Disentangling heterogeneity gaps and pure performance differences in composite indexes over time: the case of the Europe 2020 strategy*

September 12, 2022

Abstract

Composite indexes have been widely used in different contexts to evaluate the performances of entities. The popularity of these indexes come mainly from their ease of use, interpretation and computation. In practice, heterogeneity could be present in the performance evaluation exercise due to differences between entities. For example, they could differ with respect to their ownership, geographical localization, economic infrastructure, resource endowments, social environment, and so on. As a result, composite indexes fail to capture the performance differences, as they are biased by the presence of heterogeneity. In this paper, we suggest a simple procedure to disentangle the heterogeneity gaps and pure performance differences in the composite indexes over time. We apply our procedure to the case of the Europe 2020 strategy by distinguishing between old and new European members.

Keywords: Composite Index; Heterogeneity gap; Meta Index; Europe 2020 strategy.

*We thank the Editor in Chief Filomena Maggino and three anonymous referees for their comments that improved the paper substantially.

1 Introduction

The concept of the composite index has been widely used in the context of evaluating the performances of entities. The popularity of this type of index could be explained by two main reasons. On the one hand, composite indexes are easy to construct. In fact, constructing a composite index requires choosing the normalization, the weighting, and the aggregation of the selected indicators. The normalization makes sure that indicators have the same unit. Different options are possible: the minimize-maximum, the distance to a reference, or the percentage change methods. Next, for the weighting, we have to choose between exogenous or endogenous weights. Finally, popular aggregation techniques include the arithmetic and geometric averages, and the non-compensatory multi-criteria approach. Refer to OECD (2008) for more discussion on how to construct composite indexes. On the other hand, composite indexes are unit free, and can thus be used to compare and benchmark entities.

In practice, entities could differ in several aspects. For example, they can differ with respect to their ownership, geographical localization, economic infrastructure, resource endowments, social environment, operational settings, and so on. These differences between entities imply the presence of heterogeneity in the performance evaluation exercise. Therefore, whatever the choices made for the normalization, the weighting, and the aggregation of the selected indicators, the resulting composite index will fail to capture only the performance differences between entities, as it will be influenced by the presence of heterogeneity. In a sense, the presence of heterogeneity between entities induces a bias for the composite indexes.

In this paper, we suggest a simple procedure to disentangle heterogeneity gaps from pure performance differences in the composite indexes. The procedure we suggest relies on the benefit-of-the-doubt approach for composite indexes (see, for example, Cherchye et al (2007b) for more discussion). The benefit-of-the-doubt approach does not require making a choice for the normalization, is based on endogenous weights, and uses the arithmetic average for the aggregation scheme. In fact, it suffices to compute two composite indexes to isolate heterogeneity in the performance evaluation exercise. As a result, our procedure shares the virtues of this approach: weights are given by solving mathematical models, and normalization problems may be sidestepped. However, there are also drawbacks: the weights could be too extreme or unrealistic. At this point, we want to emphasize that our procedure could also be seen as a potential solution for this drawback. See Section 2 for more detail.

We apply our methodology to the case of the Europe 2020 strategy. Using a composite index to quantify, measure and monitor the progress of the European countries towards the achievement of the objectives of this strategy has been considered by Saltelli et al (2011), Colak and Ege (2013), Pasimeni (2012, 2013), Rappai (2016), and Walheer (2017). While the composite indexes used in these works might differ (in terms of normalization, weighting, and/or aggregation methods), they all find the same main result: old European members present, on average, better performances than new European members. A natural question is therefore to investigate whether the differences between the two groups found by these authors is due to pure performance differences or due to the presence of heterogeneity between the two groups. Heterogeneity between the two groups could be due to, for example, institutional differences, the date of joining the EU, geographical localization, economic infrastructure, resource endowments, growth capacity, or social environment. At this point, we remark that Pasimeni and Pasimeni (2016) asked a similar question to ours by investigating whether differences in economic growth, public finances, and institutional factors have an impact on the performance differences of the countries. As such, our empirical study might be seen as a complement to their work.

The rest of this paper unfolds as follows. In Section 2, we present our simple procedure to disentangle heterogeneity from pure performance differences in the composite index. In Section 3, we apply the methodology to the case of the Europe 2020 strategy. In Section 4, we present our conclusions.

2 Methodology

We assume that we observe N entities (e.g. countries, firms, etc.) during T periods of time. The distinguishing feature of our methodology is to recognize the presence of heterogeneity between entities. Indeed, they can differ with respect to their ownership, geographical localization, economic infrastructure, resource endowments, social environment, operational settings, and so on. In practice, heterogeneity is captured by assuming that the entities are partitioned into groups. We assume that there are K groups, and that each group k contains N_k entities. As such, the sample size is alternatively given by $N_1 + \cdots + N_K = N$. Moreover, we assume that for every entity j in each group k at time t, we observe I indicators, denoted by

$\mathbf{x}^k_{jt} = (x^k_{1jt}, \ldots, x^k_{ijt}, \ldots, x^k_{Ijt})$.

