- efficiency analysis: A new methodology for evaluating units through composite indices

Greco, Salvatore and Ishizaka, Alessio and Tasiou, Menelaos and Torrisi, Gianpiero

University of Catania, University of Portsmouth

2 January 2018

σ-μ efficiency analysis: A new methodology for evaluating units through composite indices

Salvatore Greco [a,b], Alessio Ishizaka[b], Menelaos Tasiou[c], and Gianpiero Torrisi [a,c]

[a]*Department of Economics and Business, University of Catania, Catania, Italy*
[b]*University of Portsmouth, Portsmouth Business School, Centre of Operations Research and Logistics, UK*
[c]*University of Portsmouth, Portsmouth Business School, Portsmouth, UK*

Abstract

We propose a new methodology to employ composite indicators for performance analysis of units of interest using Stochastic Multiattribute Acceptability Analysis. We start evaluating each unit by means of weighted sums of their elementary indicators in the whole set of admissible weights. For each unit, we compute the mean, μ, and the standard deviation, σ, of its evaluations. Clearly, the former has to be maximized, while the latter has to be minimized as it denotes instability in the evaluations with respect to the variability of weights. We consider a unit to be Pareto-Koopmans efficient with respect to μ and σ if there is no convex combination of μ and σ of the rest of the units with a value of μ that is not smaller, and a value of σ that is not greater, with at least one strict inequality. The set of all Pareto-Koopmans efficient units constitutes the first Pareto-Koopmans frontier. By removing this set and computing the efficiency frontier for the rest of the units, one could obtain the second Pareto-Koopmans frontier. Analogously, the third, fourth and so on Pareto-Koopmans frontiers can be defined. This permits to assign each unit to one of this sequence of Pareto-Koopmans frontiers. We measure the efficiency of each unit not only with respect to the first Pareto-Koopmans frontier, as in the classic Data Envelopment Analysis, but also with respect to the rest of the frontiers, thus enhancing the explicative power of the proposed approach. To illustrate its potential, we apply it to a case study of world happiness based on the data of the homonymous report, annually produced by the United Nations' Sustainable Development Solutions Network.

Keywords: OR in societal problem analysis · Composite Indicators · Weighting · Sigma-Mu efficiency · Stochastic Multiattribute Acceptability Analysis · Data Envelopment Analysis.

1 Introduction

In recent years, composite indicators are witnessed as increasingly popular tools for evaluating the performance of units such as countries and institutions (Becker et al., 2017). In fact, there are over 500 official composite indicators evidenced to date, mainly produced by institutions, scholars and universities, with the aim of assessing countries in a complex socio-economic phenomenon (Bandura, 2011; Yang, 2014). Understandably, their adoption by global institutions (e.g. the OECD, UN, World Bank etc.) over the past years has gradually drawn the attention of the media and policy-makers around the globe (Saltelli, 2007), and the number of applications in the literature has surged ever since (Greco et al., 2018). This spiral of attention raises several flags on issues that are still debated in the literature, mainly regarding two stages in the construction of an index; namely, the weighting and aggregation. There is a wide variety of methods available for a developer of an index to choose in these steps, with each bringing forward a solution, but with a given limitation (Gan et al., 2017). Undeniably, the choice of the proper approach lies in the developer's craftsmanship and the objective of the index (OECD, 2008). Nevertheless, these issues are still in great need of consideration; especially when something as crucial as a policy is to be drawn on the basis of a synthetic measure that could easily be 'manipulated' (see Grupp and Schubert, 2010; Abberger et al., 2017).

A fundamental step in the construction of composite indices regards the weighting of elementary indicators. Very often, this point is not taken into account and a non-weighted mean -typically the arithmetic (Karagiannis, 2017), but sometimes also the geometric one- is considered (Van Puyenbroeck and Rogge, 2017). This results in giving the same weight to all the dimensions taken into account in the composite index. By contrast, sometimes the dimensions are weighted by taking into account reasonable differences in the importance of considered dimensions (Decancq and Lugo, 2013). Either way, at first sight this procedure of weighting the indicators -with, or without equal weights- could appear as a neutral approach to the problem of aggregating the different dimensions, given a single, well-determined vector of weights. Of course, this implicitly assumes a representative agent (Hartley and Hartley, 2002), summing up in itself the preferences of all the individuals potentially interested in the composite index. However, one has to admit that in a miscellaneous group of people, each one may assign a radically different importance to the considered dimensions. Consequently, in order to ensure that the composite index is meaningful, the diversity of existing viewpoints has to be considered (Decancq et al., 2013). Undeniably, the hypothesis of the representative agent is rather stringent. Moreover, it has been long criticized in economics with the so-called *"fallacy of composition"*, proposed by Kirman (1992), who gave an example in which the representative agent disagrees with all individuals in the economy (a similar point can be found in Blackburn and Ukhov (2013), examining the relationship between individual and aggregate risk preferences in the financial markets). Besides the observation of a plurality of preferences corresponding to the individuals interested in the composite index, one has to take into account that each individual can be seen as a multiplicity of 'selves' that she is composed of (see, e.g., Elster, 1987). Several researchers have acknowledged the relevance of this point in economics (see, e.g., Ainslie, 2001; Schelling, 1980; McClure et al., 2004), so that even to represent an individual's preferences, we need to consider a set of weight vectors for the considered dimensions. Something similar happens in Multiple Criteria Decision Aiding (MCDA) (for an updated survey see Greco et al., 2016). Indeed, some recently-introduced MCDA models consider a plurality of value functions compatible with the preferences expressed by a decision maker (see, e.g., Greco et al., 2008, 2010; Corrente et al., 2013), or even a probability distribution in the set of value functions (see, e.g., Corrente et al., 2016b). This can be interpreted as a plurality of selves for each individual, from the point of view that each considered value function is a specific 'self'. Similar arguments hold for multi-prior models proposed for decisions under uncertainty, where each individual takes a decision considering a plurality of probability distributions on the state of the words (see, for example, Gilboa and Schmeidler, 1989; Bewley, 2002; Gilboa et al., 2010). These arguments suggest to abandon the idea of a single, allegedly well-defined weighting of dimensions. Indeed, by taking into account the whole set of admissible weight vectors, one can consider the whole spectrum of preferences of individuals, as well as multiple selves within each individual interested in the composite index. With respect to the domain of composite indices, this approach was recently proposed by Greco et al. (2017a) using Stochastic Multiattribute Acceptability Analysis (SMAA) (Lahdelma et al., 1998; Lahdelma and Salminen, 2001). More specifically, by considering a probability distribution on the set of feasible weight vectors, SMAA reveals the probability that a unit attains a given ranking position, as well as the probability that a given unit is better than another.

In this paper we will argue that another possible use of the plurality of weight vectors is to consider for each unit the mean value (μ) of the composite index and its variability -measured by the standard deviation (σ)- in the space of feasible weight vectors. Of course, the former is supposed to be maximized, while the latter is to be minimized, as higher values of σ denote more volatile overall performance attributed to changes in the weight vectors. Consequently, by considering the mean value and the standard deviation, it is straightforward to define a dominance relation as follows: unit a is $\sigma - \mu$ Pareto-dominating unit b if the mean value of a is not smaller than that of b and the standard deviation of a is not greater than the that of b, with at least one of these two inequalities being strict. Thus, unit a will be $\sigma - \mu$ Pareto-efficient if there is no other unit $\sigma - \mu$ Pareto-dominating it with respect to the former inequalities. Analogously obtaining the set of all efficient units permits to constitute the $\sigma - \mu$ Pareto-efficiency frontier. Consideration of the mean value and the standard deviation along with the related dominance and efficiency concepts clearly reminds the Markowitz mean-variance analysis (Markowitz, 1952), which formed the foundations of modern portfolio theory (Elton et al., 2009). However, we are not only interested in finding dominating solutions (i.e. alternatives lying on the Pareto-efficiency frontier), but in measuring the efficiency of each unit with respect to the frontier. In the domain of Operations Research this naturally leads to the consideration of Data Envelopment Analysis (DEA) (Charnes et al., 1978a; Cooper et al., 2011), which brings us to acknowledge another definition of efficiency, taking into account this time the possibility to combine different units. More specifically, in this case, unit a is Pareto-Koopmans efficient (Charnes et al., 1985) if there is no linear combination of the mean values (μ) and standard deviations (σ) of the rest of units dominating a. Moreover, following an approach that was recently presented in a companion paper for the whole DEA methodology (see Greco et al., 2017b), we are interested in decomposing the set of considered units in a family of Pareto-efficiency frontiers, as well as in a family of Pareto-Koopmans efficiency frontiers. For instance, considering the Pareto efficiency, the first frontier is the $\sigma - \mu$ Pareto efficiency frontier above-introduced, the second frontier is the $\sigma - \mu$ Pareto-efficiency frontier obtained once the units of the previous frontier have been removed, and so on until all the remaining units are $\sigma - \mu$ efficient. Of course, an analogous procedure holds for the computation of all Pareto-Koopmans efficiency frontiers. This idea of a sequence of Pareto frontiers has been considered within the celebrated evolutionary multi-objective optimization algorithm NSGA-II (Deb et al., 2002). In this case, we adopt this idea of successive efficiency frontiers not to guide a multi-objective optimization process, but to measure and analyze the efficiency of units with respect to the considered composite indicators. More specifically, we introduce an efficiency measure with respect to each frontier in the above-mentioned sequence. This measure takes a positive value when the unit is dominating the considered efficiency frontier, and a negative value if the unit is dominated by the efficiency frontier instead. Moreover, we define an overall efficiency by aggregating the efficiency measures corresponding to the efficiency frontiers in the sequence.

This paper introduces the σ-μ efficiency analysis, illustrating its potential in a case study of world happiness, based on the homonymous report by Helliwell et al. (2017). In the following, Section 2 describes in more detail the issues of weighting in the construction of a composite index. Section 3 introduces the σ-μ efficiency analysis, followed by a brief didactic example to illustrate its application on a step-by-step basis in Section 4. Section 5 contains the case study of world happiness and Section 6 provides conclusive remarks and future direction of research.

2 Composite Indicators: Issues on weights and aggregation

2.1 Weighting dimensions in composite indicators

Despite the severe criticism in their inauguration (Sharpe, 2004), the use of composite indicators is constantly growing by the day, with an ever-increasing number of composite measures produced every year by global institutions, academics and media around the world (Bandura, 2011; Yang, 2014). This is mainly owed to their irresistible property of summarizing complex phenomena with a sole number that can be easily interpreted as a benchmark (Saisana et al., 2005). Of course, this can be seen as both an asset and a liability at the same time. More specifically, lack of transparency in their construction allows significant room for 'manipulation' (Grupp and Schubert, 2010; Abberger et al., 2017). The reason being is that there exists a sequence of steps in the construction of an index, and admittedly, different choices in each step might radically alter the final outcome. As one would expect, not a single step in the construction of an index lacks criticism (Booysen, 2002); nevertheless, the paramount critique lies in two stages, namely the weighting and aggregation. The former refers to the process of declaring the importance of index dimensions, whereas the latter refers to the final synthesis of the overall index. In this paper we are engrossed with the former, thus discussion will solely revolve around it.

The basic model of composite indicators is the following. There exists a set of units $I = \{1, \ldots, n\}$ to be evaluated with respect to the set of dimensions $J = \{1, \ldots, m\}$, the values of which are x_{ij}. For each unit $i \in I$, the vector $\mathbf{x}_i = [x_{i1}, \ldots, x_{im}]$ collects the values assigned to that unit in the dimensions from J. To each dimension $j \in J$, a weight, w_j, is attached such that $w_j \geq 0$ for all $j \in J$ and $\sum_{j=1}^{m} w_j = 1$. Given a weight vector $\mathbf{w} = [w_1, \ldots, w_m]$, the composite index assigns the following value to each unit $i \in I$:

$$CI(\mathbf{x}_i, \mathbf{w}) = \sum_{j=1}^{n} x_{ij} w_j.$$

The authoritative *Handbook on Constructing Composite Indicators* (OECD, 2008) lists several approaches regarding the weighting procedure in the construction of a composite index (for a recent review of existing methodologies, criticism and proposed solutions, see Greco et al., 2018), with equal weighting being the most frequent approach (Paruolo et al., 2013). This, however, also appears to be the most criticized (Decancq and Lugo, 2013). More specifically, assignment of equal weights can be seen as a convenient solution of the last resort (Chowdhury and Squire, 2006), mainly when there is no scientific basis to justify peculiar weighting, or when an alleged 'objectivity' is desired (OECD, 2008). This rationale has been contradicted in the literature for the following two reasons. First, equal weights could be reasonably considered subjective as well as objective (see, e.g., Ray, 2008; Mikulić et al., 2015). Second, there are other, potentially more realistic solutions to deal with uncertainty in the lack of decision-makers' preferences on weights (see, e.g., Doumpos et al., 2016, 2017; Greco et al., 2017a). Other past solutions revolve around two sets of approaches, often characterized as 'subjective', and 'objective' respectively (Decancq and Lugo, 2013). The former involve participatory techniques such as the Budget Allocation Process (BAP) (see OECD, 2008, p.96) or Analytic Hierarchy Process (AHP) (Saaty, 1977, 1980). These engage a single, or a number of stakeholders (e.g. a panel of experts) to decide upon the weights to be assigned, according to their beliefs/expertise (hence, the term 'subjective'). These approaches appear to be ideal where a well-defined framework for national policy exists (see Munda, 2005b). Yet, they might yield radically different results (see Saisana et al., 2005, p.314, for a comparison between AHP and BAP), while in the presence of many criteria, they can give 'cognitive stress' decision-makers that is amplified in AHP, due to the number of pairwise comparisons required (Ishizaka and Nemery, 2013). The second set of approaches are awarded their epithet 'objective' from the fact that they do not rely on human judgement, but rather on the use of data-driven techniques (e.g. Multiple linear regression analysis, Principal Component Analysis (Pearson, 1901), Factor Analysis (Spearman, 1904), or Data Envelopment Analysis (Charnes et al., 1978b)). These have been conceptually criticized for being disoriented from the objective at hand, or that they provide unrealistic results (Decancq and Lugo, 2013), while they have a few methodology-related drawbacks that need to be addressed (Greco et al., 2018).

