CrossMark

Revisiting Worst-Case DEA for Composite Indicators

Stergios Athanassoglou[1]

Abstract Composite indicators are becoming increasingly influential tools of environmental assessment and advocacy. Nonetheless, their use is controversial as they often rely on ad-hoc and theoretically problematic assumptions regarding normalization, aggregation, and weighting. Nonparametric data envelopment analysis (DEA) methods, originating in the production-economics literature, have been proposed as a means of addressing these concerns. These methods dispense with contentious normalization and weighting techniques by focusing on a measure of best-case relative performance. Recently, the standard DEA model for composite indicators was extended to account for worst-case analysis by Zhou et al. (Ecol Econ 62(2):291–297, 2007) [hereafter, ZAP]. In this note we argue that, while valid and interesting in its own right, the measure adopted by ZAP may not capture, in a mathematical as well as practical sense, the notion of worst-case relative performance. By contrast, we focus on the strict worst-case analogue of standard DEA for composite indicators and show how it leads to tractable optimization problems. Finally, we compare the two methodologies using data from ZAP's Sustainable Energy Index case study, demonstrating that they occasionally lead to divergent results.

Keywords Composite indicator · Sustainability index · DEA · Worst-case · Convex optimization

1 Introduction

Composite indicators are becoming increasingly influential tools of public policy. Usually taking the form of a weighted arithmetic average of normalized indicators, these indices condense complex multidimensional information into a single number. As such, they are

✉ Stergios Athanassoglou
   athanassoglou@gmail.com

[1] IEFE, Bocconi University, Milan, Italy

easy to compute and to interpret. Furthermore, they allow for the computation of rankings to assess the comparative standing of different entities (countries, regions, governmental policies under consideration). This conceptual simplicity facilitates communication with the press and public, and thus aids in generating awareness regarding the issue that the composite indicator is meant to address.

Despite their increasing popularity, composite indicators are often strongly criticized by official statisticians and economists, including those interested in the measurement of sustainability (Bohringer and Jochem 2007; Ravallion 2012b). These critiques come in different varieties, both conceptual and methodological. On the conceptual side, it is argued that the underlying issues that composite indicators address are often ill-defined and open to excessive interpretation. Ravallion (2012b) cites Newsweek magazine's "best country rankings" as an intuitive example of this kind of definitional haziness. Statisticians further complain that the process of constructing a composite index discards useful statistical information by reducing multidimensional data to an aggregate measure. While we agree that these are important issues, we do not dwell on them as the focus of this paper is primarily methodological. Here, critics argue that integral modeling assumptions behind the construction of composite indices such as the choice of normalization procedure, aggregation function, and weighting scheme, fail to be grounded in economic theory or a coherent analytic framework (Ravallion 2012b). What is more, the ad-hoc nature of these choices may lead to unintended theoretical consequences such as unacceptable tradeoffs (Ravallion 2012a) and problematic measurement-theoretic implications (Ebert and Welsch 2004). Finally, the indices themselves may be very sensitive to changes in these subjective choices so that any insights or rankings that are generated can be highly non-robust.

One way of addressing the dependence of composite indices on arbitrary assumptions on normalization and weighting (while maintaining the linearity of the aggregation function) is via the nonparametric framework of data envelopment analysis (DEA).[1] First developed by Charnes et al. (1978) in the field of production economics, DEA was primarily conceived as a methodology for measuring the relative efficiency of different decision-making units. Since then, DEA has been the subject of extensive research in both economics and operations research (Cooper et al. 2007). Its application in composite index construction, known as the "benefit of the doubt" (BOD) method, was proposed by Cherchye et al. (2007b). For each entity (country, region, policy) to be assessed, the BOD method searches for its "most favorable" set of weights, defined as the maximizers of the ratio of its score to that of the highest-performing member of the group. Thus, weights are determined endogenously and may differ between entities. Furthermore, it is important to note that DEA takes as input non-normalized data and its scores and rankings are invariant to ratio-scale transformations (i.e., multiplicative changes in units). DEA-like methods are being increasingly used to build composite indices for a variety of applications ranging from market structure and technology (Cherchye et al. 2007a, 2008), to gender issues (Dominguez-Serrano and Blancas 2011), to development and environmental policy and assessment (Fare et al. 2004; Zaim 2004; Zhou et al. 2007, 2010; Zhang et al. 2013; Jin et al. 2014).

---

[1] It is worth noting that there exist other nonparametric frameworks that impose less restrictive assumptions on aggregation. This increased generality usually introduces ambiguity to the index results. For instance, in the context of multidimensional welfare measurement, Anderson et al. (2011) mpose solely monotonicity and quasiconcavity on the aggregation function and derive upper and lower bounds on index scores, not precise values and rankings.

In a recent paper Zhou et al. (2007) [hereafter, ZAP] extended the DEA framework of Cherchye et al. (2007b) to account for worst-case analysis. In particular, they propose a model with which to compute an entity's "least favorable weights" and corresponding worst-case relative performance. The yardstick of performance becomes the ratio of an entity's score to that of the worst member in the group. They then go on to propose a hybrid DEA methodology in which convex combinations of their normalized best- and worst-case DEA scores are considered.