Our procedure to disentangle heterogeneity from pure performance differences in the composite index is easy to use and composed of three main steps. We first consider decomposition for static settings, and next extend it to dynamic settings.

2.1 Static settings

Step 1: composite index. The first step is to compute the composite indexes for every entity. Following the benefit-of-the-doubt approach, the composite index for entity j in group k at time t is obtained using the following linear program:

$$
\begin{aligned}
CI^k_{jt}(\mathbf{x}^k_{jt}) = & \max_{\omega^k_{ijt}\ (i \in \{1,\ldots I\})} \sum_{i=1}^{I} \omega^k_{ijt} x^k_{ijt} \\
& \textbf{(C-1)} \ \sum_{i=1}^{I} \omega^k_{ijt} x_{ist} \leq 1, \ \text{for } s = 1, \ldots, N, \\
& \textbf{(C-2)} \ \omega^k_{ijt} \geq 0, \ \text{for } i = 1, \ldots, I. \qquad (1)
\end{aligned}
$$

$CI^k_{jt}(\mathbf{x}^k_{jt})$ is by definition between 0 and 1. A value of 1 reflects a situation where entity j in group k at time t has the best performance. Smaller values indicate worse performances. Also, note that the indicators are not required to be between 0 and 1. This is an advantage of the benefit-of-the-doubt approach. See, for example, Freudenberg (2003) for more discussion about normalization of indicators, and its consequence for the composite index. Finally, we point out that (1) gives full flexibility to the weights when computing the composite index. That is, the entities are evaluated in the best possible light. To avoid too extreme weights, additional constraints for the weights could be included in (1). See Cherchye et al (2007b, 2008) for more discussion, and our empirical study in Section 3 for an illustration. Finally, note that t on CI^k_{jt} captures the time period of the weights, and thus for the peers used in **(C-1)**.

Step 2: group composite index. The second step is to compute the group composite indexes for every entity. Attractively, these indexes can also be obtained using a linear program. The group composite index for entity j in group k at time t is obtained as follows:

$$
\begin{aligned}
GCI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = \max_{\delta_{ijt}^{k}\ (i\in\{1,\ldots I\})} \sum_{i=1}^{I} \delta_{ijt}^{k} x_{ijt}^{k} \\
\textbf{(C-1)}\ \sum_{i=1}^{I} \delta_{ijt}^{k} x_{ist}^{k} \leq 1,\ \text{for } s = 1, \ldots, N_k, \\
\textbf{(C-2)}\ \delta_{ijt}^{k} \geq 0,\ \text{for } i = 1, \ldots, I.
\end{aligned} \tag{2}
$$

The linear program (2) looks very similar to the linear program (1), the only difference is that in (2), the comparison peers are restricted to group k. While this difference might seem anecdotic, it is the key point of the suggested approach. Indeed, by restricting the comparison to the entities that belong to the same group, heterogeneity is removed from the evaluation process. As a result, $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k})$ captures the pure performance difference of entity j in group k at time t. $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k})$ has to be interpreted in an analogous manner to $CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, but, this time, the comparison partners are restricted to group k. Therefore for every entity j in group k at time t, we have, by construction, that $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k}) \geq CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$. That is, when comparing within the group, the performances can only be better than when comparing with all entities. Clearly, the remark made previously for (1) about adding extra constraints for the weights also applies to (2). Finally, we point out that once more the subscript t on GCI_{jt}^{k} captures the time period of the weights, and thus for the peers used in **(C-1)**.

Step 3: meta index. To decompose the composite index $CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, we introduce the notion of a meta index defined for every entity j in group k at time t as follows:

$$
MI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = \frac{CI_{jt}^{k}(\mathbf{x}_{jt}^{k})}{GCI_{jt}^{k}(\mathbf{x}_{jt}^{k})}. \tag{3}
$$

$MI_{jt}^{k}(\mathbf{x}_{jt}^{k})$ captures the heterogeneity gap for entity j in group k at time t. As noted previously, the group composite index is by construction larger than the composite index: $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k}) \geq CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, making the meta index smaller than 1: $MI_{jt}^{k}(\mathbf{x}_{jt}^{k}) \leq 1$. When $MI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = 1$, it means that $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, revealing the absence of heterogeneity. When $MI_{jt}^{k}(\mathbf{x}_{jt}^{k}) < 1$, a heterogeneity gap is present. A smaller value for $MI_{j}^{k}(\mathbf{x}_{jt}^{k})$ induces a greater heterogeneity gap for entity j in group k at time t.

