(\sqrt{\frac{y_{ri,t+1}}{y^B_{r,t+1}}\,\frac{y_{ri,t}}{y^B_{r,t}}}\right)^{\omega^*_{ri,t}}} \\ &= \Delta OWN_i \times \Delta BP_i \times \Delta W_i \end{aligned}$$

The first two factors average out the effect of exponent changes and look at the change in sub-indicator values resp. base performance standards. As such the abbreviations 'ΔOWN_i' and 'ΔBP_i' respectively stand for 'own performance change' and 'base performance change'. Ceteris paribus, $\Delta OWN_i > 1$ (< 1) thus indicates aggregate improvement (deterioration) in a country's sub-indicators, whereas $\Delta BP_i < 1$ (> 1) indicates a similar aggregate performance improvement (deterioration) of the base performance values. As for the latter, and recalling property (A1) of the generic geometric quantity index (1), a country's CI indeed should deteriorate (improve), everything else held fixed, if there is an improvement (a deterioration) in the base performance standard. Combining the two factors evidently provides an answer to the question whether or not a country's aggregate sub-indicator values have increased more than those of the according base performance values.

The third factor, ΔW_i, mirrors the other two in that it neutralizes the effect of changes in sub-indicator and base performance values to focus on the impact of changes in the BoD sub-indicator shares. As such, the ΔW_i -component measures the changes in the BoD-estimated sub-indicator shares (hence, the label 'ΔW_i' or weight change). If the value of ΔW_i is larger (resp. smaller) than

Table 3
Intertemporal geometric mean quantity indices for EU Member States based on $CI_i{}^i$ (period 2006–2010).

Country	$CI^i_{i,2006}$	$CI^i_{i,2010}$	$PC_i{}^i$	$\triangle OWN_i{}^i$	$\triangle PB_i{}^i$	$\triangle W_i{}^i$
EU27	1.0000	1.0000	1.0000	0.9876	1.0125	1.0000
Belgium	1.1273	1.1428	1.0138	1.0198	0.9777	1.0167
Bulgaria	0.7585	0.7367	0.9712	1.0381	1.0181	0.9189
Czech Republic	2.6760	2.6020	0.9724	1.0654	0.9127	1.0000
Denmark	1.8741	1.4996	0.8001	0.8652	0.9794	0.9442
Germany	1.3070	1.1395	0.8719	0.9402	1.0032	0.9244
Estonia	1.3976	1.0277	0.7353	0.8052	1.0073	0.9065
Ireland	1.3657	1.0970	0.8033	0.9297	0.9604	0.8997
Greece	1.4897	1.1956	0.8026	0.7641	1.0503	1.0001
Spain	1.6978	0.9633	0.5674	0.6060	1.0131	0.9242
France	1.2646	1.2360	0.9774	0.9803	0.9798	1.0176
Italy	1.2782	1.0608	0.8299	0.8390	1.0295	0.9609
Cyprus	1.9058	1.7409	0.9135	0.8698	1.0503	1.0000
Latvia	1.2527	0.8943	0.7139	0.6904	1.0416	0.9928
Lithuania	1.3369	1.5194	1.1364	0.7935	0.9732	1.4715
Hungary	1.0266	1.2163	1.1847	1.2357	0.9757	0.9827
Malta	1.4573	1.1602	0.7962	0.9109	1.0062	0.8686
Netherlands	2.9202	2.1249	0.7276	1.0707	0.9491	0.7160
Austria	1.6946	1.6468	0.9718	1.0358	1.0225	0.9176
Poland	2.2863	2.1929	0.9591	1.0509	0.9127	1.0000
Portugal	1.6338	1.3403	0.8204	0.7811	1.0503	1.0000
Romania	0.8727	0.9162	1.0499	0.9985	1.0493	1.0021
Slovenia	2.4900	2.3605	0.9480	1.0492	0.9224	0.9796
Slovakia	1.6328	2.5404	1.5559	1.2644	0.9319	1.3206
Finland	1.8913	1.5975	0.8447	1.0506	0.9896	0.8124
Sweden	3.9046	5.1463	1.3180	1.5182	0.8682	1.0000
United Kingdom	1.3397	1.0434	0.7789	0.9734	0.9428	0.8487

one, this suggests that for the evaluated country the weighting system has changed such that policy dimensions in which it performs relatively good as compared to other countries are rewarded more (resp. less) generously than in a previous period.[5]