Irrespectively of classification ('subjective', or 'objective'), all the above approaches produce a single weight vector that is used in the stage of aggregation to synthesize the composite index. While this procedure is common practice in the domain of composite indicators (OECD, 2008), either unwittingly or deliberately, the developer assumes that the obtained univocal set of weights is representative of the whole population interested in the composite index. Understandably, one could argue that this is a rather stringent assumption, as in a miscellaneous group of people, each individual may assign a radically different importance to each dimension, and the representativeness assumption may be only valid for a very small part of the population, or it could even become infeasible overall. Decancq et al. (2013) argue that when a policy-maker chooses a weight-vector there are several individuals who are inevitably 'worse-off'. This situation highly resembles the case of the representative agent in economics (see e.g. Hartley and Hartley, 2002), which has been long criticized in the literature by Kirman (1992). Kirman provides an example in which, quaintly to his title, the 'representative' agent disagrees with all the individuals in the economy. Acknowledging this confounding situation, Greco et al. (2017a) recently proposed the use of SMAA (Lahdelma et al., 1998; Lahdelma and Salminen, 2001) to take into account the whole set of possible weight vectors in the evaluation process. According to the authors, the standard procedure of choosing a single weight vector produces a single, allegedly 'representative' ranking for the units evaluated which "*amalgamates different preferences in the population*" (p.6). SMAA permits the inclusion of several potential viewpoints in the decision-making process, e.g. in the form of weight vectors, enriching in this way the single ranking that is obtained from a single preference. In terms of output, probabilistic rankings are assigned to each unit, expressing its probability to be ranked first, second etc.; or, its probability to be preferred from another unit. The use of SMAA in this exercise seems alluring, whether it is applied to take into account potential representations of citizens' preferences (Greco et al., 2017a), or simply to deal with uncertainty in the lack of information about decision-makers' preferences (see e.g. Doumpos et al., 2016, 2017). Since SMAA is the fundamental framework that we take into account in this paper, we present it in more detail in the following subsection.

2.2 Stochastic Multiattribute Acceptability Analysis (SMAA)

SMAA offers a solid solution to real-world decision-making that is surrounded by any source of uncertainty. In the domain of composite indicators, such an example would involve a decision-maker that is unable to provide the parameters required for the evaluation process (see e.g. Doumpos et al., 2016, 2017). In this paper we are engrossed with the step of weighting, hence, we are solely considering this source of uncertainty. Essentially, SMAA takes it into account by considering a probability distribution f_w over the space of all weight vectors

$$W = \{\mathbf{w} = [w_1, \dots, w_m] : w_j \geq 0, j = 1, \dots, m, \sum_{j=1}^{m} w_j = 1\}.$$

Understandably, if a different importance has to be assigned to the dimensions from J, the space W

is transformed accordingly. For instance, if the dimension $j_{(1)}$ is the most important, $j_{(2)}$ the second most important and so on until the least important, $j_{(m)}$; we have to assign higher weights to the more important dimensions, thus the space W is transformed as follows:

$$W = \{\mathbf{w} = [w_1, \ldots, w_m] : w_{j_{(1)}} \geq w_{j_{(2)}} \geq \ldots w_{j_{(m)}} \geq 0, j = 1, \ldots, m, \sum_{j=1}^{m} w_j = 1\}.$$

As the composite index $CI(\mathbf{x}_i, \mathbf{w})$ provides a ranking for each $\mathbf{w}$ in W, SMAA calculates the position attained by each unit, i, as follows:

$$\text{rank}(i, \mathbf{w}) = 1 + \sum_{i' \neq i} \rho(CI(\mathbf{x}_{i'}, \mathbf{w}) > CI(\mathbf{x}_i, \mathbf{w})),$$

where $\rho(\text{true}) = 1$, $\rho(\text{false}) = 0$. Likewise, for every $i \in I$, SMAA defines the favorable rank weights of unit $i \in I$

$$W_i^r = \{\mathbf{w} \in W : rank(i, \mathbf{w}) = r\}$$

being the set of feasible weights that position unit i in the r^{th} place, $r = 1, ..., n$, in the final rank. Finally, SMAA delivers the ranking acceptability indices, the central weight vectors and the pair-wise winning indices as follows:

- Ranking Acceptability Index (RAI) for unit $i \in I$ and r^{th} position, $r = 1, ..., n$,

  $$b_i^r = \int_{\mathbf{w} \in W_i^r} f_w(\mathbf{w})\, d\mathbf{w}$$

  RAI illustrates the proportion of weight vectors $\mathbf{w} \in W$ giving unit i the r^{th} position in the obtained final ranking. For instance, b_i^1 represent the share of weight vectors for which unit i takes the first position.

- Provided $b_i^1 \neq 0$, Central Weight Vector (CWV) for unit i

  $$\mathbf{w}_i^c = \frac{1}{b_i^1} \int_{\mathbf{w} \in W_i^1(\xi)} f_w(\mathbf{w})\mathbf{w}\, d\mathbf{w}$$

  CWV represents the weight vector of a potential decision-maker, according to whom unit i is the best.

- Pairwise Winning Index (PWI) for units i and i'

  $$p_{ii'} = \int_{w \in W : rank(i, \mathbf{w}) > rank(i', \mathbf{w})} f_w(\mathbf{w}) d\mathbf{w}$$

  PWI (Tervonen et al., 2009b; Leskinen et al., 2006) shows the probability that unit i is better than unit i'.

For some recent papers utilizing SMAA in the MCDA context, the reader is referred, among others, to Durbach (2009); Lahdelma and Salminen (2009); Tervonen et al. (2009a,c); Menou et al. (2010);

Aertens et al. (2011); Corrente et al. (2014); Angilella et al. (2015), while for a comprehensive review, see Tervonen and Figueira (2008). SMAA was only recently introduced in the field of composite indicators. More specifically, Doumpos et al. (2016) use it to deal with the uncertainty arising from the lack of information regarding the parameters to be used in the evaluation process of some financial institutions. Using 10,000 uniformly distributed random weights and marginal value functions, the authors evaluate the overall financial strength of 1,200 commercial banks through an additive value function setting, given five financial characteristics from the CAMEL framework. A similar application is found in Doumpos et al. (2017), comparing the overall financial strength of Islamic and conventional banks. Greco et al. (2017a) propose the use of SMAA in the context of composite indicators as a way to deal with the issue of representativeness inherent in the single weight vector. The authors evaluate the 20 regions of Italy, based on 65 socio-economic criteria. By enlarging the space of weight vectors, they refrain from the classic setting of the univocal set of weights, including 1,000,000 uniformly distributed weight vectors. In an alternative interpretation, these could be potentially seen as an expression of several decision-makers' preferences, e.g. ranging from policymakers to citizens, regarding the importance of the index dimensions. This involvement of a 'multiplicity of participants', or even 'selves' (see Elster, 1987) could indeed be enriching to consider in such an exercise. Quoting Munda (2005a, p.132): *"when science is used in policy, the appropriate management of quality has to be enriched to include this multiplicity of participants and perspectives"*. While the author's point refers to the context of a sustainability policy exercise (regarding the objectives and scales of such an analysis and the set of dimensions to be used in the evaluation process), the intended allegory is astonishingly fit to the context of the decision-makers' number and preferences respectively.

3 The σ-μ efficiency

We stand by the principle that a meaningful composite index should ideally reflect a multiplicity of viewpoints. Technically speaking, this can be achieved in the weighting stage, in which individuals that the index is concerning can participate, by expressing their preferences on the importance of index dimensions. These individuals could constitute different clusters, e.g. experts, policy-makers, or even citizens at whom policies are addressed. Therefore, the main driver of this concept refrains from the classic scheme of a single, allegedly representative weight vector in the construction of an index, by taking into account all these individuals' viewpoints. With this aim in mind, we re-consider the framework of SMAA, though, instead of focusing on the probability of obtaining a given ranking position, or the probability that a unit is better than another; for each unit, $i \in I$, we synthesize the distribution of its composite indicators values, $CI(\mathbf{x}_i, \mathbf{w})$, by computing its mean value μ_i and standard deviation σ_i in the weight vector space W, that is

$$\mu_i = \int_{\mathbf{w} \in W} f_w(\mathbf{w}) CI(\mathbf{x}_i, \mathbf{w}) \, d\mathbf{w}, \tag{1}$$

$$\sigma_i = \sqrt{\int_{\mathbf{w} \in W} f_w(\mathbf{w}) \left[CI(\mathbf{x}_i, \mathbf{w}) - \mu_i \right]^2 \, d\mathbf{w}}. \tag{2}$$

Understandably, μ_i is intended to be maximized because it represents the average evaluation of a unit taking into account the variability of the weight vectors $\mathbf{w}$. Instead, σ_i has to be minimized, as it exhibits the instability in the overall evaluations with respect to the variability of weights. Let us observe that, in some form, this reminds us of the same reasoning explicit in the Markowitz model (Markowitz, 1952). Following his influential theory, by taking into account the mean, μ_i, and the standard deviation,

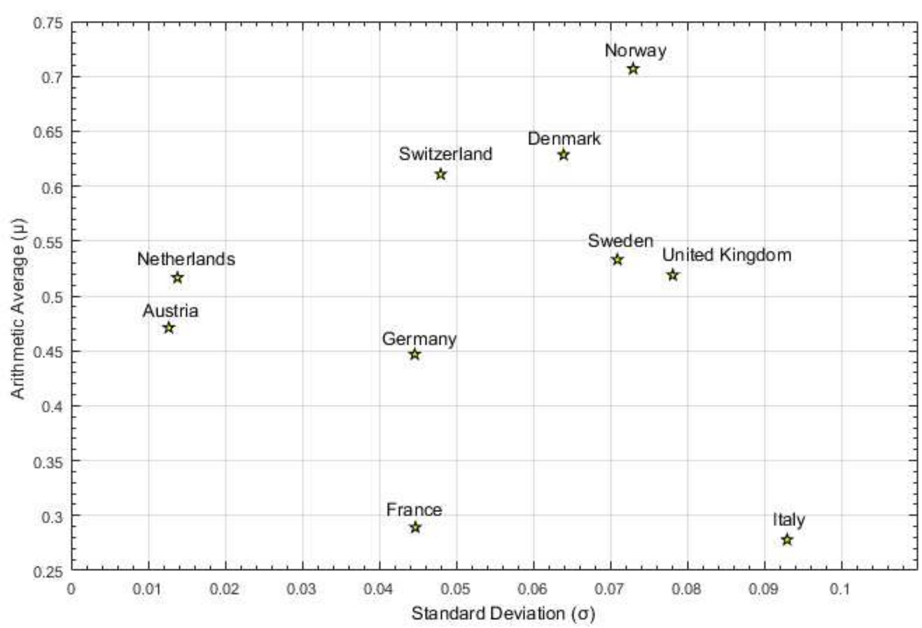
σ_i, one can draw a plane that units $i \in I$ are plotted on, pending evaluation. To be consistent with the proposed concept of $\sigma - \mu$ efficiency analysis, we will refer to this throughout the text as 'The $\sigma - \mu$ plane' (illustrated in Figure 1) which shows the standard deviation σ (on the x axis) and the mean μ (on the y axis) of ten European countries with respect to the data of the 2017 World Happiness Report (WHR) (Helliwell et al., 2017) that will be detailed in Section 4. Moreover, one can define a $\sigma-\mu$ Pareto dominance relation on the set of units I as follows: for all $i, i' \in I$, unit i is Pareto dominating unit i' if $\mu_i \geq \mu_{i'}$ and $\sigma_i \leq \sigma_{i'}$, with at least one of the two inequalities being strict. A unit $i \in I$ is $\sigma - \mu$ Pareto efficient if there is no other unit dominating it. The set of all Pareto efficient units constitutes the Pareto frontier. A concept stricter than $\sigma-\mu$ Pareto efficiency is the $\sigma-\mu$ Pareto-Koopmans efficiency (Charnes and Cooper, 1962). A unit $i \in I$ is $\sigma - \mu$ Pareto-Koopmans efficient if there is no convex combination of $\mu_{i'}$ and $\sigma_{i'}$ of the remaining units, $i' \neq i$, with a mean value μ that is not smaller, and a standard deviation σ that is not higher, with at least one of these inequalities being strict. Formally, a unit $i \in I$ is $\sigma - \mu$ Pareto-Koopmans efficient if for all $[\lambda_{i'}, i' \neq i]$, with $\lambda_{i'} \geq 0$, for all $i' \neq i$ and $\sum_{i' \neq i} \lambda_{i'} = 1$, neither (3) nor (4) hold:

$$\sum_{i' \neq i} \lambda_{i'} \mu_{i'} > \mu_i \text{ and } \sum_{i' \neq i} \lambda_{i'} \sigma_{i'} \leq \sigma_i \tag{3}$$

$$\sum_{i' \neq i} \lambda_{i'} \mu_{i'} \geq \mu_i \text{ and } \sum_{i' \neq i} \lambda_{i'} \sigma_{i'} > \sigma_i. \tag{4}$$