Building on our notion of a meta index, we can decompose the composite index for every entity j in group k at time t in two parts as follows:

$$CI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = GCI_{jt}^{k}(\mathbf{x}_{jt}^{k}) \times MI_{jt}^{k}(\mathbf{x}_{jt}^{k}). \tag{4}$$

This relationship decomposes the composite index $CI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, for every entity j in group k at time t, as a product of the group composite index $GCI_{jt}^{k}(\mathbf{x}_{jt}^{k})$ and the meta index $MI_{jt}^{k}(\mathbf{x}_{jt}^{k})$. In words, this equation allows us to disentangle pure performance differences, captured by the group composite index, from the heterogeneity gap, captured by the meta index. When there is no heterogeneity gap, i.e. $MI_{jt}^{k}(\mathbf{x}_{jt}^{k}) = 1$, the composite index correctly measures the performance difference. The higher the heterogeneity gap, i.e. smaller than 1 is $MI_{jt}^{k}(\mathbf{x}_{jt}^{k})$, the worse the measure of the composite index of the pure performance differences. Note that, in that case, the composite index does not reflect pure performance differences between entities as it is, in a sense, biased by the presence of heterogeneity.

We end this part by four remarks. Firstly, linear programs (1) and (2) are special cases of the linear programs to compute efficiency scores, as suggested by Charnes, Cooper and Rhodes (1978). In fact, (1) is formally equivalent to the input-oriented efficiency model with multiple outputs and a dummy input of one of Charnes, Cooper and Rhodes (1978). (2) may also be seen as a special case of their linear program when peers are restricted to group k. Next, we label $MI_{j}^{k}(\mathbf{x}_{jt}^{k})$ the meta index given its connection with the meta production function of Hayami and Ruttan (1970), the meta frontier of Battese and Rao (2002), and the meta technology gap of Battese, Rao, and O'Donnell (2004). As in those methods, the meta index tries to take the heterogeneity between entities (between production units in their cases) into account. Afterwards, our decomposition in (4) could also be seen as a potential solution for the extreme or unrealistic weights sometimes given by the benefit-of-the-doubt approach. Indeed, by removing heterogeneity in the performance evaluation exercises, entities are compared to entities that are closer to them. As such, it is less likely to find extreme solutions. Finally, a similar decomposition has recently been suggested by Karagiannis and Karagiannis (2018). In fact, while our starting points are slightly different and we label indexes in a different manner, we end with the same decomposition of the composite index. The distinguishing feature of our procedure is that we also consider the decomposition for dynamic settings as discussed in the next Section.

2.2 Dynamic settings

Step 1: composite index change. Assume we are interested in the change in the composite index between two periods: b (the base period) and c (the current period). To avoid choosing a particular year for the weights, Cherchye et al (2007a) suggested defining the change in composite index as the geometric average of the changes where years b and c are taken as the reference year for the weights (and thus for the peers):

$$\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \left[\Delta CI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta CI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\right]^{1/2}, \tag{5}$$

where $\Delta CI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{CI_{jb}^k(\mathbf{x}_{jc}^k)}{CI_{jb}^k(\mathbf{x}_{jb}^k)}$ is the composite index change when year b is chosen for the weights, and $\Delta CI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{CI_{jc}^k(\mathbf{x}_{jc}^k)}{CI_{jc}^k(\mathbf{x}_{jb}^k)}$ when year c is chosen. The benchmark value for $\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ is one, meaning that a value greater than unity implies a performance progress while the converse reflects a performance regress.

Attractively, we can decompose $\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ into two components as follows:

$$\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k), \tag{6}$$

where

$$\Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{CI_{jc}^k(\mathbf{x}_{jc}^k)}{CI_{jb}^k(\mathbf{x}_{jb}^k)}, \tag{7}$$

$$\Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \left[\frac{CI_{jb}^k(\mathbf{x}_{jc}^k)}{CI_{jc}^k(\mathbf{x}_{jc}^k)} \times \frac{CI_{jb}^k(\mathbf{x}_{jb}^k)}{CI_{jc}^k(\mathbf{x}_{jb}^k)}\right]^{1/2}. \tag{8}$$

$\Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ captures the catching-up of entity j in group k between the base and current periods, and $\Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ measures change in the best possible performance between periods b and c, i.e. change in the policy environment. Both components also have a benchmark value of unity.

As discussed before, the composite index does not capture pure performance of the entities, since it is biased by the presence of heterogeneity. Clearly, the same holds true for $\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$, and also for the two components of its decomposition $\Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ and $\Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$. Fortunately, we can decompose $\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ and its two components into a pure performance change and a heterogeneity gap change in a similar fashion to that in (4) when considering static settings.