As our social inclusion and poverty data illustrate, the three change components could well move in different directions. At the general level, our data in Table 3 reveal a performance decline for twenty countries. At the same time, one notes mostly limited changes in PC_i, i.e. values close to 1, and its components for several countries, which may not be that surprising given the rather short time span between our two observation periods. Yet, more outspoken changes can be observed for some countries, which can be either negative (e.g., Spain) or positive (e.g., Slovakia). Moreover, even upon comparing two countries with similar PC_i-values, such as France and Austria, one sees that the underlying reasons for the total change in country-specific performance can be due to different reasons. The case of France indicates a minor decline due to a decline in the sub-indicator relatives ($0.9803 \times 0.9798 \approx 0.96$) that is only partially compensated by a favorable change in the sub-indicator share values. The Austria case demonstrates an improvement of its underlying relative-to-the-EU-27 ratio both in numerator and denominator terms, that is strongly counterbalanced by less favorable $CI_i{}^i$ exponents in the t+1 period.

The figures for Denmark are rather typical for the majority of countries in our dataset: $\triangle OWN_i$-values are below one for sixteen of our observations, thus revealing a deterioration in these countries' own social inclusion data. For some countries (e.g., Spain, Greece, Latvia and Portugal) the decline in own social inclusion data is considerable. For half of the countries the $\triangle PB_i$-components in Table 3 are below one, thus indicating that at the same time, using these countries' BoD-sub-indicator exponents, the aggregate change in base performance is in itself positive, which by (A1) however implies a downward effect on the aggregate quantity relatives. In point of fact, three out of the four sub-indicator EU-27 base performance values improved (slightly) between 2006 and 2010, the exception being the percentage of jobless households (y_2). Unsurprisingly then, those countries with a high own ω_2 in Table 1 are prone to 'benefit' from this increased gap in base performance (denominator) values, as showing up by $\triangle PB_i > 1$ for e.g. Bulgaria, Germany, etc.

Taking both effects together, i.e. looking at the aggregate effect of the changes in quantity relatives ($\triangle OWN_i \times \triangle PB_i$) reveals that for some countries (e.g. Belgium, the Netherlands) their own improvement virtually equals that of the EU-27 base performance standard, therefore implying no country-specific *relative* improvement. (Evidently, we have that $\triangle OWN_i \times \triangle PB_i = 1$ by definition for the EU-27 pseudo-observation). Other countries, e.g., Sweden and Slovakia did fare better in these relative terms, whereas others (e.g. Spain, Latvia, Lithuania) can be associated with a decline in relative-to-EU-27 performance, even using their own country-specific BoD-exponents.

The $\triangle W_i$ -components in Table 3 also reveal different scenarios. First, for the Czech Republic, Cyprus, Poland, Portugal, and Sweden one observes that $\triangle W_i = 1$, indicating that for these countries the own country-specific BoD-exponents remained unaltered. In

[5] Although the above decomposition of our intertemporal version of the geometric mean quantity index may conceptually look somewhat similar to the more familiar decomposition of a standard Malmquist Productivity index (MPI), there are evident and notable differences between both approaches. The former decomposes as ($\triangle OWN_i \times \triangle PB_i$) $\times$ $\triangle W_i$, i.e. own performance change relative to some exogenous benchmark (possibly evolving itself) *times* a weight change effect. In a direct BoD-setting, one would rather compare the resulting BoD "efficiency" scores over different time periods and use an MPI and its decomposition to assess changes (e.g. Cherchye, Lovell et al., 2007). In the BoD-setting, an MPI can radially (only) be decomposed as the product of an efficiency change effect and a technical change/frontier shift effect (see Karagiannis & Lovell, 2016). The efficiency change effect essentially hinges on relative comparisons of own performance with endogenous (and radially defined) benchmarks, rather than exogenous benchmarks. The technical change effect essentially looks at the evolution of such endogenous benchmarks between two or more periods. Computationally, there are also obvious differences: our measure builds on weights as obtained from two BoD-models, whereas the technical change effect in an MPI typically requires additional computation of distances between an observation at time t (resp. t+1) and a benchmark situated on a t+1 (resp. t) frontier.