Figure 1: The $\sigma - \mu$ plane

Units $i \in I$ are plotted in the plane with coordinates (σ_i, μ_i). The $\sigma - \mu$ analysis hereby presented concerns ten EU countries evaluated with respect to the data of the 2017 World Happiness Report (WHR) (Helliwell et al., 2017) as explained in Section 4.



The set of all $\sigma-\mu$ Pareto-Koopmans efficient units constitutes the $\sigma-\mu$ Pareto-Koopmans frontier. The membership of a unit $i \in I$ to the Pareto-Koopmans effciency frontier can be verified with a direct or an indirect procedure described below.

The direct procedure verifies that there exist no unit -obtained as linear combination of mean $\mu_{i'}$ and standard deviation $\sigma_{i'}$- dominating unit i. This is obtained by considering the following LP problem:

$$
\begin{aligned}
&\varepsilon_i^* = \text{Max } \varepsilon \\
&s.t. \\
&\begin{cases}
\sum_{i' \neq i} \lambda_{i'} \mu_{i'} \geqslant \mu_i + \varepsilon \\
\sum_{i' \neq i} \lambda_{i'} \sigma_{i'} \leqslant \sigma_i - \varepsilon \\
\lambda_i' \geqslant 0, \ \forall i' \neq i \\
\sum_{i' \neq i} \lambda_{i'} = 1
\end{cases}
\end{aligned}
$$

where unit i is $\sigma - \mu$ Pareto-Koopmans efficient if $\varepsilon_i^* \leqslant 0$.

The indirect procedure to test the $\sigma - \mu$ Pareto-Koopmans efficiency requires to consider the following LP problem:

$$
\begin{aligned}
&\delta_i^* = \text{Max } \delta \\
&s.t. \\
&\begin{cases}
\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i'} - \beta\sigma_{i'} + \delta, \ \forall i' \neq i \\
\alpha, \beta \geqslant 0 \\
\alpha + \beta = 1
\end{cases}
\end{aligned} \tag{5}
$$

which can be interpreted as follows. An evaluation $\alpha\mu_{i'} - \beta\sigma_{i'}$, with $\alpha, \beta \geqslant 0$ and $\alpha + \beta = 1$, is assigned to all units $i' \in I$. The non-negative coefficient α for the mean $\mu_{i'}$ and the non-positive coefficient $-\beta$ for the standard deviation $\sigma_{i'}$ are coherent with the idea that $\mu_{i'}$ is intended to be maximised and $\sigma_{i'}$ is intended to be minimised. Therefore, ideally the greater $\alpha\mu_{i'} - \beta\sigma_{i'}$, the better the unit i' performs with respect to $\mu_{i'}$ and $\sigma_{i'}$. The LP problem verifies whether a pair (α, β) exists, for which unit $i \in I$ receives an evaluation that is not worse than the remaining units, $i' \neq i$, that is if $\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i'} - \beta\sigma_{i'} + \delta, \ \forall i'$, with a non-negative value of δ. This happens if $\delta_i^* \geqslant 0$ which, for the units belonging to the $\sigma - \mu$ Pareto-Koopmans efficiency frontier, represents the margin that can be subtracted to the overall evaluation $\alpha\mu_i - \beta\sigma_i$ of unit i maintaining the maximality of its evaluation with respect to all other units $i' \neq i$. For all units $i \in I$ that do not belong to the $\sigma - \mu$ Pareto-Koopmans efficiency frontier, the greater the absolute value of δ_i^*, the greater the margin that has to be added to $\alpha\mu_i - \beta\sigma_i$, in order to attain the evaluation $\alpha\mu_{i'} - \beta\sigma_{i'}$ of the units belonging to the $\sigma - \mu$ Pareto-Koopmans efficiency frontier. In this sense, the value of δ_i^* can be interpreted as a measure of efficiency of unit $i \in I$ with the following characteristics:

- if δ_i^* is non-negative, then unit i is efficient, with higher values of δ_i^* indicating higher efficiency for i,
- if δ_i^* is non-positive, then unit i is inefficient, with higher values of $|\delta_i^*|$ indicating greater inefficiency for i.

For this reason, in the following we shall refer to δ_i^* as the $\sigma - \mu$ Pareto-Koopmans efficiency of unit i.

The following proposition enunciates the equivalence between the direct and the indirect test of the $\sigma - \mu$ Pareto-Koopmans efficiency.

Proposition 1. $\delta_i^* \geqslant 0$ if and only if $\varepsilon_i^* \leqslant 0$

Proof.
Let us start by proving that if $\delta_i^* \geqslant 0$ then $\varepsilon_i^* \leqslant 0$.
If $\delta_i^* \geqslant 0$, then there exists $\alpha, \beta \geqslant 0$, with $\alpha + \beta = 1$, for which:

$$\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i'} - \beta\sigma_{i'} \text{ for all } i' \neq i.$$

Therefore, for all $\lambda = [\lambda_{i'}, i' \neq i]$ with $\lambda_{i'} \geqslant 0$, for all $i' \neq i$, and $\sum_{i' \neq i} \lambda_{i'} = 1$, we have:

$$\lambda_{i'}(\alpha\mu_i - \beta\sigma_i) \geqslant \lambda_{i'}(\alpha\mu_{i'} - \beta\sigma_{i'}) \text{ for all } i' \neq i \tag{6}$$

By (6) we can get the following:

$$\sum_{i' \neq i} \lambda_{i'}(\alpha\mu_i - \beta\sigma_i) \geqslant \sum_{i' \neq i} \lambda_{i'}(\alpha\mu_{i'} - \beta\sigma_{i'})$$

and, consequently,

$$\alpha\mu_i - \beta\sigma_i \geqslant \alpha \sum_{i' \neq i} \lambda_{i'}\mu_{i'} - \beta \sum_{i' \neq i} \lambda_{i'}\sigma_{i'}$$

.
This implies that the following condition is not verified

$$\begin{cases} \sum_{i' \neq i} \lambda_{i'}\mu_{i'} \geqslant \mu_i \\ \sum_{i' \neq i} \lambda_{i'}\sigma_{i'} \leqslant \sigma_i \end{cases}$$

with at least one strict inequality.

This amounts to the Pareto-Koopmans efficiency of unit i, so that we have $\varepsilon^* \leq 0$. Thus, we proved that if $\delta_i^* \geqslant 0$, then $\varepsilon_i^* \leqslant 0$. Let us now prove that if $\varepsilon_i^* \leqslant 0$, then $\delta_i^* \geqslant 0$.

For a given unit, i, let us consider the pair (σ_i, μ_i) and the two following sets:

- the set $P^+(\sigma_i, \mu_i)$ of all the pairs $(\sigma, \mu) \in \mathbf{R}_+^2$ Pareto dominating (σ_i, μ_i), that is

$$P^+(\sigma_i, \mu_i) = \{(\sigma, \mu) \in \mathbf{R}_+^2 : \sigma \leq \sigma_i \text{ and } \mu \geq \mu_i \text{ with at least one strict inequality } \}$$

- the set $P^-(\sigma_i, \mu_i)$ given by the convex hull of the pairs $(\sigma_{i'}, \mu_{i'})$ with $i' \neq i$, that is

$$P^-(\sigma_i, \mu_i) = \{(\sum_{i' \neq i} \lambda_{i'}\mu_{i'}, \sum_{i' \neq i} \lambda_{i'}\sigma_{i'}) : \lambda_{i'} \geq 0 \text{ for all } i' \neq i \text{ and } \sum_{i' \neq i} = 1\}.$$

Let us remember that the condition $\varepsilon_i^* \leqslant 0$ implies that (σ_i, μ_i) is Pareto-Koopmans efficient. This means that there is no pair $(\sigma, \mu) \in \mathbf{R}_+^2$ being a convex combination of the pairs $(\sigma_{i'}, \mu_{i'}) \in \mathbf{R}_+^2$, $i' \neq i$ that is dominating (σ_i, μ_i). As the set of pairs $(\sigma, \mu) \in \mathbf{R}_+^2$ dominating (σ_i, μ_i) is $P^+(\sigma_i, \mu_i)$ and the set of convex combinations of the pairs $(\sigma_{i'}, \mu_{i'}), i' \neq i$, is $P^-(\sigma_i, \mu_i)$, the Pareto-Koopmans efficiency of (σ_i, μ_i) amounts to the condition that $P^+(\sigma_i, \mu_i)$ and $P^-(\sigma_i, \mu_i)$ are disjoint. Let us point out that both $P^+(\sigma_i, \mu_i)$ and $P^-(\sigma_i, \mu_i)$ are convex sets in $\mathbf{R}^2$. Therefore, for the hyperplane separating theorem (see e.g. Boyd and Vandenberghe (2004), there must be a hyperplane separating $P^+(\sigma_i, \mu_i)$ from $P^-(\sigma_i, \mu_i)$

in the $\sigma - \mu$ space. In fact, this means that there exists a straight line $\alpha\mu - \beta\sigma = \gamma$, such that:

$$\alpha\mu - \beta\sigma > \gamma$$

for all $(\sigma, \mu) \in P^+(\sigma_i, \mu_i)$, and

$$\alpha\mu - \beta\sigma < \gamma$$

for all $(\sigma, \mu) \in P^-(\sigma_i, \mu_i)$. For contradiction, suppose now that $\delta_i^* < 0$. This means that for all $\alpha, \beta \geq 0$ we have

$$\alpha\mu_i - \beta\sigma_i < \alpha\mu_{i'} - \beta\sigma_{i'}$$

for at least one $i' \neq i$. Thus, for all $\gamma \in \mathbf{R}$

$$\alpha\mu_i - \beta\sigma_i > \gamma$$

implies

$$\alpha\mu_{i'} - \beta\sigma_{i'} > \gamma$$

for at least one $i' \neq i$. But $(\sigma_{i'}, \mu_{i'}) \in P^-(\sigma_i, \mu_i)$ and therefore, there cannot exist any hyperplane

$$\alpha\mu - \beta\sigma = \gamma$$

separating $P^+(\sigma_i, \mu_i)$ from $P^-(\sigma_i, \mu_i)$. Thus, in this case the pair (σ_i, μ_i) is not $\sigma - \mu$ Pareto-Koopmans efficient. So, if $\varepsilon_i^* \leqslant 0$ and, consequently (σ_i, μ_i) is efficient, then $\delta_i^* \geqslant 0$.
$\square$

The $\sigma - \mu$ Pareto-Koopmans efficiency δ_i^* of unit $i \in I$ refers to the $\sigma - \mu$ Pareto-Koopmans efficiency frontier. However, for a unit that is quite remote from the $\sigma - \mu$ Pareto-Koopmans efficiency frontier, it might not be very meaningful to compare it with units of that frontier, as they could be seen as potentially implausible benchmarks. Instead, it could be useful to compare these remote units with their counterparts that are closer to them in the $\sigma - \mu$ plane, and as such, constitute more realistic benchmarks. This suggests taking into consideration the idea of a sequence of efficiency frontiers considered within the celebrated evolutionary multi-objective optimization algorithm NSGA-II (Deb et al., 2002).