Step 2: group composite index change. Given its similarity with the composite index discussed previously, we can apply a similar reasoning for the change in group composite index:

$$\Delta GCI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \left[\Delta GCI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta CGI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\right]^{1/2}, \tag{9}$$

$$= \Delta GCU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta GEC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k), \tag{10}$$

where $\Delta GCI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{GCI_{jb}^k(\mathbf{x}_{jc}^k)}{GCI_{jb}^k(\mathbf{x}_{jb}^k)}$, $\Delta GCI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{GCI_{jc}^k(\mathbf{x}_{jc}^k)}{GCI_{jc}^k(\mathbf{x}_{jb}^k)}$, $\Delta GCU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{GCI_{jc}^k(\mathbf{x}_{jc}^k)}{GCI_{jb}^k(\mathbf{x}_{jb}^k)}$, and $\Delta GEC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \left[\frac{GCI_{jb}^k(\mathbf{x}_{jc}^k)}{GCI_{jc}^k(\mathbf{x}_{jc}^k)} \times \frac{GCI_{jb}^k(\mathbf{x}_{jb}^k)}{GCI_{jc}^k(\mathbf{x}_{jb}^k)}\right]^{1/2}$. All these components have to be interpreted in a similar fashion to the decomposition defined for the composite index, but at the level of group k. It turns out that these new index changes capture a pure performance change in group k.

Step 3: meta index change. We cannot apply the same procedure for the meta index since this index is not computed using a linear program, but rather defined a posteriori using the composite and group composite indexes, as stated in (3). Nevertheless, a similar decomposition can be obtained using our previous results. An initial observation is that we can generalize (4) when period b or c is chosen for the weights:

$$\Delta CI_{ju}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) = \frac{GCI_{ju}^k(\mathbf{x}_{jc}^k) \times MI_{ju}^k(\mathbf{x}_{jc}^k)}{GCI_{ju}^k(\mathbf{x}_{jb}^k) \times MI_{ju}^k(\mathbf{x}_{jb}^k)},$$

$$= \Delta GCI_{ju}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MI_{ju}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k), \text{ for } u = \{b, c\}. \tag{11}$$

$\Delta MI_{ju}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ is the change in meta index between years b and c when period u is chosen for the weights. Again, the benchmark value is one, and a larger value implies an improvement while smaller value a regress. Combining the previous equation with

(5) gives us:

$$\begin{aligned}\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) &= \big[\big(\Delta GCI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\big) \\ &\quad \times \big(\Delta GCI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\big)\big]^{1/2}, \\ &= \big[\big(\Delta GCI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta GCI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\big) \\ &\quad \times \big(\Delta MI_{jb}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MI_{jc}^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\big)\big]^{1/2}, \\ &= \big[\Delta GCI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)\big]^{1/2}. \end{aligned} \tag{12}$$

$\Delta MI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ is the change in the meta index irrespective of the chosen period for the weights. Therefore, (12) parallels (4) for dynamic settings. In words, the composite index change ($\Delta CI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$) is decomposed into a pure performance change ($\Delta GCI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$) and a heterogeneity gap change ($\Delta MI_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$).

Finally, we can also decompose the catching-up and environmental change components in a similar fashion as follows (using (11)):

$$\begin{aligned}\Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) &= \frac{GCI_{jc}^k(\mathbf{x}_{jc}^k) \times MI_{jc}^k(\mathbf{x}_{jc}^k)}{GCI_{jb}^k(\mathbf{x}_{jb}^k) \times MI_{jb}^k(\mathbf{x}_{jb}^k)}, \\ &= \Delta GCU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MCU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k). \end{aligned} \tag{13}$$

$$\begin{aligned}\Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) &= \left[\frac{GCI_{jb}^k(\mathbf{x}_{jc}^k) \times MI_{jb}^k(\mathbf{x}_{jc}^k)}{GCI_{jc}^k(\mathbf{x}_{jc}^k) \times MI_{jc}^k(\mathbf{x}_{jc}^k)} \times \frac{GCI_{jb}^k(\mathbf{x}_{jb}^k) \times MI_{jb}^k(\mathbf{x}_{jb}^k)}{GCI_{jc}^k(\mathbf{x}_{jb}^k) \times MI_{jc}^k(\mathbf{x}_{jb}^k)}\right]^{1/2}, \\ &= \left[\frac{GCI_{jb}^k(\mathbf{x}_{jc}^k)}{GCI_{jc}^k(\mathbf{x}_{jc}^k)} \times \frac{GCI_{jb}^k(\mathbf{x}_{jb}^k)}{GCI_{jc}^k(\mathbf{x}_{jb}^k)}\right]^{1/2} \times \left[\frac{MI_{jb}^k(\mathbf{x}_{jc}^k)}{MI_{jc}^k(\mathbf{x}_{jc}^k)} \times \frac{MI_{jb}^k(\mathbf{x}_{jb}^k)}{MI_{jc}^k(\mathbf{x}_{jb}^k)}\right]^{1/2}, \\ &= \Delta GEC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k) \times \Delta MEC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k). \end{aligned} \tag{14}$$

$\Delta MCU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ and $\Delta MEC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ are the change in the heterogeneity gap for the catching-up and environmental components, respectively. They give the option to decompose the catching-up and the environmental change ($\Delta CU_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$ and $\Delta EC_j^k(\mathbf{x}_{jb}^k, \mathbf{x}_{jc}^k)$) into a pure index change and a change in the heterogeneity gap.