fact, recalling Table 1, these are 'extreme' countries in that they assign the maximum weight of 85 percent to one of the four underlying indicators, and the minimum weight of 5 percent to the other three indicators. All other countries have different sets of BoD-exponents in both evaluations, suggesting that their relative strengths and weaknesses relative to the other countries changed from 2006 to 2010. To recall, ΔW_i -values below 1 point to an unfavorable change in the country-specific weighting (i.e., policy dimensions in which a country performs relatively good as compared to other countries are assigned relatively lower weights than in a previous period). The most extreme example of such countries is the Netherlands. Looking at this case in more detail, the only shift in its own ω-set pertains to giving increased importance to jobless households ($\omega_{2,2006} = 0.2516$; $\omega_{2,2010} = 0.5972$) at the expense of the material deprivation dimension ($\omega_{4,2006} = 0.6484$; $\omega_{4,2010} = 0.3028$). This shift is understandable once one recognizes that what can be regarded as an outstanding *relative* performance vis-à-vis other countries for material deprivation in 2006 was still very good five years later, but (even despite a minor improvement in that indicator) was in the same relative terms upset by the improvement that other countries had managed to secure with respect to this dimension (as also shows up in the deterioration of the material deprivation Netherlands/EU-27 *ratio*). Or in other words, by 2010 the Netherlands had less to gain, given the performance improvement of the other countries, by putting that much (2010 BoD-)weight on material deprivation. It is essentially this loss in comparative advantage which is revealed by the Netherlands' low ΔW_i -value. At the other extreme, following a doubling of its jobless household rate from 2006 to 2010 and a concomitant decline in relative position among the EU-27 for this dimension, Lithuania's dramatic shift from $\omega_{2,2006} = 0.7817$ to the minimal value of $\omega_{2,2010} = 0.05$ is understandable. The shift was mirrored in ω_3 (going to the maximum of 0.85 from 0.1183) for which Lithuania scored reasonably well in both time periods.

We end this section by noting that, while we have presented the performance change measure and its decompositions on the basis of the basic BoD-measure CI_i^i as presented in Eq. (3), it should be clear that similar change measures and factors can readily be constructed starting from multilateral (common weight) indices such as CI_i^B, CI_i^{MGT} or CI_i^{cross} (Eqs. (3′), (4) and (5) respectively). In these cases, the ΔBP_i-component is the same for all countries by construction, although it can of course take on different values depending on the specific common weight set used. In our application, and recalling from Table 2 the qualitative differences between the 2010 set $\omega_{rB}{}^*$ on the one hand and $M(\omega_r{}^*)$ and Ω_r on the other, such different values can be expected. Appendices B–D display the outcomes of the intertemporal analysis for these common weight indices in more detail.

6. Conclusions and future research

The present paper advocated a two-step procedure ('indirect approach') to combine both multiplicative aggregation and BoD-weighting in the construction of CIs. In a first step, information on the importance weights of the sub-indicators in the CI are derived from a linear BoD-model. In a second step, this weight information is used to impute shadow price based budget share expressions in the construction of the CI as a geometric quantity index. This indirect approach clearly differs from direct approaches that have been proposed in the literature which compute multiplicative BoD-weighted CIs in one computation step. Throughout the paper, we argued and illustrated that our proposal, while indirect in terms of its construction, has some comparative advantages, notably when taking the perspective that a CI ideally ought to embed some intuitive (axiomatic) characteristics.

While our empirical results primarily serve to illustrate the proposed method, they are interesting in their own right. The observation that there is a broadly shared EU-wide concern to combat poverty and social inclusion while, at the same time, the different traditions and instruments to achieve this goal are, under the subsidiarity principle, still largely to be situated at the level of *national* social policy, marks a setting in which performance benchmarking fits uneasily with the idea of some 'imposed' policy priority weighting scheme. Hence the potential value added of BoD-techniques in this setting. Specifically the recurring finding in the CI-rankings computed by the model specifications (4)–(6), that Romania, Bulgaria, Lithuania, and Latvia are among the worst performing member states, may be a valuable trigger for further action by the national states concerned, by national poverty agencies, NGO's, etc. Conversely, we identified EU Member States such as Sweden, the Czech Republic, the Netherlands, Slovenia and Austria as well performing countries, thereby effectively recognizing that different social policy models might overall lead to (comparatively) good results.