Using ZAP's work as a springboard we argue that, while interesting in its own right, the worst-case measure that they adopt may not capture, in a theoretical as well as practical sense, the notion of worst-case relative performance. We propose an alternative measure that is, in a strict mathematical sense, the worst-case analogue of the BOD model of Cherchye et al. (2007b). While the mathematical structure of this measure differs significantly to that of the BOD method, we show how it can nonetheless be tractably computed, even under general convex restrictions on the weights.[2] We then compare the two methodologies using data from ZAP's Sustainable Energy Index case study, demonstrating that they occasionally lead to notably different results.

1.1 Paper Outline

The structure of the paper is as follows. Section 2 sets up the formal model and relevant DEA framework. It goes on to discuss ZAP's approach to modeling worst-case relative performance and to suggest, by means of a stylized example, how it may result in undesirable conclusions. Section 3 introduces and analyzes an alternative optimization problem that is the strict worst-case analogue of traditional DEA for composite indices. Section 4 applies the proposed procedure to the case study of the original ZAP paper, showing how the two methodologies can lead to divergent results. Section 7 provides conclusions. All mathematical proofs, tables, and figures are collected in the "Appendix".

2 Model Description

Suppose we are given a set $\mathcal{A} = \{a_1, a_2, \ldots, a_A\}$ of A agents and a set $\mathcal{I} = \{i_1, i_2, \ldots, i_I\}$ of I indicators. Moreover let x_{ai} denote agent a's value for indicator i. All indicator values x_{ai} for $a \in \mathcal{A}$ and $i \in \mathcal{I}$ are assumed to be positive. Indicators are weighted with a non-negative column vector of weights $\boldsymbol{w} \in \mathfrak{R}_+^{\mathfrak{I}}$, where w_i denotes the weight assigned to indicator i. Consistent to classical DEA, an agent a's score under weights $\boldsymbol{w}$ is given by the corresponding weighted sum of the non-normalized indicators: $\sum_{i=1}^{I} w_i x_{ai}$.

Now, let us introduce the main concept behind the use of DEA-like methods in composite indicators. Consider an individual agent $a_j \in \mathcal{A}$ and suppose that weights $\boldsymbol{w}$ are chosen. The relative standing of this agent among her peers, given the chosen weights w, is captured via the ratio of her performance to that of the highest-performing agent of group $\mathcal{A}$. Denoting it by a function $f_a(\boldsymbol{w})$, it equals:

---

[2] An additional advantage of the proposed approach is that it results in worst-case DEA scores that share a similar 0–1 scale to that of best-case DEA scores. Thus, there is no need for potentially contentious normalization procedures when taking the aforementioned convex combinations of best- and worst-case DEA scores.

$$f_{a_j}(\boldsymbol{w}) \equiv \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{1}$$

Equation (1) ranges between 0 and 1; the higher it is, the closer agent a_j is to the top performer. If it equals 1, then for this choice of $\boldsymbol{w}$, agent a_j has the top score. The DEA approach to the construction of composite indicators uses exactly this measure of relative standing as its measuring stick of performance. In particular, it searches for the set of weights that maximize the function $f_a(\boldsymbol{w})$, for each agent $a \in \mathcal{A}$. Applied to agent a_j, it solves the following optimization problem:

$$f_{a_j}^* \equiv \max_{\boldsymbol{w} \geq \boldsymbol{0}} f_{a_j}(\boldsymbol{w}) = \max_{\boldsymbol{w} \geq \boldsymbol{0}} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}} \tag{2}$$

Optimization problem (2) determines the weights, subject to a non-negativity constraint, resulting in the best-case relative performance of agent a_j. These are known as the "most favorable weights" for agent a_j.

From a mathematical standpoint, the tractability of problem (2) is crucially dependent on the fact that it may be reduced to the following, equivalent linear-fractional program:

$$f_{a_j}^* = \max_{\boldsymbol{w} \geq \boldsymbol{0}, z} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{z}$$
$$\text{s.t.} \quad \sum_{i=1}^{I} w_i x_{ai} \leq z, \quad a \in \mathcal{A}, \tag{3}$$

which, in turn, can be shown to be equivalent (see section 4.5.2 in Boyd and Vandenberghe (2004) to the linear program

$$f_{a_j}^* = \max_{\boldsymbol{w} \geq \boldsymbol{0}} \quad \sum_{i=1}^{I} w_i x_{a_j i}$$
$$\text{s.t.} \quad \sum_{k=1}^{I} w_i x_{ai} \leq 1, \quad \text{for all } a \in \mathcal{A}. \tag{4}$$

Linear program (4) is the familiar "benefit of the doubt" method for composite indicators discussed in Cherchye et al. (2007b) and applied in many contexts since (Zhou et al. 2007; Cherchye et al. 2007a, 2008; Hatefi and Torabi 2010; Dominguez-Serrano and Blancas 2011; Rogge 2012).