A first sequence of $\sigma - \mu$ efficiency frontiers can be defined by taking into consideration the Pareto dominance. In this perspective, the set of all $\sigma - \mu$ Pareto-efficient units constitutes the first $\sigma - \mu$ Pareto efficiency frontier, denoted by PF_1. Removing PF_1 from I and computing again the $\sigma - \mu$ Pareto efficiency frontier for the remaining units, we get the second $\sigma - \mu$ Pareto-efficiency frontier denoted by PF_2. The third $\sigma - \mu$ Pareto efficiency frontier, PF_3, and the following ones can be computed analogously.

The sequence of Pareto efficiency frontiers $PF_1, PF_2, \ldots$ based on the concept of Pareto dominance is used in NSGA-II (Deb et al., 2002). However, for the sake of our analysis, an analogous sequence of efficiency frontiers based on the concept of Pareto-Koopmans dominance seems more appropriate. We call the efficiency frontiers of this new sequence first $\sigma - \mu$ Pareto-Koopmans efficient frontier, denoted by PKF_1, second $\sigma - \mu$ Pareto-Koopmans efficiency frontier, denoted by PKF_2, and so on. Let us denote by $\mathbf{PKF} = \{PKF_1, \ldots, PKF_p\}$ the set of all the $\sigma - \mu$ Pareto-Koopmans efficiency frontiers. For each unit $i \in I$, and for each $\sigma - \mu$ Pareto-Koopmans efficiency frontier $PKF_k \in \mathbf{PKF}$, we can define a $\sigma - \mu$ Pareto-Koopmans efficiency δ_{ik} with respect to PKF_k as follows:

$$
\begin{aligned}
&\delta_{ik} = \text{Max } \delta \\
&s.t. \\
&\begin{cases}
\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i'} - \beta\sigma_{i'} + \delta, \ \forall i' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h \\
\alpha, \beta \geqslant 0 \\
\alpha + \beta = 1
\end{cases}
\end{aligned} \tag{7}
$$

The above LP problem verifies whether there exists a pair (α, β), for which unit $i \in I$ receives an evaluation $\alpha\mu_i - \beta\sigma_i$ which is not worse than the analogous evaluation of the rest of the units $i' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h$, that is, all the units i' belonging to the k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier, or to a better $\sigma - \mu$ Pareto-Koopmans efficiency frontier. This happens if $\delta_{ik} \geqslant 0$. Instead, if $\delta_{ik} < 0$, then unit i belongs to a $\sigma - \mu$ Pareto-Koopmans efficiency frontier worse than PKF_k, that is, if $i \in PKF_h$ with $h = k+1, \ldots, p$. The interpretation of δ_{ik} with respect to the k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier is analogous to the interpretation δ_i^* with respect to the overall $\sigma - \mu$ Pareto-Koopmans efficiency frontier. More precisely, for the units in the k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier or better, $\delta_{ik} \geqslant 0$ represents the margin that can be subtracted from the overall evaluation $\alpha\mu_i - \beta\sigma_i$ of unit i maintaining an evaluation that is superior to all units in the k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier or worse. Instead, for all units $i \in I$ belonging to the k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier or worse, the absolute value of $\delta_i^* < 0$ represents the margin that has to be added to $\alpha\mu_i - \beta\sigma_i$, in order to obtain the same evaluation of at least one unit belonging to k-th $\sigma - \mu$ Pareto-Koopmans efficiency frontier or better. Therefore, as δ_i^* constitutes an efficiency measure with respect to the overall $\sigma - \mu$ Pareto-Koopmans efficiency frontier (that, in fact, corresponds to the first $\sigma - \mu$ Pareto-Koopmans efficient frontier), δ_{ik} constitutes an efficiency measure with respect to the overall k^{th} $\sigma - \mu$ Pareto-Koopmans efficiency frontier. For this reason, in the following we shall refer to δ_{ik} as $\sigma - \mu$ Pareto-Koopmans efficiency of unit i with respect to the k^{th} frontier.

The following proposition gives a simple, yet useful result with respect to the $\sigma - \mu$ Pareto-Koopmans efficiency corresponding to the k^{th} frontier.

Proposition 2. The $\sigma - \mu$ Pareto-Koopmans efficiency respects the $\sigma - \mu$ Pareto dominance, that is, for all $i, i' \in I$ if $\mu_i \geqslant \mu_{i'}$ and $\sigma_i \leqslant \sigma_{i'}$, then $\delta_{ik} \geqslant \delta_{i'k}$ for any $k = 1, \ldots, p$.

Proof. As $\mu_i \geqslant \mu_{i'}$ and $\sigma_i \leqslant \sigma_{i'}$, $\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i'} - \beta\sigma_{i'}$ for all $\alpha, \beta \geqslant 0$ with $\alpha + \beta = 1$. Consequently,

$$\alpha\mu_{i'} - \beta\sigma_{i'} \geqslant \alpha\mu_{i''} - \beta\sigma_{i''} + \delta$$

implies

$$\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i''} - \beta\sigma_{i''} + \delta$$

for any $i'' \in I$ and any $\delta \in \mathbf{R}$. Therefore

$$\alpha\mu_{i'} - \beta\sigma_{i'} \geqslant \alpha\mu_{i''} - \beta\sigma_{i''} + \delta_{i'k}, \ \forall i'' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h$$

implies

$$\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i''} - \beta\sigma_{i''} + \delta_{i'k}, \ \forall i'' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h.$$

Consequently, since δ_{ik} is the maximum δ satisfying

$$\alpha\mu_i - \beta\sigma_i \geqslant \alpha\mu_{i''} - \beta\sigma_{i''} + \delta, \; \forall i'' \in I \setminus \bigcup_{h=1}^{k-1} PKF_h,$$

we have to conclude that $\delta_{ik} \geqslant \delta_{i'k}$. $\square$

To all units $i \in I$, we can assign an overall $\sigma - \mu$ Pareto-Koopmans efficiency score, denoted by sm_i, that reflects its efficiency with respect to all frontiers from **PKF**, as follows:

$$sm_i = \sum_{k=1}^{p} \delta_{ik}. \tag{8}$$

The following corollary of Proposition 2 ensures that overall $\sigma - \mu$ Pareto - Koopmans efficiency score sm_i respects the $\sigma - \mu$ Pareto dominance.

Proposition 3. For all $i, i' \in I$ if $\mu_i \geqslant \mu_{i'}$ and $\sigma_i \leqslant \sigma_{i'}$, then $sm_{ik} \geqslant sm_{i'k}$.

Proof. By Proposition 2: $\mu_i \geqslant \mu_{i'}$ and $\sigma_i \leqslant \sigma_{i'}$ implies $\delta_{ik} \geqslant \delta_{i'k}$ for all $k = 1, \dots, p$. Consequently, we have

$$sm_i = \sum_{k=1}^{p} \delta_{ik} \geqslant \sum_{k=1}^{p} \delta_{i'k} = sm_{i'}.$$

$\square$

In the following we supply some remarks related to the application of our approach in real life problems. As usual for the other indices of SMAA, the integrals defining the mean value μ_i and the standard deviation σ_i, $i \in I$, can be approximated by using a random sampling of q vectors of weights - with q being a relatively large number; for instance, following the suggestions of Tervonen and Lahdelma (2007), q could equal $10,000$-. The q random extracted weight vectors $\mathbf{w_h} = [w_{1h}, \dots, w_{mh}], h = 1, \dots, q$ can be collected in the following $m \times q$ **RW** matrix:

$$\underset{m\times q}{\mathbf{RW}} = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1q} \\ w_{21} & w_{22} & \cdots & w_{2q} \\ \vdots & \vdots & \dots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mq} \end{pmatrix}$$

Using the weight vector matrix **RW**, a composite index $CI(\mathbf{x}_i, \mathbf{w}_h)$ can be computed for each unit $i \in I$ and each weight vector $\mathbf{w}_h$, and the obtained results can be ordered in the following $n \times q$ matrix **CI** shown below:

$$\underset{n\times q}{\mathbf{CI}} = \begin{pmatrix} CI(\mathbf{x}_1, \mathbf{w}_1) & CI(\mathbf{x}_1, \mathbf{w}_2) & \cdots & CI(\mathbf{x}_1, \mathbf{w}_q) \\ CI(\mathbf{x}_2, \mathbf{w}_1) & CI(\mathbf{x}_2, \mathbf{w}_2) & \dots & CI(\mathbf{x}_2, \mathbf{w}_q) \\ \vdots & \vdots & \dots & \vdots \\ CI(\mathbf{x}_n, \mathbf{w}_1) & CI(\mathbf{x}_n, \mathbf{w}_2) & \cdots & C(\mathbf{x}_n, \mathbf{w}_q) \end{pmatrix}$$

Using the values collected in **CI**, for each unit $i \in I$ one can compute the approximated values $\widetilde{\mu}_i$ and $\widetilde{\sigma}_i$ for the mean μ_i and the standard deviation σ_i as follows:

$$\widetilde{\mu}_i = \frac{1}{q}\sum_{h=1}^{q} CI(\mathbf{x}_i, \mathbf{w}_h)$$

$$\widetilde{\sigma}_i = \sqrt{\frac{1}{q}\sum_{h=1}^{q}(CI(\mathbf{x}_i, \mathbf{w}_h) - \widetilde{\mu}_i)^2}$$

4 The σ-μ efficiency analysis step by step: A didactic example

The present section illustrates the application of $\sigma-\mu$ efficiency analysis with a concise didactic example. We consider a sample of the dataset supplied by the 2017 World Happiness Report (WHR) (Helliwell et al., 2017) that will be analyzed in its entirety as a case study in the following section. The WHR provides an evaluation of life satisfaction in more than 150 countries, based on citizens' responses to a Gallup World Poll survey. The report further supplies data on six key variables, analysing their relation with life satisfaction. For this didactic example, we take into consideration a sub-set of ten European countries (namely, *Austria, Denmark, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland and United Kingdom*) for the latest available year (data regarding the year 2016) to be evaluated through $\sigma-\mu$ efficiency analysis. For the sake of simplicity, we only consider three of the six key variables, and more precisely, *GDP per capita, Social support* and *Perceptions of corruption*. We report these in Table 1.

We start by normalizing the raw data reported in Table 1 following the methodology proposed in Greco et al. (2017a) that we recall in the following. Let us denote by y_{ij}, $i \in I, j \in J$ the raw value assumed for unit i with respect to dimension j. For each dimension $j \in J$, the mean value M_j and the standard deviation s_j can be computed as follows:

$$M_j = \frac{\sum_{i=1}^{n} y_{ij}}{n},$$

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}(y_{ij} - M_j)^2}{n}}.$$

Using the mean M_j and the standard deviation s_j we obtain the z-score

$$z_{ij} = \frac{y_{ij} - M_j}{s_j}$$

for each $i \in I$ and $j \in J$. Finally, we compute the normalized values x_{ij} as follows:

$$x_{ij} = \begin{cases} 0, & \text{if } y_{ij} \leqslant M_j - 3s_j \\ 0.5 + \dfrac{z_{ij}}{6}, & \text{if } M_j - 3s_j < y_{ij} < M_j + 3s_j \\ 1, & \text{if } y_{ij} \geqslant M_j + 3s_j \end{cases}$$

The normalization is applicable to positively-oriented dimensions, that is, dimensions for which the greater the raw value the better (e.g. *GDP per capita* and *Social Support*). Instead, for negatively-oriented dimensions, for which the greater the raw value the worse for a unit's performance (e.g. *Perception of corruption*), the normalization is formulated as follows:

Table 1: Raw and normalized values of the considered dimensions.