As the basic geometric CI of Eq. (3) rests on information yielded by a traditional linear BoD-model, there are many existing variants of the latter that can further be incorporated in our geometric index approach. The cross-efficiency variant CI_i^{cross} illustrated above is one of these, but one could for instance also straightforwardly implement other versions of the BoD-model in the present framework of constructing CIs as geometric mean quantity index numbers. We have already pointed at the potential use of a common set of weight approach (Roll & Golany, 1993) as an alternative, but other examples include variants of the BoD-model that enable deriving robust sub-indicator share exponents (building e.g. on a robust order-m variant of (P1), see e.g. Cazals, Florens, & Simar, 2002), correcting for country background characteristics (i.e., conditional BoD-models, see Verschelde & Rogge, 2012 for an application), handling undesirable policy performance indicators (e.g., Zanella, Camanho, & Dias, 2015), or limiting compensation among performance indicators (e.g., Fusco, 2015) in the construction of CIs. Another variant is to combine the shadow price information of a regular BoD-model with sub-indicator shares ω_i^- as derived from a linear programming model looking for the set of 'worst possible' country-specific weights (Athanassoglou, 2015; Rogge, 2012), and arrive at a compromise CI:

$$CI_i^c(y_i, y_B, \omega_i^*, \omega_i^-) = \prod_{r=1}^{s} \left(\frac{y_{ri}}{y_{rB}}\right)^{\frac{\omega_i^* + \omega_i^-}{2}}, \tag{7}$$

which could then further be decomposed. Combining maximizing (BoD-) and minimizing weights in an overall CI based on multiplicative aggregation is also done by Zhou et al. (2010). However, their approach is clearly different from (7) in that they combine CI-scores from (two) 'direct' multiplicative models to arrive at an overall index. We leave the implementation of these and other variants of the BoD-model in the present framework for future work.

Of course, our alternative proposals are less than perfect in many other ways. We have already indicated that non-unique shadow prices for particular (extremal) observations imply multiple solutions for their CI-values. The problem may be limited in terms of its magnitude, may be further limited by additional weight restrictions, by bootstrapping approaches, or additional selection mechanisms. In any case, given that many different additional selection mechanisms have been proposed (next to Cooper et al., 2007; Doyle & Green, 1994, see e.g. Liang, Wu, Cook, & Zhu, 2008; Wang & Chin, 2010) a further analysis of the issue, while falling beyond the scope of the current paper, seems warranted. In addition, referring to Section 4, we point out that there are still other conceivable multilateral indices that we did not explore here but leave for future research. In fact, the variants that we did discuss are certainly imperfect from an axiomatic point of view. For instance, while (4) is surely more attractive than (3′) if one insists on the (rather compelling) strong transitivity requirement, it is also true that CI_i^{MGT} has its own weaknesses, e.g. because it may violate the mean value property (see e.g. Hill, 1997).

Nevertheless, and to conclude, besides the fact that our indirect use of BoD allows to sidestep a few problems related to direct multiplicative BoD-models, it should be recalled that within the context for which CIs are employed, their informational value *qua* composite virtually never lies in an absolute number (what do we know by the fact that the US' Human Development Index is 0.914?), but in their comparative use. In that sense, one could claim that relative merging functions may be at least as informative –and at least as 'direct'– as weighted averages. The view upheld in this paper is that, once such a viewpoint is taken, axioms such as (A1)–(A5) or others may be further helpful, and can still be combined with a BoD-approach.

Acknowledgments

We thank three anonymous referees for insightful comments and suggestions. This paper is an offshoot of the Impulsproject IMP/14/011 of the KU Leuven (Belgium).