Importantly, additional linear constraints may be imposed to the weights in optimization problem (2) at no conceptual or computational cost. Particularly compelling weight restrictions come in the form of so-called "pie shares" (see Cherchye et al. 2007a, b), which set lower and upper bounds on the contribution of any single indicator to the agent's total score. To wit, given a set of numbers L_i, U_i for all $i \in \mathcal{I}$ the corresponding pie-share constraints to be appended to problem (2), and ultimately also to its linear equivalent (4), are given by

$$L_i \leq \frac{w_i x_{a_j i}}{\sum_{k=1}^{I} w_k x_{a_j k}} \leq U_i, \quad \text{for all } i \in \mathcal{I}. \tag{5}$$

The above constraints hold theoretical as well as practical appeal. Theoretically, their imposition does not compromise the very desirable property of ratio-scale invariance of

DEA, also known as "units invariance" (Cherchye et al. 2007b; Cooper et al. 2007). That is, DEA scores and their resulting rankings remain unchanged under incomparable (i.e., non-identical across indicators) ratio-scale transformations of the original indicators. For example, if an index is composed of three indicators i_1, i_2, i_3 and we multiple indicator i_1 by 2, i_2 by 8 and i_3 by 0.1, this will have no effect on the corresponding DEA scores and rankings. This property is particularly compelling in the case of environmental indices (Ebert and Welsch 2004).[3] Meanwhile, on a practical level pie shares are pure numbers whose meaning is easy to grasp and on whose values experts can usually come to an agreement (Cherchye et al. 2007a, b).

2.1 The Worst-Case Model of ZAP

ZAP take as a starting point the above standard DEA model and extend it to account for worst-case relative performance. Considering again an agent $a_j \in \mathcal{A}$, they draw on previous work by Zhu (2004) and Takamura and Tone (2003) and (implicitly) define this agent's "least favorable weights" as the solution of the following optimization problem:

$$g_{a_j}^{ZAP} \equiv \min_{w \geq 0} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\min_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{6}$$

That is, they define the worst-case DEA weights to be such that they *minimize* the ratio of an agent's performance to that of the *worst-case* performer in the group. Optimization problem (6) retains the nice properties of problem (2) in that it too can be reduced to a linear-fractional program, and ultimately to the following linear program (which, in turn, is the formulation that appears in ZAP's work):

$$g_{a_j}^{ZAP} = \min_{w \geq 0} \quad \sum_{i=1}^{I} w_i x_{a_j i}$$
$$\text{s.t.} \quad \sum_{i=1}^{I} w_i x_{ai} \geq 1, \quad a \in \mathcal{A}. \tag{7}$$

In a formal sense, problem (6), and thus also its linear equivalent (7), does not correspond to worst-case DEA for composite indicators. This is because it abandons the measure of relative performance $f_{a_j}(w)$ of Eq. (1), which constitutes the objective function of problem (2), in favor of an alternative measure, namely

$$g_{a_j}(w) \equiv \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\min_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{8}$$

The ratio $g_{a_j}(w)$ is no smaller than 1 and unbounded above; the smaller it is, the closer agent a_j is to the bottom performer. If it equals 1, then for this choice of w, agent a_j has the worst score. Thus, worst-case DEA as defined by ZAP searches for the set of weights w that minimize the ratio $g_{a_j}(w)$.

While problem (6) is interesting in its own right, and the underlying optimization problem has identical structure to the standard DEA context (and is thus readily solvable using similar techniques), it is not the worst-case analogue of standard DEA. Moreover, it

---

[3] The reader may refer to Example 1 in Ebert and Welsch (2004) for a concrete demonstration of this fact in the context of an index measuring eutrophication.

Table 1 x_{ai} values for Example 1

	i_1	i_2
a_1	9999	1
a_2	5000	5000

may sometimes fail to capture the essence of worst-case relative performance. The following, highly stylized, example illustrates this fact.

Example 1 [counter-intuitive implications of ZAP's model]. Consider the setting described in Table 1 summarizing an instance of the problem for $\mathcal{A} = \{a_1, a_2\}$ and $\mathcal{I} = \{i_1, i_2\}$.

The standard best-case DEA model of Eq. (2) results in identical scores for a_1 and a_2, since $f_{a_1}^* = f_{a_2}^* = 1$. Let us now consider worst-case performance. According to ZAP's model of Eq. (6), agents a_1 and a_2 are equal as they both get the absolute minimum score of 1. This is because there exist weight vectors that equalize their performance (e.g., $w = (1/2, 1/2)'$), thus implying that they are simultaneously the worst performers of the two-member group $\mathcal{A}$. By definition of problem (6), this means that they both get the worst possible score, i.e., $g_{a_1}^{ZAP} = g_{a_2}^{ZAP} = 1$, and so ZAP's methodology cannot discriminate between them. This result does not, arguably, accord with intuition. Indeed, we would expect agent a_2's balanced performance across indicators, in combination with a_1's extremely unbalanced one, to be recognized and rewarded. Furthermore, note that the exact numbers here are not important. Similar results would obtain if we make $x_{a_1 i_1} \geq 0$ as large and $x_{a_1 i_2} \geq 0$ as small as we like, and set $x_{a_1 i_1} + x_{a_1 i_2} = x_{a_2 i_1} + x_{a_2 i_2}$ and $x_{a_2 i_1} = x_{a_2 i_2}$. □