Raw Data				Normalized values			
Country	**Log of GDP per capita**	**Social support**	**Perceptions of corruption**	**Country**	**Log of GDP per capita**	**Social support**	**Corruption free**
Austria	10.69	0.93	0.52	Austria	0.48	0.49	0.44
Denmark	10.68	0.95	0.21	Denmark	0.47	0.70	0.71
France	10.54	0.88	0.62	France	0.33	0.18	0.35
Germany	10.70	0.91	0.45	Germany	0.49	0.34	0.51
Italy	10.43	0.93	0.90	Italy	0.23	0.50	0.11
Netherlands	10.76	0.93	0.43	Netherlands	0.54	0.49	0.52
Norway	11.07	0.96	0.41	Norway	0.84	0.74	0.54
Sweden	10.74	0.91	0.25	Sweden	0.53	0.38	0.68
Switzerland	10.92	0.93	0.30	Switzerland	0.70	0.50	0.63
United Kingdom	10.57	0.95	0.46	United Kingdom	0.37	0.70	0.50
Average	10.71	0.93	0.46				
Standard Deviation	0.17	0.02	0.19				

Data: 2017 World Happiness Report (WHR), obtained from: http://worldhappiness.report/ed/2017/. The data regard the year 2016. The detailed description and the sources of the considered dimensions can be found in Helliwell et al. (2017, p.17).

$$x_{ij} = \begin{cases} 0, & \text{if } y_{ij} \geqslant M_j + 3s_j \\ 0.5 - \dfrac{z_{ij}}{6}, & \text{if } M_j - 3s_j < y_{ij} < M_j + 3s_j \\ 1, & \text{if } y_{ij} \leqslant M_j - 3s_j \end{cases}$$

With respect to the creation of the weight vector matrix **RW**, in this didactic example we consider the following two scenarios, where $w_{GDP}, w_{Soc}, w_{Corr}$ denote weights for GDP per capita, Social support and Perception of corruption respectively:

- Scenario 1: No definite ranking importance of the three considered dimensions, so that the set of feasible weight vectors is

$$\mathbf{W} = \{[w_{GDP}, w_{Soc}, w_{Corr}] : w_{GDP} \geqslant 0, w_{Soc} \geqslant 0, w_{Corr} \geqslant 0, w_{GDP} + w_{Soc} + w_{Corr} = 1\};$$

- Scenario 2: Social support is more important than Perception of corruption that in turn is more important than GDP per capita, so that the set of feasible weight vectors is

$$\mathbf{W} = \{[w_{GDP}, w_{Soc}, w_{Corr}] : w_{Soc} \geqslant w_{Corr} \geqslant w_{GDP} \geqslant 0, w_{GDP} + w_{Soc} + w_{Corr} = 1\}.$$

For both scenarios, a set of 10,000 weight vectors $\mathbf{w}_h$, $h = 1, \ldots, 10,000$, was randomly sampled from a uniform distribution on the feasible set of weight vectors **W** and collected in the matrix $\mathbf{RW} = [w_{jh}, j = 1, 2, 3, h = 1, \ldots, 10,000]$. The weight vectors from **RW** and the normalized values x_{ij}, $i = 1, \ldots, 10$, $j = 1, 2, 3$, are then used to compute the composite indices

$$CI(\mathbf{x}_i, \mathbf{w}_h) = w_{GDP} x_{i,GDP} + w_{Soc} x_{i,Soc} + w_{Corr} x_{i,Corr},$$

$h = 1, \ldots, 10,000$.

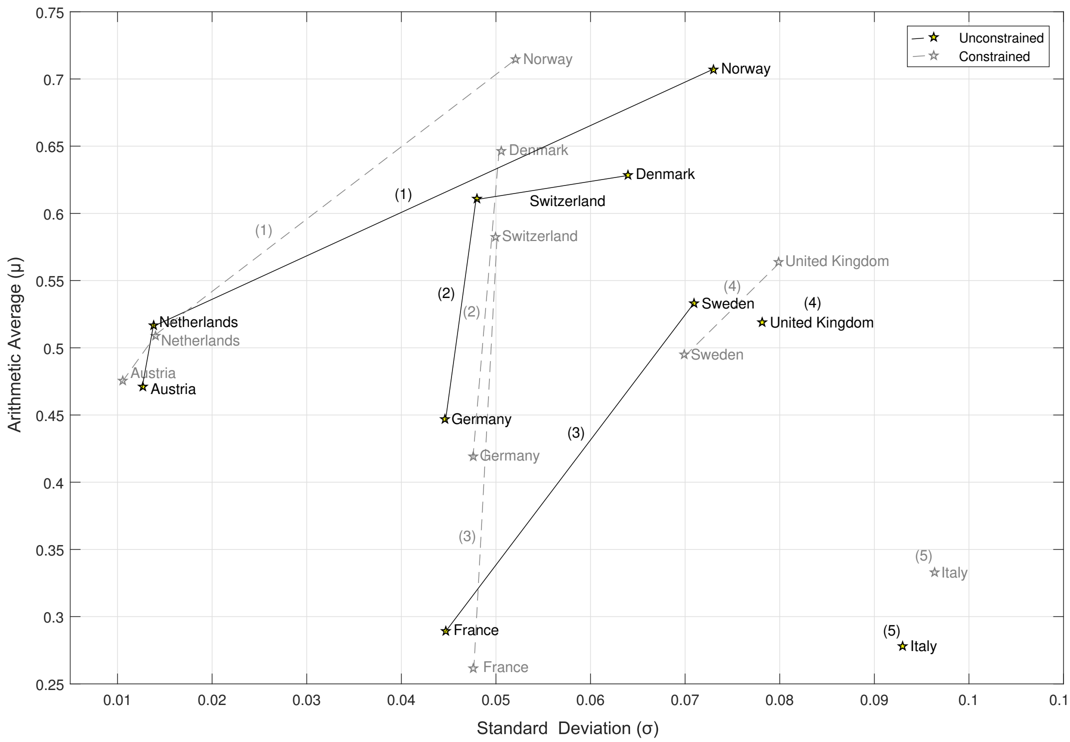
Using the values $CI(\mathbf{x}_i, \mathbf{w}_h)$, $i = 1, \ldots, 10, h = 1, \ldots, 10,000$, the approximation of the mean value $\widetilde{\mu}_i$ and the standard deviation $\widetilde{\sigma}_i$ of composite indices were calculated for each considered country. For the sake of simplicity, we refer to them as μ_i and σ_i, respectively. These two measures are reported in Table 2 and plotted in Figure 2 for both scenarios considered. Figure 2 also delineates the $\sigma - \mu$ Pareto-Koopmans efficiency frontiers; five were found in each scenario.

The $\sigma - \mu$ Pareto-Koopmans efficiency δ_{ik} of the considered countries with respect to the different $\sigma - \mu$ Pareto-Koopmans efficiency frontiers is given in Table 3. In both scenarios examined, the $\sigma - \mu$ Pareto-Koopmans family of frontiers consists of five frontiers. For the first scenario, that without a definite ranking of importance for the considered dimensions, the five frontiers are the following: PKF_1 = {Norway, Netherlands, Austria} , PKF_2 = {Denmark, Switzerland, Germany}, PKF_3 = {Sweden, France}, PFK_4 = {United Kindom}, PKF_5 = {Italy}. In the second scenario, the $\sigma - \mu$ Pareto-Koopmans frontiers are the same with the exceptions of Switzerland, that was in the second $\sigma - \mu$ Pareto-Koopmans efficiency frontier in the first scenario but descended to the third frontier in this one. Similarly, Sweden, which was in the third $\sigma - \mu$ Pareto-Koopmans efficiency frontier in the first scenario has been now descended to the fourth one.

In terms of overall efficiency, Norway presents the highest overall $\sigma - \mu$ Pareto-Koopmans efficiency score sm_i, while the second highest value is attributed to Denmark in both scenarios. It is worthwhile to observe that Denmark is not in the first $\sigma - \mu$ Pareto-Koopmans efficiency frontier, which, instead, is the case for Netherlands and Austria. Therefore, we can say that even if Denmark is in a worse Pareto-Koopmans efficiency frontier with respect to Netherlands and Austria, overall it compares better with

Figure 2: Illustrative example of the $\sigma - \mu$ plane in the two scenarios considered

Black colour represents $\sigma - \mu$ efficiency analysis output in the unconstrained case (i.e. scenario 1), while grey colour represents $\sigma - \mu$ efficiency analysis output in the constrained case (i.e. scenario 2). Numbers in parentheses denote respective $\sigma - \mu$ Pareto-Koopmans efficiency frontier (PKF_i).



respect to the whole set of efficiency frontiers (as shown by the overall efficiency score, sm_i). Let us also observe that in both scenarios Italy is the only country for which the efficiency score, sm_i, is negative. On the other hand, Italy is also the only country in the worst efficiency frontier.

Table 2: Evaluating the units with $\sigma - \mu$ under the two alternative scenarios

	Scenario 1 Unconstrained weights			**Scenario 2** Constrained weights		
Country	$\boldsymbol{\mu_i}$	$\boldsymbol{\sigma_i}$	$\mathbf{sm_i}$	$\boldsymbol{\mu_i}$	$\boldsymbol{\sigma_i}$	$\mathbf{sm_i}$
Austria	0.471	0.013	0.338	0.475	0.011	0.281
Denmark	0.628	0.064	0.561	0.646	0.051	0.514
France	0.289	0.045	0.076	0.262	0.048	0.037
Germany	0.447	0.045	0.188	0.419	0.048	0.074
Italy	0.278	0.093	-0.188	0.333	0.096	-0.209
Netherlands	0.517	0.014	0.393	0.509	0.014	0.303
Norway	0.707	0.073	0.948	0.715	0.052	0.802
Sweden	0.533	0.071	0.219	0.495	0.070	0.081
Switzerland	0.611	0.048	0.512	0.582	0.050	0.287
United Kingdom	0.519	0.078	0.394	0.564	0.080	0.204

$\boldsymbol{\mu_i}$ and $\boldsymbol{\sigma_i}$ are the means and standard deviations of the composite index $CI(\mathbf{x}_i, \mathbf{w})$ in the $10,000$ extractions accordingly. $\mathbf{sm_i}$ is the overall score computed as in eq.8.

Table 3: Measuring $\sigma - \mu$ Pareto-Koopmans efficiency

Unconstrained weights						**Constrained weights**					
	$\sigma - \mu$ Pareto-Koopmans efficiency						$\sigma - \mu$ Pareto-Koopmans efficiency				
	PKF1	**PKF2**	**PKF3**	**PKF4**	**PKF5**		**PKF1**	**PKF2**	**PKF3**	**PKF4**	**PKF5**
Country	δ_{i1}	δ_{i2}	δ_{i3}	δ_{i4}	δ_{i5}	**Country**	δ_{i1}	δ_{i2}	δ_{i3}	δ_{i4}	δ_{i5}
Austria	0.001	0.032	0.047	0.065	0.193	Austria	0.003	0.037	0.038	0.059	0.143
Denmark	-0.012	0.018	0.095	0.110	0.350	Denmark	-0.009	0.064	0.064	0.083	0.313
France	-0.032	0.000	0.026	0.033	0.048	France	-0.037	0.000	0.002	0.022	0.049
Germany	-0.032	0.002	0.015	0.034	0.169	Germany	-0.037	0.001	0.001	0.022	0.086
Italy	-0.080	-0.048	-0.045	-0.015	0.000	Italy	-0.086	-0.049	-0.048	-0.026	0.000
Netherlands	0.008	0.032	0.050	0.064	0.239	Netherlands	0.002	0.034	0.035	0.056	0.176
Norway	0.078	0.078	0.174	0.188	0.429	Norway	0.068	0.068	0.132	0.151	0.382
Sweden	-0.040	-0.024	0.014	0.014	0.255	Sweden	-0.049	-0.021	-0.020	0.010	0.162
Switzerland	-0.004	0.013	0.078	0.092	0.333	Switzerland	-0.019	0.000	0.028	0.028	0.249
United Kingdom	-0.049	-0.031	-0.008	0.241	0.241	United Kingdom	-0.047	-0.030	-0.019	0.069	0.231

PKF1-5 denote respective $\sigma - \mu$ Pareto-Koopmans frontiers illustrated in Figure 2. $\delta_{\mathbf{ik}}$ shows the (in)efficiency of Country i, with respect to the k^{th} frontier.

5 Case Study: World Happiness Index

In this section, we apply $\sigma - \mu$ efficiency analysis to the whole set of data supplied by the 2017 Report of 'World Happiness'. On a general note, happiness is an age-old concept that can be traced back to Aristotle's 'eudaimonia', a word commonly translated as 'welfare' (Shin and Johnson, 1978). Central concept of the Aristotelian ethics, welfare was seen as the ultimate human good (Robinson, 1989), which, more than two millennia after Aristotle's era appears to be at the centre of academics and policy-makers' discussions. More specifically, world-renowned economists have recently criticized the use of traditional, economic output measures like the GDP as a proxy for welfare (see e.g. Costanza et al., 2009; Stiglitz et al., 2009). In April 2012, an initiative of a group of independent experts -in support of the United Nations' High Level Meeting on happiness and well-being- further paved this way. Through the Sustainable Development Solutions Network of the UN, they published the first 'World Happiness Report' (Helliwell et al., 2012). Since 2012, these reports have gained considerable attention, while, in the authors' words (Helliwell et al., 2017, p.3): "*happiness is now increasingly considered the proper measure of social progress and the goal of public policy*". In fact, on a recent OECD meeting at the ministerial level (OECD, 2016, p.12), the OECD committed to "*redefine the growth narrative to put people's well-being at the center of governments' efforts*".