Appendix A. Social inclusion and poverty data for 2006 and 2010

Country	y_1^{2006}	y_2^{2006}	y_3^{2006}	y_4^{2006}	y_1^{2010}	y_2^{2010}	y_3^{2010}	y_4^{2010}
EU27	**16.5**	**19.6**	**15.5**	**9.9**	**16.4**	**21.1**	**14**	**8.4**
Belgium	14.7	26.3	12.6	6.4	14.6	24.7	11.9	5.9
Bulgaria	18.4	27.1	17.3	57.7	20.7	27	13.9	45.7
Czech Republic	9.9	15.3	5.1	9.6	9	14.5	4.9	6.2
Denmark	11.7	12.7	9.1	3.1	13.3	18	11	2.7
Germany	12.5	21.4	13.7	5.1	15.6	18.7	11.9	4.5
Estonia	18.3	13.5	13.4	7	15.8	25.4	11	9
Ireland	18.5	19.3	12.1	4.8	15.2	34.3	11.5	5.7
Greece	20.5	12	15.5	11.5	20.1	16.6	13.7	11.6
Spain	20.3	11.6	30.3	4.1	21.4	22.6	28.2	4.9
France	13.2	19.7	12.4	5	13.3	20.1	12.5	5.8
Italy	19.6	15.2	20.6	6.3	18.2	19.2	18.8	6.9
Cyprus	15.6	9.1	14.9	12.6	15.6	10.9	12.7	11.2
Latvia	23.5	13.8	14.8	31.3	20.9	22.6	12.9	27.6
Lithuania	20	13.7	8.8	25.3	20.5	27.8	7.9	19.9
Hungary	15.9	25.5	12.6	20.9	12.3	29	10.5	21.6
Malta	14.2	17.2	32.2	3.9	15.5	17	23.8	6.5
Netherlands	9.7	13.8	12.6	2.3	10.3	12.6	10	2.2
Austria	12.6	13.6	9.8	3.6	12.1	12.7	8.3	4.3
Poland	19.1	24.3	5.4	27.6	17.6	19	5.4	14.2
Portugal	18.5	10.4	39.1	9.1	17.9	14.2	28.7	9
Romania	24.8	20.6	17.9	36.5	21.1	21	18.4	31
Slovenia	11.6	10.8	5.6	5.1	12.7	12.2	5	5.9
Slovakia	11.6	21.6	6.6	18.2	12	19.4	4.7	11.4
Finland	12.6	14.4	9.7	3.3	13.1	13.9	10.3	2.8
Sweden	12.3	19.6	8.6	2.1	12.9	20.3	6.5	1.3
United Kingdom	19	27.3	11.3	4.5	17.1	29.8	14.9	4.8

Note: we apply the *inverse* figures in our calculations; see also footnote 1.

Appendix B. Intertemporal geometric mean quantity indices for EU Member States based on CI_i^B, 2006–2010

Country	$CI^B_{i,2006}$	$CI^B_{i,2010}$	PC_i^B	ΔOWN_i^B	ΔPB_i^B	ΔW_i^B
EU27	1.0000	1.0000	1.0000	0.9876	1.0125	1.0000
Belgium	0.9593	1.0356	1.0795	1.0356	1.0125	1.0295
Bulgaria	0.7520	0.7418	0.9864	0.9648	1.0125	1.0098
Czech Republic	1.4959	1.6978	1.1350	1.0979	1.0125	1.0210
Denmark	1.5422	1.2709	0.8241	0.8153	1.0125	0.9983
Germany	1.1396	1.1152	0.9786	0.9544	1.0125	1.0126
Estonia	1.1481	0.9651	0.8406	0.8393	1.0125	0.9893
Ireland	1.0019	0.9064	0.9047	0.8772	1.0125	1.0186
Greece	1.1098	0.9573	0.8626	0.8948	1.0125	0.9521
Spain	1.1520	0.8363	0.7260	0.7459	1.0125	0.9612
France	1.1585	1.1696	1.0096	0.9807	1.0125	1.0167
Italy	1.0392	0.9710	0.9344	0.9477	1.0125	0.9737
Cyprus	1.4175	1.2819	0.9043	0.9448	1.0125	0.9453
Latvia	0.9344	0.8081	0.8649	0.8868	1.0125	0.9633
Lithuania	1.0482	0.7916	0.7552	0.7618	1.0125	0.9790
Hungary	0.8835	1.0150	1.1489	1.0917	1.0125	1.0393
Malta	1.1464	1.0972	0.9570	0.9508	1.0125	0.9940
Netherlands	1.6213	1.6821	1.0375	1.0190	1.0125	1.0056
Austria	1.4301	1.4982	1.0476	1.0479	1.0125	0.9873
Poland	0.8521	1.0192	1.1960	1.1866	1.0125	0.9955
Portugal	1.1967	1.0518	0.8789	0.9148	1.0125	0.9489
Romania	0.7528	0.8055	1.0701	1.0849	1.0125	0.9742
Slovenia	1.6601	1.4935	0.8996	0.9090	1.0125	0.9774
Slovakia	1.1438	1.2723	1.1124	1.0673	1.0125	1.0293
Finland	1.4019	1.4045	1.0018	0.9992	1.0125	0.9902
Sweden	1.2763	1.3198	1.0341	1.0000	1.0125	1.0213
United Kingdom	0.8570	0.8882	1.0364	1.0023	1.0125	1.0213

Note: 2006-weights for EU-27: $\omega_1^{B*}=0.4662$; $\omega_2^{B*}=0.4338$; $\omega_3^{B*}=0.05$; $\omega_4^{B*}=0.05$; 2010-weights for EU-27 are shown in Table 2.