Finally, in order to construct a DEA measure combining best- and worst-case performance, ZAP normalize the results of (2) and (6) via max-min rescaling. This is necessary because the scales of the two measures clearly differ; one ranges from 0 to 1, the other from 1 to $+\infty$. This normalization introduces an undesirable source of subjectivity, which arguably goes against the normalization-free essence of DEA. In any event, given an agent a_j and $\lambda \in [0, 1]$, ZAP propose to consider the following family of convex combinations of normalized best- and worst-case DEA scores:

$$CI_{a_j}^{ZAP}(\lambda) = \lambda \frac{f_{a_j}^* - \min_{a \in \mathcal{A}} f_a^*}{\max_{a \in \mathcal{A}} f_a^* - \min_{a \in \mathcal{A}} f_a^*} + (1 - \lambda) \frac{g_{a_j}^{ZAP} - \min_{a \in \mathcal{A}} g_a^{ZAP}}{\max_{a \in \mathcal{A}} g_a^{ZAP} - \min_{a \in \mathcal{A}} g_a^{ZAP}}. \quad (9)$$

We close this section by noting that, while relatively recent, ZAP's model has already been quite influential in the literature. Indeed, a number of studies have adopted ZAP's approach to worst-case DEA for the construction of composite indices (see Hatefi and Torabi 2010; Dominguez-Serrano and Blancas 2011; Rogge 2012, among others).

3 An Alternative Approach to Worst-Case DEA

An alternative way of modeling worst-case relative performance is to maintain the structure of optimization problem (2) (i.e., its objective function and constraints) but make it a minimization as opposed to a maximization. This would involve solving the following optimization problem:

$$g_{a_j}^* \equiv \min_{w \geq 0} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{10}$$

Problem (10) is the strict worst-case analogue of problem (2). Not surprisingly, when applied to the data of Example 1 it clearly points to a_2's far superior worst-case relative performance since we have $g_{a_2}^* = 5000/9999$ vs. $g_{a_1}^* = 1/5000$).

Analytically, problem (10) is not as straightforward as (2) or (6). This is because we cannot do the same trick of Eq. (3) to reduce it to an equivalent linear-fractional program. Nonetheless, it is possible to argue from first principles that it too admits a simple and tractable solution.

For expository reasons, before going into the statement and proofs of the following results, we generalize Eq. (10) to incorporate arbitrary constraints on the weights. Letting $W_j \subseteq \mathfrak{R}_+^I$ denote an arbitrary subset of the non-negative orthant, define the optimization problem:

$$g_{a_j}^*(W_j) \equiv \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{11}$$

The pie-share bounds of Eq. (5) correspond to sets W_j that are *polyhedra*, i.e., they can be expressed as systems of linear inequalities [see Chapter 2 in Bertsimas and Tsitsiklis (1997)]. This is because Eq. (5) are equivalent to the system of linear inequalities

$$\left\{ w_i x_{a_j i} - L_i \sum_{k=1}^{I} w_k x_{a_j k} \geq 0, \quad w_i x_{a_j i} - U_i \sum_{k=1}^{I} w_k x_{a_j k} \leq 0 \right\}, \quad \text{for all } i \in \mathcal{I}.$$

We are now ready to state the paper's first theorem.

Theorem 1 *Consider optimization problem* (11) *with m linear constraints on the weights given by* $W_j = \{ w \in \mathfrak{R}^I : w \geq 0, G^j \cdot w \leq h^j \}$, *where* $G^j \in \mathfrak{R}^{m \times \mathfrak{I}}$ *and* $h^j \in \mathfrak{R}^m$. *We have*

$$g_{a_j}^*(W_j) = \min_{a \in \mathcal{A}} \left\{ \begin{array}{c} \min_{\substack{w \geq 0, \, y \geq 0 \\ G^j w - h^j y \leq 0 \\ \sum_{i=1}^{I} w_i x_{ai} = 1}} \sum_{i=1}^{I} w_i x_{a_j i} \end{array} \right\}. \tag{12}$$

Proof See "Appendix". □

Theorem 1 establishes that problem (11) is highly tractable for arbitrary polyhedral restrictions on weights. Indeed, its solution simply amounts to solving A linear programs, the inner minimizations of Eq. (12) for each $a \in \mathcal{A}$, and picking the optimal solution which is the smallest. Specifically, this means that the pie-share weight restrictions of Eq. (5) can be easily accommodated in problem (10).

Corollary 1 establishes an easy consequence of Theorem 1. In particular, when there are no constraints on the weights, problem (11) can be trivially solved by simply enumerating the ratios $\frac{x_{a_j i}}{x_{ai}}$ for all $i \in \mathcal{I}$ and $a \in \mathcal{A}$ and picking the minimum value.

Corollary 1 *Consider the setting of Theorem* 1 *with no constraints on the weights except for non-negativity (i.e.,* $W_j = \mathfrak{R}_+^I$*). In this case Eq.* (12) *can be simplified to*:

$$g_{a_j}^*(\mathfrak{R}_+^{\mathfrak{J}}) \equiv g_{a_j}^* = \min_{a \in \mathcal{A}, \, i \in \mathcal{I}} \frac{x_{a_j i}}{x_{ai}}. \tag{13}$$

Let $\mathcal{I}^*$ *denote the set of indicators that attain the minimum in Expression* (13)*. Any vector* $w_{a_j}^* \geq 0$ *such that* $\sum_{i \in \mathcal{I}^*} w_{a_j i^*}^* > 0$ *for* $i^* \in \mathcal{I}^*$ *and* $w_{a_j i}^* = 0$ *otherwise, is an optimal solution of problem* (10)*.*

Finally, it is worth noting that the positive result of Theorem 1 extends to the case of arbitrary convex constraints, which has not been previously mentioned in the DEA literature.