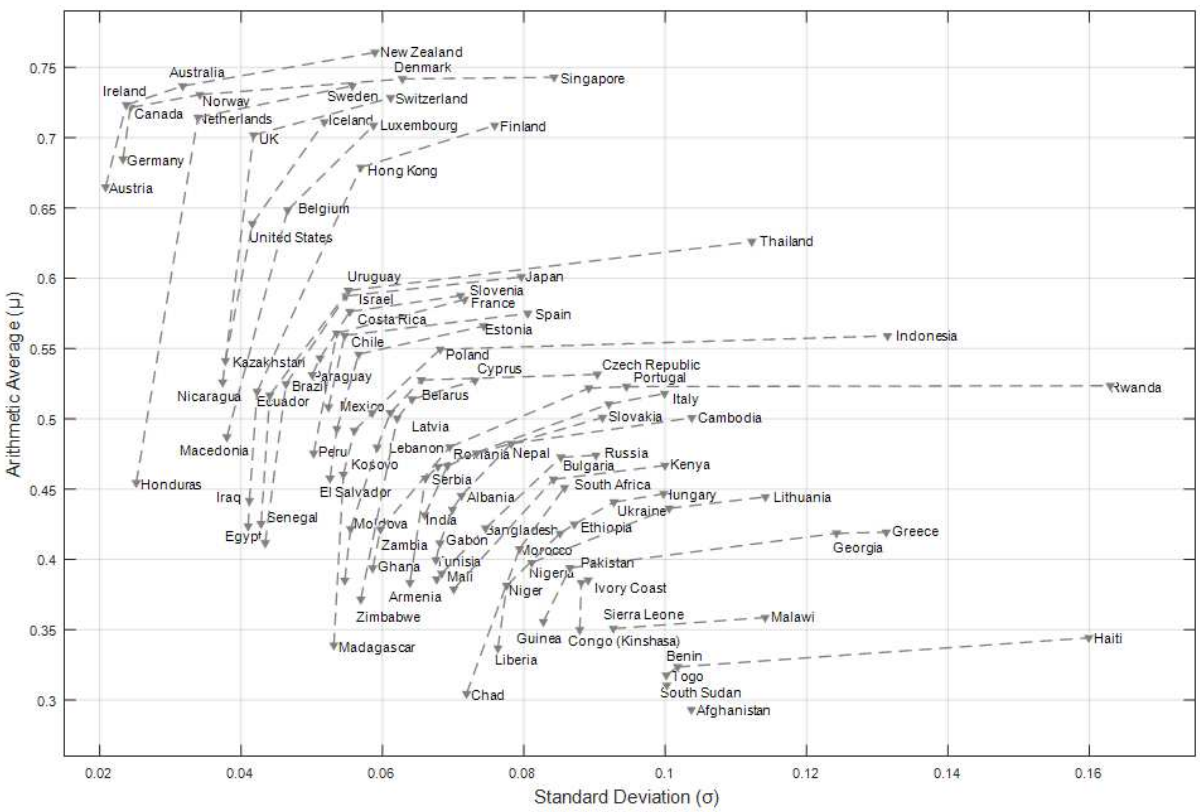
The 'World Happiness' report presents and analyses the data of a survey question conducted by the Gallup World Poll. More specifically, 3,000 respondents in each of the -roughly- 150 countries considered, evaluate their lives on a 0-10 scale which is known as 'Cantril Ladder' (see Helliwell et al., 2017, p.123). The authors use a three-year rolling window of the average response in each country to rank them accordingly. For instance, the 2016 ranking is based on the average response of the three-year period 2014-2016. We shall refer to the results of the survey as Subjective Well Being (SWB). According to the report, 6 key variables (namely *GDP per capita, healthy life expectancy at birth, social support, freedom to make life choices, generosity* and *perceptions of corruption*) used as proxies for 6 socio-economic aspects respectively, may on average explain 75% of the respondents' subjective evaluations (Pooled OLS regression). Detailed information about the description and sources of the 6 key variables can be found in Helliwell et al. (2017, Technical Box 2, p.17). We applied $\sigma - \mu$ efficiency analysis adopting the same procedure extensively described in the previous section, which considered a sub-sample of 10 European countries, apart from the following step. We use a three-year rolling-window for the six variables, in order to be consistent with the procedure used by the World Happiness Report for the subjective evaluation. This means that the values we consider in each dimension in year 2016 are in fact non-weighted arithmetic averages of the period 2014-2016. We restrict the sample to only these countries that possess data for all 6 dimensions for the 2016 and at least one of the years 2014 and 2015. After this cleansing procedure we are left with a final sample of 119 countries.

In applying the proposed approach, we find that the family of $\sigma - \mu$ Pareto-Koopmans frontiers consists of 31 frontiers, which are illustrated in Figure 3. We computed the $\sigma - \mu$ Pareto-Koopmans efficiency δ_{ik} with respect to all 31 frontiers for each country. However, due to a large number of countries and frontiers in our sample, we will hereby discuss and report only the efficiency of the top-10 ranked countries of the 2017 'World Happiness' report. The results for the rest of the countries (e.g. efficiencies, overall scores and rankings) are disclosed in the on-line supplementary appendix (available here: `https://goo.gl/URBRuC`). According to the 2017 report, the countries found in the top ten rankings are the following: *Norway, Denmark, Iceland, Switzerland, Finland, Netherlands, Canada, New Zealand, Australia and Sweden*, which are ranked in this exact order. In our analysis, these 10 countries are found to be spread in the first seven frontiers, which will therefore be the focus of our analysis for the rest of this section.

The countries spread over the first seven frontiers are reported in Table 4, ordered according to the

Figure 3: Family of $\sigma - \mu$ Pareto-Koopmans frontiers.

The 119 countries in our sample are spread over 31 $\sigma - \mu$ Pareto-Koopmans efficiency frontiers (PKF). Further details about the coordinates, efficiency with respect to each PKF, overall $\sigma - \mu$ efficiency and rankings of each country are given in the on-line supplementary appendix.



rankings attributed to them in the WHR (denoted by 'WHR rank' respectively). Also reported in the table are the mean score (μ_i) and the standard deviation (σ_i) of the countries' scores in the $10,000$ extractions, the $\sigma - \mu$ Pareto-Koopmans efficiency of each country with respect to the efficiency frontier $PKF_k, k = 1, \ldots, 7$, δ_{ik}, and the overall efficiency score sm_i with its corresponding and ranking (denoted by '$\sigma - \mu$ rank').

First of all, we should note that it is by definition reasonable to observe a shuffle, or even entirely different patterns between the SWB ('WHR rank') and the $\sigma - \mu$ efficiency rankings ('$\sigma - \mu$ rank'). The first expresses peoples' own subjective beliefs, while the latter refers to the aggregation of 6 variables that are considered to explain SWB well on average. Moreover, there is a whole ongoing discussion between the difference of SWB and objective conditions attributed to psychological reasons and cultural differences (see Kroll and Delhey, 2013). In other words, the two rankings are not directly comparable, nor should they necessarily be; though one could make a few interesting inferences. To start with, it is notable, that the countries which are self-claimed to be ranked in the top-10 positions (i.e. having the top-10 highest subjective evaluation) are positioned in our top-10 list as well, with the exception of Iceland and Finland, which we position in the 11^{th} and 13^{th} places accordingly.

A second interesting point relates to the measurement of efficiency with respect to the frontiers, and how the dynamics of these might change under some circumstances. Consider for instance Finland, a country that is ranked 13^{th} according to our overall $\sigma - \mu$ Pareto-Koopmans efficiency, and which participates in the $\sigma - \mu$ Pareto-Koopmans family by lying on the 7^{th} frontier. The reason Finland is not participating in the previous frontier (i.e. PKF6) can be better clarified when it is compared to Luxembourg. The latter clearly dominates the former in terms of standard deviation ($\sigma_{Luxembourg} = 0.059$ versus $\sigma_{Finland} = 0.076$), but only marginally dominates in terms of average performance ($\mu_{Luxembourg} = 0.70865$ versus $\mu_{Finland} = 0.70864$ - in Table 4 both are rounded to three decimals). Therefore, if Finland slightly increases its average performance to surpass that of Luxembourg, it will then move to frontier 6 *ceteris paribus*. This is also clear by looking at the efficiency of Finland with respect to the 6^{th} frontier (Table 4: $\delta_{Finland,6} = -0.00001$), which is almost zero. Following this line of reasoning, one could be interested to compare Finland with Iceland ($\mu_{Iceland} = 0.7111$ versus $\mu_{Finland} = 0.70864$), e.g. by looking at the (in)efficiency of the former with respect to the frontier that the latter is lying on (Table 4: $\delta_{Finland,5} = -0.002$).

Another interesting point arises from tracking the frontiers' formation from a dynamic viewpoint. More specifically, one could be interested in tracing changes in the performance of units in the $\sigma - \mu$ plane within a time period and thus, how were the frontiers re-structured accordingly. This could be accomplished in several ways. For instance, one could trace all, or a subset of the $\sigma - \mu$ PKF, or even trace the frontiers and performance of only certain countries. An example is given in Figure 4, which illustrates how the first two frontiers were changed from 2015 (illustrated in gray) to the following year (illustrated in black). It quickly becomes obvious that Singapore did not participate in the first two frontiers in 2015, but it joined the second in 2016. Moreover, one can distinguish how the performance of the countries lying in the first two $\sigma - \mu$ PKF changed during this time period. For instance, almost all countries exhibit a drop from 2015 to the following year. In few countries this is less and in others more noticeable. Exception in this rule are Germany, Luxembourg and Singapore, with the latter meeting with such an improvement that positioned the country in the second frontier. Of course this can be attributed to both a remarkable improvement in the elementary indicators, and the fact that the performance of the surrounding countries was deteriorated (e.g. see Denmark in Fig. 4). This highlights the fact that even if a unit's performance remains steady through a time period examined, the distance with respect to other frontiers might alter either due to an improvement, or a downturn of the surrounding units. In this particular example, from a policy-maker's perspective, two consecutive years might not be enough; thus, the time period examined in the plane could be re-considered to that of specific 'goalposts' (i.e. the start and end dates of a scheduled policy period, see Mazziotta and Pareto, 2016, p.989). Nonetheless,

Table 4: Case study results for the first seven frontiers.