Combining (10) with (13) and (14), we obtain a fourpartite decomposition of the composite index change $\Delta CI^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc})$ between periods b and c:

$$\begin{aligned}\Delta CI^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) =& \Delta CU^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) \times \Delta EC^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}),\\ =& \Delta GCU^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) \times \Delta MCU^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc})\\ & \times \Delta GEC^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) \times \Delta MEC^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}).\end{aligned} \tag{15}$$

This decomposition is attractive since it gives us the option to better understand the reasons of the change in composite index. We point out that all the components can be computed using linear programs (1) and (2). In fact, it suffices to change the years of the evaluated entity and the peers to b and/or c to obtain all the components. We note that the components for the meta index are always obtained a posteriori once the other components have been computed. Finally, combining all previous equations gives us the decomposition of the meta index change into two parts:

$$\Delta MI^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) = \Delta MCU^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}) \times \Delta MEC^k_j(\mathbf{x}^k_{jb}, \mathbf{x}^k_{jc}). \tag{16}$$

3 The case of the Europe 2020 strategy

We apply our methodology to the case of the Europe 2020 strategy. This strategy was introduced by the European Union to replace the Lisbon Strategy that ended in 2010 (Refer to European Commission (2010) for more detail). Contrary to the Lisbon Strategy, the Europe 2020 strategy is accompanied by clear and measurable targets for the objectives. Recently, composite indexes have been used to evaluate, manage, and monitor the performances of the European countries towards the achievement of this strategy. Refer to Saltelli et al (2011), Colak and Ege (2013), Pasimeni (2012, 2013), Rappai (2016), and Walheer (2017).

While the composite index used in these works may differ (in terms of normalization, weighting, and/or aggregation methods), they all find the same conclusion: old European members (i.e. Belgium, Denmark, Germany, Ireland, Greece, Spain, France, Italy, Luxembourg, Netherlands, Portugal, Finland, Sweden, and the United Kingdom) present, on average, better performances than the new members (i.e. Bulgaria, Czech Republic, Estonia, Croatia, Cyprus, Latvia, Lithuania, Hungary, Malta, Austria, Poland, Romania, Slovenia, and Slovakia). As such, we ask the following question: are the composite index differences between old and new members due to pure performance differences or are they due to heterogeneity between the two groups? For policy makers, this distinction is of high importance as it will allow them to better target their policy implementations, and to better assess the potential pay-offs of such policies.

To present our empirical application, we first present the data and discuss the normalization of the indicators. Subsequently, we present the results of the composite indexes and the decomposition into pure performance differences and the heterogeneity gaps.

Data and normalization of the indicators. Eurostat, the official statistical institutional of the European Union, provides tailored data to study the Europe 2020 strategy from 2004 to 2015. In particular, data are divided into the three pillars as follows:

- Smart Growth:
  - Making sure that at least 40% of youngsters have a degree or diploma: measured by the percentage of people, for the age group 30-34, that have successfully completed university or university-like education.
  - Raising the investment in R&D to 3% of the GDP: measured by the ratio of the expenditure of R&D over GDP.
- Sustainable Growth:
  - Cut greenhouse gas emissions by 20%: given as an index taking 1990 as the base year.
  - Source 20% of its energy needs to be from renewable sources: measured as the share of renewable energy in gross final energy consumption.
  - Increase energy efficiency by 20%: measured by the gross inland consumption of energy divided by GDP.
- Inclusive Growth:
  - Raising the employment rate of the population aged 20-64 to 75%: measured by the rate of employment in the age group 20-64.

- Reducing the share of early school leavers to under 10%: measured by the percentage of the population aged 18-24 with, at most, lower secondary education and not in further education or training.
- Reducing the number of people living below the poverty line by 25%: measured by the people at risk of poverty or social exclusion.

Note that there are two types of objective for the Europe 2020 strategy: positive and negative. The former type is when countries have to increase their indicator level to reach the target, while the latter is when countries have to decrease their indicator level to meet the requirement.

An important step for our analysis is to include the targets of the objectives into the indicators $x^k_{1jt}, \ldots, x^k_{ijt}, \ldots, x^k_{Ijt}$. A procedure introduced by Colak and Ege (2013) and used in Walheer (2017) defines the indicators as follows:

$$x^k_{ijt} = \begin{cases} \frac{y^k_{ijt} - y_{im}}{z^k_{ij} - y_{im}}, & \text{for a positive objective,} \\ \frac{y_{iM} - y^k_{ijt}}{y_{iM} - z^k_{ij}}, & \text{for a negative objective,} \end{cases} \tag{17}$$