Theorem 2 *Consider optimization problem* (11) *for an arbitrary set* W_j*. We have*

$$g_{a_j}^*(W_j) = \min_{a \in \mathcal{A}} \left\{ \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{a_j i}}{\sum_{i=1}^{I} w_i x_{ai}} \right\}. \tag{14}$$

If W_j *can be written as* $W_j = \{w \in \mathfrak{R}_+^{\mathfrak{J}} \mid \mathfrak{c}_{\mathfrak{f}}(w) \leq , \mathfrak{f} = , , \ldots, \mathfrak{R}\}$*, where* $c_k(\cdot)$ *for* $k = 1, 2, \ldots, K$ *are convex functions, then the inner minimizations of Eq.* (14) *are concave fractional programs that can be efficiently solved with standard methods.*

Proof See "Appendix". □

In conclusion, the analytic results of this section establish that problem (11), in addition to being the (generalized) worst-case equivalent of (2), is highly tractable.

4 Numerical Case Study

In this section, we apply the framework developed in Sect. 3 to the original case study of ZAP. In their paper, Zhou and his co-authors applied their DEA methodology to the construction of a sustainable energy index (SEI) for the eighteen Asia Pacific Economic Development (APEC) economies in 2002. In what follows, we offer a bare-bones description of ZAP's SEI, omitting details on how the index was developed. This is because our primary objective is to briefly compare the results obtained under the two different DEA methodologies.

The three building blocks of ZAP's SEI are an energy efficiency indicator (EEI), a renewable energy indicator (REI), and a climate change indicator (CCI). The EEI is the reciprocal of the energy-to-GDP ratio, the REI is the percentage of renewable energy in total final energy consumption, and the CCI is the reciprocal of the CO_2 emissions-to-GDP ratio. More information on the rationale and data sources of the SEI can be found in Sect. 4 of ZAP. Table 2 summarizes data on the EEIs, REIs, and CCIs of the 18 APEC countries.

5 Application of DEA Methodologies

We begin by considering the simplest possible DEA setting in which there are no weight restrictions. Table 3 summarizes the results of the various DEA models for this case.[4]

---

[4] All computations in this section were performed in MATLAB. Details and programs available upon request.

Table 2 SEI data (see Zhou et al. 2007 for data sources)

Countries	EEI (10^3 US\$/toe)	REI (%)	CCI (10^3 US\$/tons)
Peru	13.825	53.6	4.51
Philippines	17.758	44.6	4.136
Papua	12.381	23.5	5.039
New Zealand	5.473	56.9	2.281
Vietnam	10.79	30	2.478
Canada	4.286	46.8	1.608
Chile	6.95	32.2	2.542
Japan	8.647	8.2	2.522
Mexico	8.424	9.5	2.059
Indonesia	8.516	7.8	1.784
Thailand	8.204	4.8	1.891
China	8.178	11.1	1.372
USA	5.901	6	1.614
Australia	6.208	5.6	1.425
Malaysia	5.767	4	1.442
Taiwan	5.539	2.6	1.391
Korea	4.683	0.6	1.437
Russia	2.453	11.5	0.652

Table 3 Best- and worst-case DEA scores for SEI case study (unconstrained weights)

Country	Best-case DEA score		Worst-case DEA score (Zhou et al.)		Worst-case DEA score (this paper)		Normalized best- and worst-case DEA score average (Zhou et al.)		Normalized best- and worst-case DEA score average (this paper)	
Peru	1.000	1	5.159	*1*	0.779	*2*	1.000	*1*	0.997	*2*
Philippines	1.000	1	4.964	2	0.784	*1*	0.977	2	1.000	*1*
Papua	1.000	1	3.577	3	0.413	*4*	0.810	3	0.760	3
New Zealand	1.000	1	2.231	6	0.308	6	0.648	4	0.692	4
Vietnam	0.642	6	3.134	*4*	0.492	*3*	0.529	5	0.584	5
Canada	0.823	5	1.747	*9*	0.241	*7*	0.477	6	0.537	6
Chile	0.597	7	2.817	*5*	0.391	*5*	0.463	7	0.491	7
Japan	0.558	8	2.103	*7*	0.144	*10*	0.353	8	0.306	8
Mexico	0.488	9	1.853	*8*	0.167	*9*	0.278	*9*	0.277	*10*
Indonesia	0.480	10	1.585	*10*	0.137	*11*	0.241	*10*	0.252	*11*
Thailand	0.462	11	1.511	*11*	0.084	*15*	0.220	*11*	0.207	*12*
China	0.461	12	1.467	*12*	0.195	*8*	0.214	*12*	0.277	*9*
USA	0.366	13	1.382	13	0.105	13	0.144	13	0.160	13
Australia	0.350	14	1.235	14	0.098	14	0.116	14	0.145	14
Malaysia	0.339	15	1.169	*15*	0.070	*16*	0.101	15	0.119	15
Taiwan	0.326	16	1.066	16	0.046	*17*	0.081	16	0.096	16
Korea	0.312	17	1.000	*17*	0.011	*18*	0.064	*17*	0.064	*18*
Russia	0.211	18	1.000	*17*	0.129	*12*	0.000	*18*	0.077	*17*