Appendix C. Intertemporal geometric mean quantity indices for EU Member States based on CI_i^{MGT}, 2006–2010

Country	$CI^{MGT}_{i,2006}$	$CI^{MGT}_{i,2010}$	PC_i^{MGT}	ΔOWN_i^{MGT}	ΔPB_i^{MGT}	ΔW_i^{MGT}
EU27	1.0000	1.0000	1.0000	1.0151	0.9851	1.0000
Belgium	1.0210	1.0301	1.0177	1.0524	0.9851	0.9817
Bulgaria	0.6081	0.6924	1.1523	1.0524	0.9851	1.1115
Czech Republic	1.5425	1.7151	1.1376	1.1243	0.9851	1.0272
Denmark	1.7619	1.3478	0.7655	0.8243	0.9851	0.9427
Germany	1.1785	1.1800	1.0023	1.0509	0.9851	0.9681
Estonia	1.2533	0.9603	0.7767	0.7977	0.9851	0.9884
Ireland	1.1769	0.8891	0.7699	0.8123	0.9851	0.9621
Greece	1.1425	1.0160	0.8860	0.8983	0.9851	1.0013
Spain	1.2248	0.8549	0.6824	0.7391	0.9851	0.9372
France	1.2349	1.1515	0.9329	0.9670	0.9851	0.9794
Italy	1.1009	0.9874	0.8855	0.9279	0.9851	0.9688
Cyprus	1.3441	1.3357	0.9798	0.9741	0.9851	1.0210
Latvia	0.8829	0.8013	0.9157	0.8765	0.9851	1.0606
Lithuania	1.0543	0.8264	0.8089	0.7834	0.9851	1.0482
Hungary	0.8220	0.8850	1.0972	1.0397	0.9851	1.0713
Malta	1.1146	1.0563	0.9234	0.9695	0.9851	0.9669
Netherlands	1.7497	1.7629	0.9998	1.0789	0.9851	0.9407
Austria	1.6177	1.6028	0.9925	1.0461	0.9851	0.9631
Poland	0.9087	1.1289	1.2841	1.2454	0.9851	1.0467
Portugal	1.0756	1.0268	0.9210	0.9428	0.9851	0.9917
Romania	0.6921	0.7690	1.1050	1.0502	0.9851	1.0680
Slovenia	1.9003	1.6855	0.9048	0.9312	0.9851	0.9863
Slovakia	1.0949	1.3059	1.2359	1.1884	0.9851	1.0557
Finland	1.6078	1.5276	0.9464	1.0178	0.9851	0.9439
Sweden	1.5788	1.5267	0.9878	1.1051	0.9851	0.9074
United Kingdom	1.0387	0.8954	0.8674	0.9271	0.9851	0.9498

Note: 2006-weights: $\omega_1^{MGT*}=0.2042$; $\omega_2^{MGT*}=0.4117$; $\omega_3^{MGT*}=0.1981$; $\omega_4^{MGT*}=0.1806$; 2010-weights are shown in Table 2.

Appendix D. Intertemporal geometric mean quantity indices for EU Member States based on CI_i^{cross}, 2006–2010