where $y^k_{1jt}, \ldots, y^k_{ijt}, \ldots, y^k_{Ijt}$ are the non-normalized indicators for entity j in group k at time t, i.e. those given by Eurostat; y_{im} is the minimum of the non-normalized indicators for objective i; y_{iM} is the maximum of the non-normalized indicators for objective i; and z^k_{ij} is the target level for objective i and country j in group k. Note that the target levels are country specific as set up by the Europe 2020 strategy. Also, note that most of the target levels (i.e. z^k_{ij}) are provided by Eurostat. Only two targets are not clearly given: the targets for the energy efficiency objective and for the people at risk-of-poverty objective. Indeed, for those objectives, the referent year is not clearly specified. As done in Colak and Ege (2013) and Walheer (2017), we define the target for the energy efficiency objective by multiplying the indicator of 2008 by 0.8, and the target for the people at risk-of-poverty objective by multiplying the indicator of 2008 by 0.75.

As defined in (17), x^k_{ijt} quantifies the distance between the non-normalized indicators and the target levels for objective i and entity j in group k at time t. A value smaller than 1 means that the objective is not achieved as the non-normalized indicator is smaller than the target level. A value of 1 reveals that the objective is reached as the non-normalized indicator coincides with the target level. Finally, a value greater than 1 implies that the target level is surpassed.

Results. We compute using the linear programs (1) and (2) the composite and group composite indexes for every year and country. Note that to avoid extreme weights when computing the indexes, we impose restriction to the weights in (1) and (2). In particular, we impose that each objective has a weight between 8% and 25%. We believe that those weights are reasonable in light of those chosen in previous works: Colak and Ege (2013) impose exogenous weights between $\frac{1}{12}$ and $\frac{1}{6}$, and Pasimeni (2012, 2013) and Walheer (2017) rely on uniform weights (i.e. $\frac{1}{8}$). We remark that the weight restrictions thus take the form of 'proportion constraints' or 'pie-share restriction' in (1) and (2); see for example Cherchye et al (2007b) for more discussion. In the following, to improve the readability of our Figures, we order the European countries using two criteria: old and new countries, and alphabetical order. Also, we make use of abbreviations for the country names. See Table 1 in the Appendix for a full list.

As an initial step, we plot the average of the composite, the group composite, and the meta index in Figure 1. We also present the averages per group in that Figure. At this point, we remark that using a simple arithmetic average to obtain an aggregate value at the group level is a coherent aggregation scheme (see Rogge (2018) for the composite index and Walheer (2018) for the meta index).

Figure 1: Averages of the composite, group composite and meta index per country

Source: author's computations

The composite index is, on average, 0.58 for the period when considering all countries. It is 0.66 when considering only old members and 0.49 for new members. The group composite index provides some light with regard to this aspect. The average of that index is 0.62 when considering all countries. That is higher than the average of the composite index (0.58). This already indicates the presence of a heterogeneity gap between the countries. To determine the direction of the heterogeneity gap, we can check the values of the group composite index for both groups separately. For the old members, the group composite index coincides, on average, with the composite index, revealing that this group represents the best practice in Europe. Note that it also indicates that old members form a homogeneous group. For the new members, the group composite index is, on average, larger than the composite index meaning that the performance of this group is probably under-estimated when comparing with the old members. In other words, the presence of heterogeneity between the two groups implies that the composite indexes do not reflect the pure performance differences, and is, in a sense, biased. The meta index confirms our initial results based on the group composite index. Indeed, it is, on average, 0.91 for the period when considering all countries. For old members the value is 1 on average, confirming the best practice status of those countries; while it is 0.83 for the new members. This clearly confirms the presence of heterogeneity between old and new members.

Sweden, Finland, and Denmark with composite indexes and group composite indexes close to 1 are the best countries for these dimensions. The worst countries for these dimensions include Romania, Malta, Cyprus, Bulgaria, and Greece. This is confirmed by both the composite and group composite indexes. The countries with the highest differences between the composite and group composite indexes are Croatia, the Czech Republic, Estonia, Lithuania, Hungary, Slovenia, and Slovakia. That is only new members. Note also the high group composite indexes of the Czech Republic, Austria, and Estonia. Except for the United Kingdom and Luxembourg with meta indexes of 0.98 and 0.99, respectively, all meta indexes for the old members are equal to 1. The new members all have a meta index less than 1. Note that the highest values for the meta index are observed for Malta, Austria, Slovenia, Estonia, Hungary, and Lithuania; and the lowest values are those of Romania, Slovakia, Croatia, and Latvia.

Next, in Figure 2, we present the cumulative changes of the composite index (CI) and of the decomposition components, i.e. the catching up component (CU) and the environmental change component (EC). We choose to rely on the cumulative changes, instead of the changes, since the former gives an overview of the overall progress over the period and is therefore more informative. Note that the cumulative changes are obtained by multiplying the year-to-year index values, and that it is one for 2004, by construction. Higher values imply more progress. As done before, we present the cumulative changes when pooling all countries and when distinguishing between old and new members.