No pie-share bounds

The presence of italics indicates differences in the two ranks

The second column of Table 3 collects the results of best-case DEA scores as defined by Eq. (2) (which are of course identical to those cited in ZAP) along with the ranks they imply. The third column collects the results of ZAP's worst-case DEA scores as per Eq. (6), while the fourth column summarizes the worst-case DEA scores proposed in this paper as per Eq. (10). The fifth column lists the average of the normalized values of Columns 2 and 3, i.e., the values of Eq. (9) for $\lambda = 1/2$; the sixth column does the same for Columns 2 and 4, i.e using the worst-case DEA scores of this paper.[5]

Examining Table 3 we see that the choice of model (6) versus model (10) results in numerous rank changes (indicated in red). Some of them can be quite dramatic, like for instance those involving Russia, which is last according to model (6) and 12th according to model (10). Indeed, to elucidate the differences between the two methodologies it is instructive to focus on Russia and contrast its performance to that of Korea. Under model (6) Russia and Korea are considered equal as there exist weight vectors that result in their having the minimum score in group $\mathcal{A}$ of APEC countries. Denoting by $w_{a_{17}}^{ZAP}$ and $w_{a_{18}}^{ZAP}$ the optimal solutions of (6) for Korea and Russia respectively, we have

$$w_{a_{17}}^{ZAP} = (0.01, 1.52, 0.03)', \; w_{a_{18}}^{ZAP} = (0.22, 0.01, 0.49)'.$$

Conversely, under model (10) we see that Korea has a far inferior worst-case performance to Russia. The optimal weights provide insight as to why. Denoting by $w_{a_{17}}^*$ and $w_{a_{18}}^*$ the optimal solutions of (10) for Korea and Russia respectively, we have

$$w_{a_{17}}^* = (0, K_1, 0), \; w_{a_{18}}^{ZAP} = (0, 0, K_2), \text{ for any } K_1, K_2 > 0.$$

Hence, we see that for Korea (Russia), worst-case weights correspond to those assigning positive weight exclusively to indicator EEI (CCI). The result now follows since Korea's performance of 0.6 in EEI (where New Zealand has the maximum value of 56.9) is, in relative terms, worse than that of Russia which has a CCI value of 0.652 (where Papua has the maximum value of 5.039).

Qualitatively similar implications persist even when we impose the uniform pie-share bounds $(L_i, U_i) = (L, U) = (0.1, 0.5)$ for all $i \in \{1, 2, 3\}$, albeit to a weaker degree. Table 4 summarizes the corresponding results.

6 Combining Best- and Worst-Case DEA Scores

Examining the fifth and sixth columns of Tables 3 and 4, we see that the rankings implied by the two methodologies converge significantly when we consider the averages of the respective normalized DEA scores, as per Eqs. (9) and (15). Taking this analysis further, Figure 1 follows the example of ZAP's Figures 2 and 3 and presents box plots of country ranks when λ is allowed to assume all values in $\{0, 0.1., 0.2., \ldots, 0.9, 1\}$, for both cases of unconstrained and constrained weights. As expected, we observe greater variability in

---

[5] That is, it lists the values of $CI_a^*(1/2)$ for all $a \in \mathcal{A}$, where

$$CI_{a_j}^*(\lambda) \equiv \lambda \frac{f_{a_j}^* - \min_{a \in \mathcal{A}} f_a^*}{\max_{a \in \mathcal{A}} f_a^* - \min_{a \in \mathcal{A}} f_a^*} + (1 - \lambda) \frac{g_{a_j}^* - \min_{a \in \mathcal{A}} g_a^*}{\max_{a \in \mathcal{A}} g_a^* - \min_{a \in \mathcal{A}} g_a^*}. \tag{15}$$

Note that, unlike ZAP, the above normalization is not necessary because the $g_{a_j}^*$ measures defined in Eq. (10) are already scaled to range between 0 and 1. Nonetheless, we still adopt it to maximize comparability of the two sets of results.

Table 4 Best- and worst-case DEA scores for SEI case study (constrained weights)