Country	$CI_{i,2006}^{cross}$	$CI_{i,2010}^{cross}$	PC_i^{cross}	ΔOWN_i^{cross}	ΔPB_i^{cross}	ΔW_i^{cross}
EU27	1.0000	1.0000	1.0000	1.0249	0.9757	1.0000
Belgium	1.0820	1.0458	0.9665	1.0492	0.9757	0.9441
Bulgaria	0.6824	0.6896	1.0106	1.0564	0.9757	0.9804
Czech Republic	1.7599	1.7602	1.0002	1.1173	0.9757	0.9174
Denmark	1.7040	1.3657	0.8015	0.8324	0.9757	0.9868
Germany	1.2029	1.1901	0.9893	1.0362	0.9757	0.9785
Estonia	1.1703	0.9765	0.8344	0.8602	0.9757	0.9942
Ireland	1.1521	0.9164	0.7955	0.8646	0.9757	0.9429
Greece	1.0451	1.0079	0.9644	0.9308	0.9757	1.0619
Spain	0.9835	0.8389	0.8531	0.7857	0.9757	1.1128
France	1.2507	1.1555	0.9239	0.9705	0.9757	0.9756
Italy	0.9864	0.9746	0.9880	0.9559	0.9757	1.0593
Cyprus	1.2190	1.3030	1.0689	0.9963	0.9757	1.0995
Latvia	0.8608	0.7998	0.9292	0.9232	0.9757	1.0315
Lithuania	1.0930	0.8529	0.7804	0.8320	0.9757	0.9613
Hungary	0.9125	0.8940	0.9797	1.0746	0.9757	0.9344
Malta	0.9746	1.0242	1.0509	0.9922	0.9757	1.0854
Netherlands	1.6574	1.7655	1.0652	1.0823	0.9757	1.0087
Austria	1.5690	1.6126	1.0278	1.0559	0.9757	0.9976
Poland	1.0971	1.1693	1.0658	1.2093	0.9757	0.9033
Portugal	0.8659	0.9796	1.1313	0.9870	0.9757	1.1746
Romania	0.7048	0.7523	1.0674	1.0540	0.9757	1.0379
Slovenia	1.9408	1.7285	0.8906	0.9462	0.9757	0.9646
Slovakia	1.3024	1.3561	1.0412	1.1903	0.9757	0.8965
Finland	1.5684	1.5363	0.9795	1.0064	0.9757	0.9975
Sweden	1.6050	1.6034	0.9990	1.1151	0.9757	0.9181
United Kingdom	1.0753	0.9107	0.8470	0.9234	0.9757	0.9400

Note: 2006-weights: $\omega_1^{cross*}=0.2975$; $\omega_2^{\ cross*}=0.2616$; $\omega_3^{\ cross*}=0.3099$; $\omega^{\ cross*}=0.1310$; 2010-weights are shown in Table 2.