The cumulative change for the composite index is 1.71 when considering all countries, while it drops to 1.36 for old members and rises to 2.11 for new members. The difference in composite index is mainly due to a catching-up effect: cumulative

Figure 2: Cumulative changes for the composite index, the catching-up and the environmental change components

Source: author's computations

change of 1.81 for old members and of 3.80 for new members. Nevertheless, it is not important enough to reach the same level as the old members. Moreover, the environmental change component is positive for the old members (1.08) due to an important increase after 2011, while it is negative for the new members (0.79) due to a negative trend for the period. This implies a more favourable policy environment for old members than for new members.

We present the cumulative changes for the composite index, the catching-up component, and the environmental change component per country in Figure 3. We clearly see important differences between old and new members in that Figure. The amplitudes for the changes are highest, in general, for the new members. We point out the important increase for Bulgaria, Malta, and Romania. These countries were among the worst countries in terms of composite index, as highlighted in Figure 1. Their important increase is only due to a positive catching-up effect, while the environmental change is negative. We also point out the importance increase for Luxembourg and Portugal in the old member group. This is explained again by a positive catching-up effect.

Figure 3: Cumulative changes for the composite index, the catching-up and the environmental change components per country

Source: author's computations

A natural question in light of these results is to know whether these differences in the composite indexes, catching-up components, and environmental change components are due to pure performance differences or because of the existence of a heterogeneity gap between the old and new member groups. To answer this question, we first present the cumulative change of the composite (CI) , group composite (GCI), and meta index (MI) when pooling countries in Figure 4, and per country in Figure 5.

At the overall level, we see that the group composite index follows a similar trend to the composite index. This is even truer for the old members. In other words, this

Figure 4: Cumulative changes for the composite, group composite, and meta index

Source: author's computations

confirms the greater importance of the heterogeneity gap for the new members. That is, the under-estimation of the new member performances is more and more an issue over time. In fact, it seems that the heterogeneity between the two groups slightly increases with time. Indeed, the group composite index is farther and farther to the composite index over years. It turns out that the role of the heterogeneity gap for the performance evaluation of the Europe 2020 strategy should not be skipped.

Figure 5: Cumulative changes for the composite, group composite, and meta index per country

Source: author's computations

The cumulative changes per country, displayed in Figure 5, also provide important information. For old members, all the countries face an improvement of their composite index, except Italy (-0.02). Note that it is due to a pure performance regress. In fact, only the United Kingdom and Luxembourg have an improvement of their meta index (+0.05 and +0.07, respectively). For the other countries, the improvement of their composite index is only due to better performance. Note also the important increase of Portugal.

For the new members, they all present an improvement of their composite index, but the causes are clearly different between countries. The highest increases are for Malta, Bulgaria, Romania, Poland, and Latvia. Note that these improvements are the highest in Europe for the period. Also, these improvements are due to an increase of both the group composite and the meta indexes in the case of Romania, Poland, and Latvia; but only due to the group composite index for Malta and Bulgaria (the meta index change is even negative). Besides those five countries, interesting patterns are highlighted by our method. Croatia, Cyprus, the Czech Republic and Slovenia present an increase of their meta index, resulting in an increase of their composite index. In the case of Croatia, the increase of the composite index is almost only due to the increase in the meta index.

We end our empirical analysis by presenting the results of the decomposition of the composite index change into four components: group and meta catching-up components (GCU and MCU), and group and meta environmental change components (GEC and MEC). In Figure 6, we present the results when grouping countries, and in Figure 7, we give the results per country.

Figure 6: Cumulative changes for the fourpartite decomposition

Source: author's computations

We once more see important differences between old and new members. The amplitude of the changes for all components is clearly larger for the latter group. Nevertheless, both groups have in common that the group catching-up component is the main reason for the increase in the composite index. The group environmental change has a positive contrition for the old members, but a negative contribution for the new members. For the heterogeneity gap, we find an improvement for the catching-up component, but not for the environmental change component. It means that the policy environmental context between old and new members differs more and more with time. Clearly, policy implementations could help here to counter the increase in heterogeneity.

Figure 7: Cumulative changes for the fourpartite decomposition per country

Source: author's computations

At the country level, we better understand the results discussed before for Bulgaria, Malta, Romania, and Poland. Their important increase is mainly due to a huge progress of their group catching-up component. The same holds true for Portugal in the new member group. There is a decrease of that component for Italy (this explains its regress found previously), Croatia, Slovenia, and Finland. It implies that these countries have a negative catching-up effect in their respective group. Policy environmental contexts at the group level have benefited Hungary, Estonia, Austria, Lithuania, Slovenia, Croatia, and Italy; but not for Malta, Bulgaria, Romania, Cyprus, and Portugal. This also explains their lowest position in their group. Improvements for the meta catching-up component are found for all the new members, except Malta and Bulgaria. For the old members, it is the status quo, expect for Luxembourg and the United Kingdom. Finally, the meta environmental change component is negative for all countries, except Malta. We point out that the highest declines occur only for the new members.