Country	Best-case DEA score		Worst-case DEA score (Zhou et al.)		Worst-case DEA score (thispaper)		Normalized best- and worst-case DEA score average (Zhou et al.)		Normalized best- and worst-case DEA score average (this paper)	
Peru	1.000	1	5.244	1	0.916	1	1.000	1	1.000	1
Philippines	1.000	1	5.050	2	0.897	2	0.977	2	0.990	2
Papua	0.852	3	3.827	3	0.597	3	0.752	3	0.741	3
New Zealand	0.660	4	2.788	6	0.412	6	0.525	5	0.533	5
Vietnam	0.636	5	3.189	4	0.571	4	0.559	4	0.608	4
Canada	0.496	7	2.115	8	0.313	7	0.356	7	0.388	7
Chile	0.574	6	2.896	5	0.484	5	0.491	6	0.526	6
Japan	0.457	8	2.120	7	0.241	10	0.335	8	0.326	8
Mexico	0.432	9	1.945	9	0.259	8	0.300	9	0.322	9
Indonesia	0.393	10	1.731	10	0.217	11	0.254	10	0.278	11
Thailand	0.343	12	1.653	11	0.149	15	0.218	12	0.213	12
China	0.371	11	1.574	12	0.257	9	0.224	11	0.288	10
USA	0.310	13	1.418	13	0.172	12	0.172	13	0.207	13
Australia	0.295	14	1.327	14	0.159	13	0.153	14	0.192	14
Malaysia	0.263	15	1.242	15	0.122	16	0.126	15	0.154	15
Taiwan	0.217	16	1.143	16	0.084	17	0.089	16	0.107	17
Korea	0.086	18	1.000	17	0.021	18	0.000	18	0.000	18
Russia	0.186	17	1.000	17	0.153	14	0.055	17	0.128	16

Uniform pie-share bounds: U = 0.5, L = 0.1

The presence of italics indicates differences in the two ranks
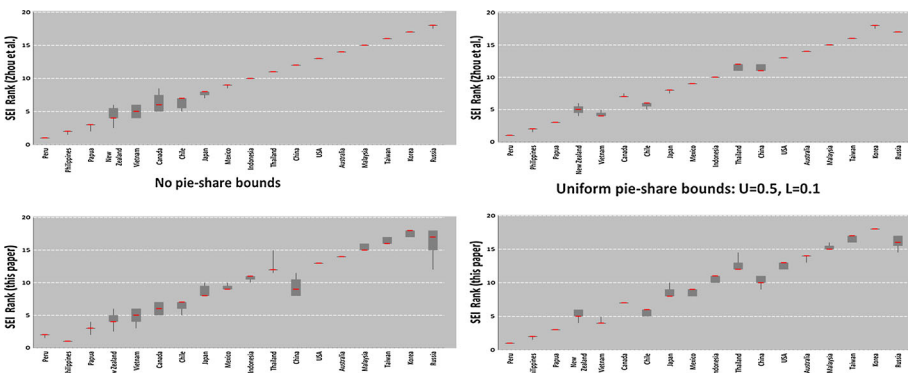

Fig. 1 Box-plots of SEI ranks when $\lambda \in \{0, 0.1., \ldots, 0.9, 1\}$. *Red lines* indicate median values, *while gray boxes* (*black lines*) denote 25th–75th (5th–95th) percentile intervals. *Left* (*right*) panels refer to the unconstrained (constrained) weights case. *Upper* (*lower*) panels plot the results corresponding to Eq. (9) [resp., Eq. (15)]. (Color figure online)

country ranks for the worst-case DEA model of Eq. (15) compared to that of (9). The effect is stronger when weights are unrestricted, but persists even upon setting the afore-mentioned pie shares.

In conclusion, this brief empirical exercise suggests that indices combining best- and worst-case DEA scores are quite sensitive to how one chooses to model worst-case relative performance.

7 Conclusion

This note has revisited the concept of worst-case performance in a nonparametric DEA framework, first introduced in the composite-indicator literature by Zhou et al. (2007) [ZAP]. We argue that, while interesting and valid in its own right, the worst-case measure adopted by ZAP does not capture, in a formal sense, the notion of worst-case DEA performance. By means of a stylized example, we showed that this theoretical inconsistency may at times lead to undesirable implications. We analyze the strict worst-case analogue of standard DEA and show how it can be tractably computed, even under general convex restrictions on the weights. Furthermore, the resulting worst-case DEA scores can be combined with their best-case analogues without requiring prior normalization. The two methodologies are compared using ZAP's Sustainable Energy Index case study, demonstrating that they occasionally lead to divergent results. Future work could incorporate the model presented herein in uncertainty and sensitivity analyses of composite indicators (Cherchye et al. 2008).

Appendix

Proofs

Theorem 1. Consider optimization problem (11). Let w^* be its optimal solution and a^* be the agent attaining the maximum in the denominator. By definition, we have

$$
\min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}} = \frac{\sum_{i=1}^{I} w_i^* x_{aji}}{\sum_{i=1}^{I} w_i^* x_{a^*i}}
$$
$$
\geq \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\sum_{i=1}^{I} w_i x_{a^*i}} \geq \min_{a \in \mathcal{A}} \left\{ \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\sum_{i=1}^{I} w_i x_{ai}} \right\}. \tag{16}
$$

On the other hand,

$$
\min_{a \in \mathcal{A}} \left\{ \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\sum_{i=1}^{I} w_i x_{ai}} \right\} \geq \min_{a \in \mathcal{A}} \left\{ \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}} \right\}
$$
$$
= \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}}. \tag{17}
$$

Putting inequalities (16)–(17) together we obtain the equality

$$
\min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\max_{a \in \mathcal{A}} \sum_{i=1}^{I} w_i x_{ai}} = \min_{a \in \mathcal{A}} \left\{ \min_{w \in W_j} \frac{\sum_{i=1}^{I} w_i x_{aji}}{\sum_{i=1}^{I} w_i x_{ai}} \right\}. \tag{18}
$$