References

Aczél, J. (1990). Determining merged relative scores. *Journal of Mathematical Analysis and Applications, 150*(1), 20–40.
Allen, R., Athanassopoulos, A., Dyson, R. G., & Thanassoulis, E. (1997). Weights restrictions and value judgements in data envelopment analysis: Evolution, development and future directions. *Annals of Operations Research, 73*, 13–34.
Amado, C. A. F., São José, J. M. S., & Santos, S. P. (2016). Measuring active ageing: A data envelopment analysis approach. *European Journal of Operational Research, 255*(1), 207–223.
Athanassoglou, S. (2015). Revisiting worst-case DEA for composite indicators. FEEM Working Paper No. 013.2015.
Balk, B. M. (2008). *Price and quantity index numbers: Models for measuring aggregate change and difference*. Cambridge University Press.
Bernini, C., Guizzardi, A., & Angelini, G. (2013). DEA-like model and common weights approach for the construction of a subjective community well-being indicator. *Social Indicators Research, 114*(2), 405–424.
Blancard, S., & Hoarau, J. F. (2013). A new sustainable human development indicator for small island developing states: A reappraisal from data envelopment analysis. *Economic Modelling, 30*, 623–635.
Blancas, F. J., Contreras, I., & Ramírez-Hurtado, J. M. (2012). Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives. *Journal of the Operational Research Society, 64*(5), 668–678.
Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics, 106*(1), 1–25.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
Charnes, A., Cooper, W. W., Seiford, L., & Stutz, J. (1983). Invariant multiplicative efficiency and piecewise Cobb–Douglas envelopments. *Operations Research Letters, 2*(3), 101–103.
Cherchye, L., Moesen, W., & Van Puyenbroeck, T. (2004). Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the EU. *Journal of Common Market Studies, 42*(5), 919–955.
Cherchye, L., Lovell, C. K., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review, 51*(3), 749–779.
Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research, 82*(1), 111–145.
Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., et al. (2008). Creating composite indicators with DEA and robustness analysis: The case of the technology achievement index. *Journal of the Operational Research Society, 59*(2), 239–251.
Cook, W. D., & Zhu, J. (2015). DEA cross efficiency. *Data envelopment analysis* (pp. 23–43). Springer US.
Cook, W. D., & Zhu, J. (2007). Within-group common weights in DEA: An analysis of power plant efficiency. *European Journal of Operational Research, 178*(1), 207–216.
Cooper, W. W., Ruiz, J. L., & Sirvent, I. (2007). Choosing weights from alternative optimal solutions of dual multiplier models in DEA. *European Journal of Operational Research, 180*, 443–458.
Cooper, W. W., Seiford, L. S., & Tone, K. (2000). *DEA: A comprehensive text with models. Applications, references and DEA-solver software*. London: Kluwer Academic Publishes.
Despić, O. (2013). Some properties of geometric DEA models. *Croatian Operational Research Review, 4*(1), 2–18.
Despić, O., Despić, M., & Paradi, J. C. (2007). DEA-R: Ratio-based comparative efficiency model, its mathematical relation to DEA and its use in applications. *Journal of Productivity Analysis, 28*(1-2), 33–44.
Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the Operational Research Society, 45*(5), 567–578.
Dyson, R. G., & Thanassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of the Operational Research Society, 39*(6), 563–576.
Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management, 47*(2), 270–283.
Emrouznejad, A., & Cabanda, E. (2010). An aggregate measure of financial ratios using a multiplicative DEA model. *International Journal of Financial Services Management, 4*(2), 114–126.
Fleming, P. J., & Wallace, J. J. (1986). How not to lie with statistics: The correct way to summarize benchmark results. *Communications of the ACM, 29*(3), 218–221.
Fusco, E. (2015). Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research, 242*(2), 620–630.
Giambona, F., & Vassallo, E. (2014). Composite indicator of social inclusion for European countries. *Social Indicators Research, 116*(1), 269–293.
Hatefi, S. M., & Torabi, S. A. (2010). A common weight MCDA–DEA approach to construct composite indicators. *Ecological Economics, 70*(1), 114–120.
Hill, R. J. (1997). A taxonomy of multilateral methods for making international comparisons of prices and quantities. *Review of Income and Wealth, 43*(1), 49–69.
Hollingsworth, B., & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters, 10*(11), 733–735.
Karagiannis, G., & Lovell, C. A. K. (2016). Productivity measurement in radial DEA models with a single constant input. *European Journal of Operational Research, 251*, 323–328.
Liang, L., Wu, J., Cook, W. D., & Zhu, J. (2008). Alternative secondary goals in DEA cross-efficiency evaluation. *International Journal of Production Economics, 113*(2), 1025–1030.
Mariano, E. B., Sobreiro, V. A., & do Nascimento Rebelatto, D. A. (2015). Human development and data envelopment analysis: A structured literature review. *Omega, 54*, 33–49.
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on constructing composite indicators: Methodology and users guide. Paris/Ispra*. OECD Publication.
Ravallion, M. (2012). Troubling tradeoffs in the human development index. *Journal of Development Economics, 99*(2), 201–209.
Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The environmental performance index. *Ecological Indicators, 23*, 143–154.
Roll, Y., & Golany, B. (1993). Alternate methods of treating factor weights in DEA. *Omega, 21*(1), 99–109.
Sexton, T. R., Silkman, R. H., & Hogan, A. J. (1986). Data envelopment analysis: Critique and extensions. In R. H. Silkman (Ed.), *Measuring efficiency: An assessment of data envelopment analysis* (pp. 73–105). San Francisco, CA: Jossey-Bass.
Sinuany-Stern, Z., & Friedman, L. (1998). DEA and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research, 111*(3), 470–478.
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*(5), 791–792.
United Nations Development Programme. (2010). *Human development report: The real wealth of nations*. New York: Palgrave Macmillan.
Verschelde, M., & Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional data envelopment analysis approach. *European Journal of Operational Research, 223*(1), 214–225.
Von der Lippe, P. (2007). *Index theory and price statistics*. Frankfurt: Peter Lang Verlag.
Wang, Y. M., & Chin, K. S. (2010). Some alternative models for DEA cross-efficiency evaluation. *International Journal of Production Economics, 128*(1), 332–338.
Whelan, C. T., & Maître, B. (2010). Comparing poverty indicators in an enlarged European Union. *European Sociological Review, 26*(6), 713–730.
Zanella, A., Camanho, A. S., & Dias, T. G. (2012). Benchmarking countries' environmental performance. *Journal of the Operational Research Society, 64*(3), 426–438.
Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research, 245*(2), 517–530.
Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research, 96*(1), 169–181.