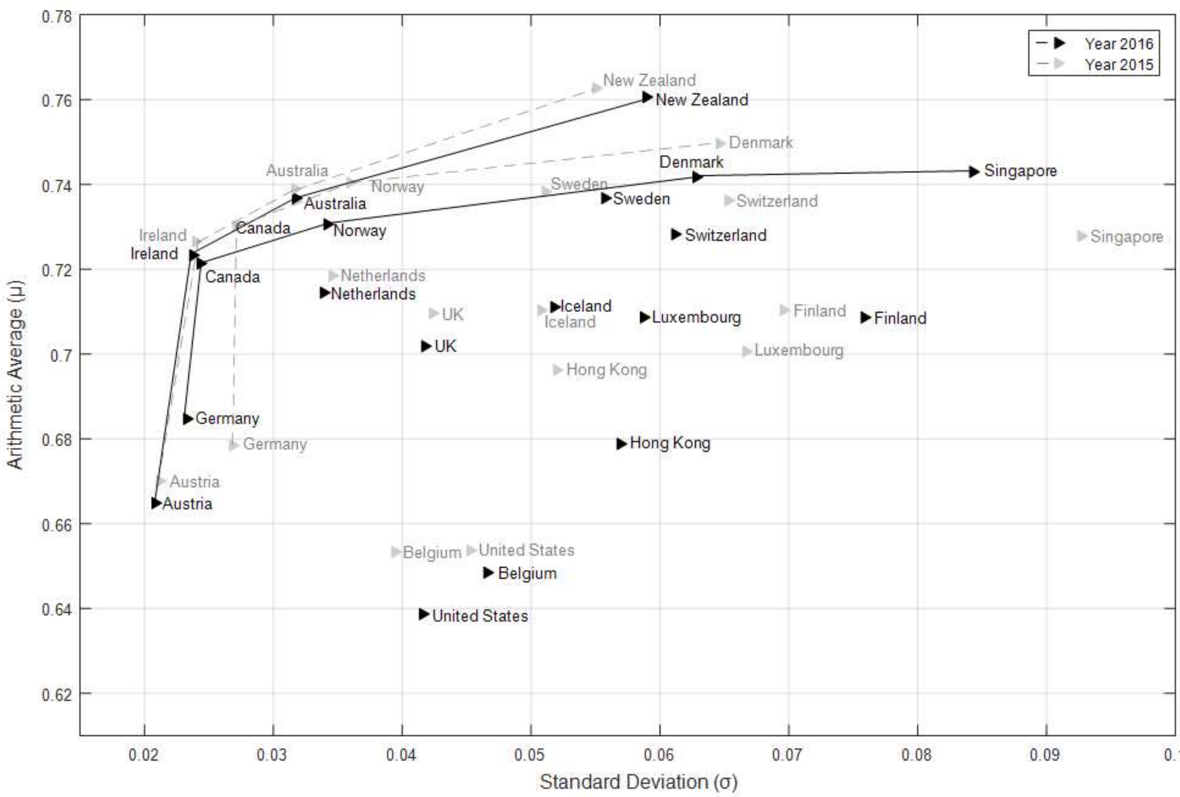
Country	WHR rank	μ_i	σ_i	$\mathrm{sm_i}$	$\sigma-\mu$ rank	$\sigma-\mu$ Pareto-Koopmans efficiency PKF1 δ_{i1}	PKF2 δ_{i2}	PKF3 δ_{i3}	PKF4 δ_{i4}	PKF5 δ_{i5}	PKF6 δ_{i6}	PKF7 δ_{i7}
Norway	1	0.731	0.034	6.040	6	-0.004	0.003	0.008	0.017	0.020	0.024	0.034
Denmark	2	0.742	0.063	6.312	3	-0.012	0.003	0.005	0.013	0.031	0.033	0.033
Iceland	3	0.711	0.052	5.445	11	-0.022	-0.019	-0.011	-0.002	0.006	0.006	0.016
Switzerland	4	0.728	0.061	5.922	7	-0.018	-0.009	-0.007	0.017	0.017	0.020	0.020
Finland	5	0.709	0.076	5.335	13	-0.036	-0.027	-0.024	-0.017	-0.002	-0.000	0.030
Netherlands	6	0.714	0.034	5.619	10	-0.010	-0.008	0.009	0.010	0.016	0.022	0.028
Canada	7	0.721	0.024	5.843	9	-0.001	0.006	0.009	0.018	0.025	0.031	0.036
New Zealand	8	0.761	0.059	6.904	1	0.018	0.018	0.024	0.032	0.050	0.052	0.052
Australia	9	0.737	0.032	6.218	4	0.002	0.005	0.012	0.021	0.026	0.028	0.038
Sweden	10	0.737	0.056	6.173	5	-0.011	-0.002	0.009	0.009	0.026	0.028	0.028
Austria	13	0.665	0.021	4.496	17	0.002	0.002	0.011	0.020	0.021	0.025	0.032
United States	14	0.639	0.042	3.726	19	-0.021	-0.018	-0.010	-0.001	0.004	0.004	0.011
Ireland	15	0.723	0.024	5.891	8	0.001	0.001	0.010	0.019	0.026	0.032	0.038
Germany	16	0.685	0.023	4.955	15	-0.001	0.001	0.009	0.018	0.022	0.025	0.031
Belgium	17	0.648	0.047	3.925	18	-0.025	-0.023	-0.014	-0.006	-0.003	0.006	0.007
Luxembourg	18	0.709	0.059	5.358	12	-0.027	-0.023	-0.015	-0.007	-0.002	0.010	0.010
United Kingdom	19	0.702	0.042	5.252	14	-0.018	-0.017	-0.008	0.008	0.008	0.013	0.018
Singapore	26	0.743	0.084	6.341	2	-0.018	0.001	0.006	0.015	0.032	0.034	0.034
Nicaragua	41	0.526	0.037	1.668	33	-0.017	-0.014	-0.010	0.000	0.000	0.003	0.005
Ecuador	44	0.519	0.042	1.496	38	-0.021	-0.019	-0.014	-0.005	-0.004	-0.002	0.002
Kazakhstan	60	0.541	0.038	1.871	30	-0.017	-0.014	-0.009	0.000	0.001	0.003	0.006
Hong Kong	71	0.679	0.057	4.592	16	-0.034	-0.033	-0.023	-0.015	-0.008	-0.004	0.012
Honduras	91	0.455	0.025	1.359	40	-0.004	-0.002	0.009	0.012	0.013	0.013	0.016
Macedonia	92	0.487	0.038	1.272	41	-0.017	-0.015	-0.011	-0.001	0.000	0.004	0.004
Egypt	111	0.424	0.041	0.786	55	-0.020	-0.018	-0.016	-0.004	-0.003	-0.003	0.000
Iraq	117	0.442	0.041	0.876	54	-0.020	-0.018	-0.016	-0.004	-0.003	-0.003	0.000

WHR is the rank attributed to Country i by the 'World Happiness' report using the Gallup World Poll surveys (i.e. 'Cantril Ladder'). $\boldsymbol{\mu_i}$ and $\boldsymbol{\sigma_i}$ are the means and standard deviations of the composite index $CI(\mathbf{x}_i, \mathbf{w})$ in the 10,000 extractions accordingly. $\mathbf{sm_i}$ is the overall score computed as in eq.8. $\boldsymbol{\sigma}$-$\boldsymbol{\mu}$ **rank** is the rank obtained based on the overall score **sm**. **PKF1-7** denote respective frontiers and δ_{ik} exhibits the (in)efficiency of Country i, with respect to the k^{th} $\sigma-\mu$ Pareto-Koopmans frontier.

our approach would eventually allow for the analysis of the impact of such a policy in comparative terms, with respect to similar units of analysis not involved in the policy programme.

Figure 4: Dynamic illustration of the frontiers.

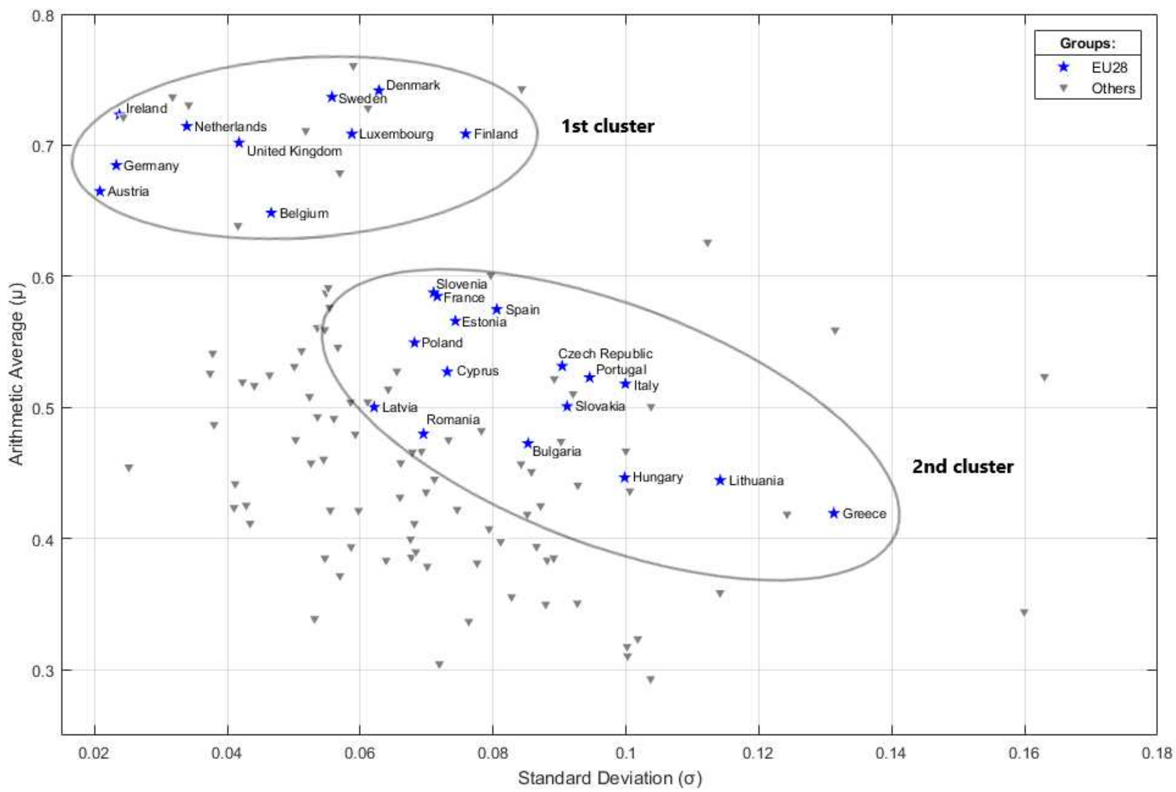
An interesting feature of $\sigma - \mu$ analysis is the comparison of units or frontiers from a dynamic viewpoint. A developer might be keen on tracking the formation of a frontier of interest, or the performance of a unit through time (e.g. either consecutive years, or a policy period of interest). This figure delineates the formation of the first two $\sigma - \mu$ Pareto-Koopmans efficiency frontiers (PKF) in two consecutive years. Black colour represents the year 2016 while grey colour represents the year 2015.



Last, but not least, interesting inferences could be made by focusing on specific clusters of units of interest within the $\sigma - \mu$ plane. For instance, a policy-maker could be interested in observing how a specific country performs in comparison to a manually-chosen group of countries. In Figure 5 we have chosen to illustrate how the EU-28 countries (with the exception of Malta and Croatia due to missing data for the period examined) perform both among them and in comparison to the rest of the countries considered in our sample. This type of manual grouping into specific clusters of units has highlighted some further structural differences among them, with EU-28 countries positioned in the north and north-west $\sigma - \mu$ Pareto-dominating those to the south and south-east. This is validated even according to the WHR rankings (unreported) that reflect the citizens' own beliefs, in which the highest-ranked EU-28 country belonging to the second cluster of countries in Figure 5 is France (ranked 31^{st} according to the report) (see Helliwell et al., 2017, figure 2.2, p.20). Obviously, the discussion about the structural differences inherent in these countries and their determinants goes beyond the scope of this study. However, we can argue that the $\sigma - \mu$ plane can provide the decision-maker with some enriched implications. This can be seen as a considerable asset of our proposed method, which illustrates alternative comparisons among units of interest. More specifically, one could be interested in benchmarking countries within their own specified cluster, entirely neglecting the rest of the units. For instance, countries belonging to the second cluster in Figure 5 could be benchmarked against each other, instead of a more holistic analysis that involves all 119 countries considered in our sample.

Figure 5: Leaders and Laggards: cluster-spotting in the $\sigma-\mu$ plane.

This figure delineates the manual grouping of the 119 countries in our sample into 'EU28' (symbolized with pentagram; Malta & Croatia missing due to data unavailability) and 'Others' (symbolized with reversed triangles). It is visually clear that the EU28 group of countries is partitioned into two clusters. One could be interested in comparing a group of countries (e.g those belonging to a predetermined cluster) with their counterparts within this group, rather than conducting a more holistic analysis.



Consequently, there are several points that could be noted from the outputs of our proposed approach. From an overall score/ranking that takes into account all potential viewpoints (i.e. space of weight vectors) and all potential benchmarks (as denoted by the family of $\sigma-\mu$ Pareto-Koopmans frontiers), to the analysis of the dynamic, or spatial performance of an unit. These could be all advantageous to both the developer of an index and the individuals interested in it. Due to a high number of countries within our sample we have limited the discussion of the results to only those countries that made the top-10 list in the 2017 World Happiness Report. For the reader interested in the remaining results, we report these in the on-line supplementary appendix (available here: `https://goo.gl/URBRuC`). We should hereby note again that subjective evaluations (i.e. those of the WHR in this case) and our own output (i.e. sm_i overall efficiency score and $\sigma-\mu$ rankings accordingly) cannot be directly compared due to the intrinsic differences in their representation.

6 Conclusion

We proposed a novel methodology called $\sigma-\mu$ efficiency analysis to deal with the issue of weighting in the construction of a composite index. In fact, quite different results can be obtained by changing the weights of the dimensions considered by the composite index. Therefore, it seems reasonable to take into account for each unit the distribution of values assumed by the composite index on the whole set of feasible weight vectors. We synthesize such distributions for each unit with its mean value μ, intended to be maximized, and its standard deviation σ, intended to be minimized, as it denotes instability in the evaluations with respect to the variability of weights. We further defined the concepts of $\sigma - \mu$ Pareto-Koopmans dominance and efficiency, which permitted us to partition the units under analysis in a sequence of efficiency frontiers and to define several types of meaningful efficiency measures. This way we outlined the $\sigma - \mu$ efficiency analysis which finds its basis in some well known Operational Research methodologies:

- Stochastic Multiattribute Acceptability Analysis (SMAA), for the idea of considering the whole set of feasible weight vectors;
- Data Envelpment Analysis (DEA), for the idea of measuring efficiency;
- Markowitz modern portfolio theory, for the idea of representing distributions in terms of mean and standard deviation.
- NSGA-II, for the idea of a sequence of Pareto frontiers.