Now since W_j is a polyhedral set with $W_j = \{w \in \mathfrak{R}^{\mathfrak{I}} \mid w \geq 0, \ G^{j}w \leq h^{j}\}$, for a given $a \in \mathcal{A}$ the minimization problem inside the brackets of Eq. (18) is a linear-fractional problem that can be transformed into the following equivalent linear program (see Section 4.3.2 in Boyd and Vandenberghe 2004):

$$
\begin{aligned}
&\min_{w \geq 0, \ y \geq 0} \quad \sum_{i=1}^{I} w_i x_{a_j i}. \\
&G^{j}w - h^{j}y \leq 0 \\
&\sum_{i=1}^{I} w_i x_{ai} = 1
\end{aligned}
\tag{19}
$$

$\square$

Corollary 1 Recall Eq. (19) and suppose $G^{j} = 0$ and $h^{j} = 0$. By standard LP theory (Bertsimas and Tsitsiklis 1997), the optimal solution of (19) will be attained (perhaps non-uniquely) at one of the extreme points of polytope $\{w \in \mathfrak{R}^{\mathfrak{I}} \mid w \geq 0, \ \sum_{i=}^{\mathfrak{I}} w_i x_{ai} = \}$. The extreme points of this polytope are given by the set of I vectors $\{w^* \in \mathfrak{R}^{\mathfrak{I}} \mid w_i^* = \frac{}{x_{ai}}, \ w_k^* = \text{ for all } k \neq i\}$. Thus, finding the optimal solution of (19) is equivalent to enumerating the terms $\frac{x_{a_j i}}{x_{ai}}$ for $i = 1, 2, \ldots I$, and picking the smallest. Hence, when $W_j = \mathfrak{R}_+^{\mathfrak{I}}$, problem (11) is equivalent to

$$
\min_{a \in \mathcal{A}} \min_{i \in \mathcal{I}} \frac{x_{i a_j}}{x_{ia}}.
$$

$\square$

Theorem 2 Eq. (14) is obtained by simply repeating Steps (16)–(17)–(18) in the proof of Theorem 1 (note that they are not in any way dependent on the structure of sets W_j). The second part of the Theorem follows from well-known results on concave fractional programming (Schaible and Shi 2004).[6]

Tables and Figures in Main Text

See Fig. 1 and Tables 2, 3 and 4.

References

Anderson, G., Crawford, I., & Leicester, A. (2011). Welfare rankings from multivariate data, a nonparametric approach. *Journal of Public Economics*, 95, 247–252.

Bertsimas, D., & Tsitsiklis, J. (1997). *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific.

Bohringer, C., & Jochem, P. E. (2007). Measuring the immeasurable—A survey of sustainability indices. *Ecological Economics*, 63(1), 1–8.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge: Cambridge University Press.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.

Cherchye, L., Knox Lovell, C. A., 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.

---

[6] Note that both the numerator and denominator of the inner minimization of Eq. (14) are affine non-zero functions.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007b). 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.

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver Software* (2nd ed.). New York: Springer.

Dominguez-Serrano, M., & Blancas, F. J. (2011). A gender wellbeing composite indicator: The best-worst global evaluation approach. *Social Indicators Research*, *102*(3), 477–496.

Fare, R., Grosskopf, S., & Hernandez-Sancho, F. (2004). Environmental performance: An index number approach. *Resource and Energy Economics*, *26*(4), 343–352.

Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management*, *47*(2), 270–283.

Hatefi, S. M., & Torabi, S. A. (2010). A common weight MCDA-DEA approach to construct composite indicators. *Ecological Economics*, *70*(1), 114–120.

Jin, J., Zhou, D., & Zhou, P. (2014). Measuring environmental performance with stochastic environmental DEA: The case of APEC economies. *Economic Modelling*, *38*, 80–86.

Ravallion, M. (2012a). Troubling tradeoffs in the human development index. *Journal of Development Economics*, *99*, 201–209.

Ravallion, M. (2012b). Mashup indices of development. *World Bank Research Observer*, *27*, 1–32.

Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The Environmental Performance Index. *Ecological Indicators*, *23*, 143–154.

Schaible, S., & Shi, J. (2004). Recent developments in fractional programming: Single-ratio and max-min case. *Nonlinear Analysis and Convex Analysis*, 493–506.

Takamura, Y., & Tone, K. (2003). A comparative site evaluation study for relocating Japanese government agencies out of Tokyo. *Socio-Economic Planning Sciences*, *37*(2), 85–102.

Zaim, O. (2004). Measuring environmental performance of state manufacturing through changes in pollution intensities: a DEA framework. *Ecological Economics*, *48*(1), 37–47.

Zhang, H. Y., Ji, Q., & Fan, Y. (2013). An evaluation framework for oil import security based on the supply chain with a case study focused on China. *Energy Economics*, *38*, 87–95.

Zhou, P., Ang, B. W., & Poh, K. L. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics*, *62*(2), 291–297.

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.

Zhu, J. (2004). A buyer-seller game model for selection and negotiation of purchasing bids: Extensions and new models. *European Journal of Operational Research*, *154*(1), 150–156.