With respect to its merits, the proposed method permits the inclusion of all potential viewpoints in the construction of a composite index, while it takes into account the distances of units from all the $\sigma-\mu$ Pareto-Koopmans frontiers lying on the plane. While there is no particular scope in this study to treat compensatory issues in the construction of an index; we should note that our methodology permits the use of non-compensatory aggregation techniques such as PROMETHEE methods (see Brans et al., 1986) or ELECTREE methods (for a survey see Figueira et al., 2016 and for a review of recent developments see Figueira et al., 2013) to be applied instead of the additive utility model illustrated in the paper. In this case, to apply the SMAA to PROMETHEE and ELECTRE methods, see the approaches proposed in Corrente et al. (2014) and Corrente et al. (2016a) respectively.

We attempted to show the potential of $\sigma-\mu$ efficiency analysis by applying it to the data supplied by the 'World Happiness' report, obtaining a few interesting results and insights. Of course, our methodology cannot be considered a 'panacea' for the many problems affecting the adoption of composite indices, in general, and the 'World Happiness' in particular (see e.g. the critical discussion on composite indices applied to wellbeing measures in Kroll and Delhey, 2013). However, we hope that this case study can convince on the many interesting analyses and insights that $\sigma - \mu$ efficiency analysis permits in this domain.

Finally, as far as its future direction of research is concerned, we believe that our methodology can be fruitfully applied to all the domains in which composite indices are considered, ranging from the ranking of universities to the measurement of competitiveness of geographical regions. Moreover, we believe that the idea of successive Pareto-Koopmans efficiency frontiers has clear implications in the domain of classic DEA, which we endeavour to explore in Greco et al. (2017b).

References

Abberger, K., Graff, M., Siliverstovs, B., and Sturm, J.-E. (2017). Using rule-based updating procedures to improve the performance of composite indicators. *Economic Modelling*, 68:127–144.

Aertens, W., Kint, V., Van Orshoven, J., and Muys, B. (2011). Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environmental Modelling & Software*, 26(7):929–937.

Ainslie, G. (2001). *Breakdown of will*. Cambridge, United Kingdom: Cambridge University Press.

Angilella, S., Corrente, S., and Greco, S. (2015). Stochastic multiobjective acceptability analysis for the choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, 240(1):172–182.

Bandura, R. (2011). *Composite Indicators and Rankings: Inventory 2011*. Technical report, New York: Office of Development Studies, United Nations Development Programme (UNDP).

Becker, W., Saisana, M., Paruolo, P., and Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, 80:12–22.

Bewley, T. F. (2002). Knightian decision theory. part i. *Decisions in Economics and Finance*, 25(2):79–110.

Blackburn, D. W. and Ukhov, A. D. (2013). Individual vs. aggregate preferences: The case of a small fish in a big pond. *Management Science*, 59(2):470–484.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, 59(2):115–151.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge, United Kingdom: Cambridge university press.

Brans, J.-P., Vincke, P., and Mareschal, B. (1986). How to select and how to rank projects: The promethee method. *European journal of operational research*, 24(2):228–238.

Charnes, A. and Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics (NRL)*, 9(3-4):181–186.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., and Stutz, J. (1985). Foundations of data envelopment analysis for pareto-koopmans efficient empirical production functions. *Journal of econometrics*, 30(1-2):91–107.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978a). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.

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

Chowdhury, S. and Squire, L. (2006). Setting weights for aggregate indices: An application to the commitment to development index and human development index. *The Journal of Development Studies*, 42(5):761–771.

Cooper, W. W., Seiford, L. M., and Zhu, J. (2011). *Handbook on data envelopment analysis*. International Series in Operations Research & Management Science. 2nd edition, Springer US.

Corrente, S., Figueira, J. R., and Greco, S. (2014). The smaa-promethee method. *European Journal of Operational Research*, 239(2):514–522.

Corrente, S., Figueira, J. R., Greco, S., and Słowiński, R. (2016a). A robust ranking method extending electre iii to hierarchy of interacting criteria, imprecise weights and stochastic analysis. *Omega*.

Corrente, S., Greco, S., Kadziński, M., and Słowiński, R. (2013). Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93(2-3):381–422.

Corrente, S., Greco, S., Kadziński, M., and Słowiński, R. (2016b). Inducing probability distributions on the set of value functions by subjective stochastic ordinal regression. *Knowledge-Based Systems*, 112:26–36.

Costanza, R., Hart, M., Posner, S., and Talberth, J. (2009). *Beyond GDP: The Need for New Measures of Progress*. Pardee Center for the Study of the Longer-Range Future, Boston.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197.

Decancq, K. and Lugo, M. A. (2013). Weights in Multidimensional Indices of Wellbeing: An Overview. *Econometric Reviews*, 32(1):7–34.

Decancq, K., Van Ootegem, L., and Verhofstadt, E. (2013). What if we voted on the weights of a multi-dimensional well-being index? an illustration with flemish data. *Fiscal Studies*, 34(3):315–332.

Doumpos, M., Gaganis, C., and Pasiouras, F. (2016). Bank Diversification and Overall Financial Strength: International Evidence. *Financial Markets, Institutions & Instruments*, 25(3):169–213.

Doumpos, M., Hasan, I., and Pasiouras, F. (2017). Bank overall financial strength: Islamic versus conventional banks. *Economic Modelling*, 64:513–523.

Durbach, I. (2009). On the estimation of a satisficing model of choice using stochastic multicriteria acceptability analysis. *Omega*, 37(3):497–509.

Elster, J. (1987). *The Multiple self*. Cambridge, United Kingdom: Cambridge University Press.

Elton, E. J., Gruber, M. J., Brown, S. J., and Goetzmann, W. N. (2009). *Modern portfolio theory and investment analysis*. John Wiley & Sons.

Figueira, J. R., Greco, S., Roy, B., and Slowinski, R. (2013). An Overview of ELECTRE Methods and their Recent Extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2):61–85.

Figueira, J. R., Mousseau, V., and Roy, B. (2016). ELECTRE methods. In Greco, S., Ehrgott, M., and Figueira, J., editors, *Multiple criteria decision analysis: State of the art surveys*, pages 155–185. Springer.

Gan, X., Fernandez, I. C., Guo, J., Wilson, M., Zhao, Y., Zhou, B., and Wu, J. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81:491–502.

Gilboa, I., Maccheroni, F., Marinacci, M., and Schmeidler, D. (2010). Objective and subjective rationality in a multiple prior model. *Econometrica*, 78(2):755–770.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, 18(2):141–153.

Greco, S., Ehrgott, M., and Figueira, J. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. International Series in Operations Research & Management Science. 2nd edition, New York: Springer.

Greco, S., Ishizaka, A., Matarazzo, B., and Torrisi, G. (2017a). Stochastic multi-attribute acceptability analysis (smaa): an application to the ranking of italian regions, *Regional Studies*, doi: 10.1080/00343404.2017.1347612, advance online publication.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2017b). Measuring efficiency with multiple frontier data envelopment analysis, *Mimeo*.

Greco, S., Ishizaka, A., Tasiou, M., and Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation and robustness, *Social Indicators Research*, doi: 10.1007/s11205-017-1832-9, advance online publication.

Greco, S., Mousseau, V., and Słowiński, R. (2008). Ordinal regression revisited: multiple criteria ranking using a set of additive value functions. *European Journal of Operational Research*, 191(2):416–436.

Greco, S., Słowiński, R., Figueira, J., and Mousseau, V. (2010). Robust ordinal regression. *Trends in multiple criteria decision analysis*, 142:241–283.

Grupp, H. and Schubert, T. (2010). Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, 39(1):67–78.

Hartley, J. E. and Hartley, J. E. (2002). *The representative agent in macroeconomics*. London, United Kingdom: Routledge.

Helliwell, J., Layard, R., and Sachs, J. (2012). *World Happiness Report 2012*. New York: Sustainable Development Solutions Network.

Helliwell, J., Layard, R., and Sachs, J. (2017). *World Happiness Report 2017*. New York: Sustainable Development Solutions Network.

Ishizaka, A. and Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. Chichester, United Kingdom: John Wiley & Sons.

Karagiannis, G. (2017). On aggregate composite indicators. *Journal of the Operational Research Society*, 68(7):741–746.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *The Journal of Economic Perspectives*, 6(2):117–136.

Kroll, C. and Delhey, J. (2013). A happy nation? opportunities and challenges of using subjective indicators in policymaking. *Social Indicators Research*, 114(1):13–28.

Lahdelma, R., Hokkanen, J., and Salminen, P. (1998). SMAA - Stochastic multiobjective acceptability analysis. *European Journal of Operational Research*, 106(1):137–143.

Lahdelma, R. and Salminen, P. (2001). SMAA-2 : Stochastic Multicriteria Acceptability Analysis for Group Decision Making. *Operations Research*, 49(3):444–454.

Lahdelma, R. and Salminen, P. (2009). Prospect theory and stochastic acceptability analysis (SMAA). *Omega*, 37(5):961–971.

Leskinen, P., Viitanen, J., Kangas, A., and Kangas, J. (2006). Alternatives to incorporate uncertainty and risk attitude in multicriteria evaluation of forest plans. *Forest Science*, 52(3):304–312.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

Mazziotta, M. and Pareto, A. (2016). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena. *Social Indicators Research*, 127(3):983–1003.

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, 306(5695):503–507.

Menou, A., Benallou, A., Lahdelma, R., and Salminen, P. (2010). Decision support for centralizing cargo at a Moroccan airport hub using stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 204(3):621–629.

Mikulić, J., Kožić, I., and Krešić, D. (2015). Weighting indicators of tourism sustainability: A critical note. *Ecological Indicators*, 48:312–314.

Munda, G. (2005a). "Measuring sustainability": A multi-criterion framework. *Environment, Development and Sustainability*, 7(1):117–134.

Munda, G. (2005b). Multiple Criteria Decision Analysis and Sustainable Development. In Greco, S., Ehrgott, M., and Figueira, J., editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 953–986.

OECD (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, Paris.

OECD (2016). Strategic Orientations of the Secretary-General for 2016 and beyond. Technical report, Meeting of the OECD Council at Ministerial Level. Retrieved from https://www.oecd.org/mcm/documents/strategic-orientations-of-the-secretary-general-2016.pdf.

Paruolo, P., Saisana, M., and Saltelli, A. (2013). Ratings and rankings: voodoo or science? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(3):609–634.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572.

Ray, A. K. (2008). Measurement of social development: an international comparison. *Social Indicators Research*, 86(1):1–46.

Robinson, D. N. (1989). *Aristotle's psychology.* New York: Columbia University Press.

Saaty, T. L. (1977). A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, 15(3):234–281.

Saaty, T. L. (1980). *The Analytic Hierarchy Process*. McGraw-Hill, New York.

Saisana, M., Saltelli, A., and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 168(2):307–323.

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, 81(1):65–77.

Schelling, T. C. (1980). The intimate contest for self-command. *The Public Interest*, (60):94.

Sharpe, A. (2004). *Literature Review of Frameworks for Macro-indicators*. Centre for the Study of Living Standards, Ottawa.

Shin, D. C. and Johnson, D. M. (1978). Avowed happiness as an overall assessment of the quality of life. *Social indicators research*, 5(1-4):475–492.

Spearman, C. (1904). "General Intelligence", Objectively Determined and Measured. *The American Journal of Psychology*, 15(2):201–292.

Stiglitz, J., Sen, A. K., and Fitoussi, J.-P. (2009). *The measurement of economic performance and social progress revisited: Reflections and Overview*. Commission on the Measurement of Economic Performance and Social Progress, Paris.

Tervonen, T., Figueira, J., Lahdelma, R., Almeida Dias, J., and Salminen, P. (2009a). A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research*, 192(1):236–242.

Tervonen, T., Figueira, J., Lahdelma, R., and Salminen, P. (2009b). SMAA-III: A simulation-based approach for sensitivity analysis of ELECTRE III. In *Real-Time and Deliberative Decision Making*, pages 241–253. Springer.

Tervonen, T. and Figueira, J. R. (2008). A survey on stochastic multicriteria acceptability analysis methods. *Journal of Multi-Criteria Decision Analysis*, 15(1-2):1–14.

Tervonen, T. and Lahdelma, R. (2007). Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research*, 178(2):500–513.

Tervonen, T., Linkov, I., Figueira, J. R., Steevens, J., Chappel, M., and Merad, M. (2009c). Risk-based classification system of nanomaterials. *Journal of Nanoparticle Research*, 11(4):757–766.

Van Puyenbroeck, T. and Rogge, N. (2017). Geometric mean quantity index numbers with benefit-of-the-doubt weights. *European Journal of Operational Research*, 256(3):1004–1014.

Yang, L. (2014). An Inventory of Composite Measures of Human Progress. Technical report, UNDP Human Development Report Office.