4 Summary and discussion

We summarize our findings in four main points:

- Our empirical study confirms the presence of heterogeneity between old and new European members. In fact, we demonstrate that the differences in composite indexes are due to both pure performance differences and heterogeneity gaps.

- Over the years, we observe a catching-up effect of the new members, while old members still have better performances on average. In general, this catching-up effect is due to both an increase of the performances and a decrease of the heterogeneity gaps; but this does not hold true for all the countries.

- We find a decrease of the heterogeneity gaps for the catching-up component, but an increase in terms of policy environment. It means that new and old members have benefited differently from policy implementations.

- We highlight the case of Malta and Bulgaria in the new member group, and of Portugal and Italy in the old member group.

All in all, we believe that our findings are important for policy makers. Indeed, by distinguishing between pure performance differences and heterogeneity gaps, it allows them to better target their policy implementations, and to better quantify the potential pay-offs of such policy implementations. Indeed, when heterogeneity is not taken into consideration it could give a biased vision of the performance differences, and thus may imply inefficient policy implementations. It also means that specificities of the European countries should be considered to obtain a fair comparison.

Finally, we highlight limitations of our study and routes for further research. Firstly, the proposed approach could be extended when considering other normalization-weighting-aggregation methods when defining the composite indexes. Next, specific weight restrictions may be developed to increase the realism of the obtained weights in our context of heterogeneity gap (inspiration can be found in Cherchye et al (2008)). Finally, our way to model heterogeneity, while attractive and simple to use, is based on the concept of partitioning entities into a number of groups. More advanced methods can be used as an alternative (for example, the concept of spatial heterogeneity, see Fusco, Vidoli and Sahoo (2018)).

5 Conclusion

Composite indexes are easy to use, interpret, and construct; explaining their popularity to evaluate the performances of entities. In this paper, we suggested a simple procedure to disentangle heterogeneity gaps and pure performance differences in the composite indexes over time. Indeed, in many empirical studies, entities differ with respect to their ownership, geographical localization, economic infrastructure, resource endowments, social environment, and so on. These differences induce the presence of heterogeneity in the performance evaluation exercise. As a consequence, composite indexes fail to capture the performance differences, as they are biased by the presence of heterogeneity. Our procedure provides a solution to this issue.

We applied our procedure to the case of the Europe 2020 by distinguishing between old and new European members. Our empirical study reveals the presence of a heterogeneity gap between those two groups that slightly increases with time. We highlighted important patterns at the country level that could be used by policy makers not only to better target their policy implementations, but also to quantify the potential gains of such policy implementations.

References

[1] Battese G. E., Rao D. P., 2002, "Technology gap, efficiency, and a stochastic metafrontier function", *International Journal of Business Economics* 1 (2), 87-93.

[2] Battese G. E., Rao D. P., O'Donnell C. J., 2004, "A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies", *Journal of Productivity Analysis* 21, 91-103.

[3] Charnes A., Cooper W. W., Rhodes E., 1978, "Measuring the efficiency of decision making units", *European Journal of Operational Research* 2, 429-444.

[4] Cherchye L., 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.

[5] Cherchye L., Moesen W., Rogge N., Van Puyenbroeck T., 2007b, "An Introduction to benefit of the doubt composite indicators", *Social Indicators Research* 82, 111-145.

[6] Cherchye L., Moesen W., Rogge N., Van Puyenbroeck T., Saisana M., Saltelli A., Liska R., Tarantola S., 2008, "Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index", *Journal of the Operational Research Society* 59, 239-251.

[7] Colak M. S., Ege A., 2013, "An Assessment of EU 2020 Strategy: Too Far to Reach?", *Social Indicators Research* 110, 659-680.

[8] European Commission, 2010, "Europe 2020: A strategy for smart, sustainable and inclusive growth", *Brussels: European Commissions.*

[9] Freudenberg M., 2003, "Composite indicators of country performance: A critical assessment", *OECD Science, Technology and Industry Working Papers* 16.

[10] Fusco E., Vidoli F., Sahoo B. K., 2018, "Spatial heterogeneity in composite indicator: A methodological proposal", *Omega* 77, 1-14.

[11] Hayami Y., Ruttan V. W., 1970, “Agricultural productivity differences among countries”, *American Economic Review* 60, 895-911.

[12] Karagiannis R., Karagiannis G., 2018, “Intra- and inter-group composite indicators using the BoD model”, *Socio-Economic Planning Sciences* 61, 44-51. osite indicators a methodology and user guide”, *Paris: OECD Publications*.

[13] OECD, 2008, “Handbook on constructing composite indicators – methodology and user guide”, *Paris: OECD Publications*.

[14] Pasimeni P., 2012, “Measuring Europe 2020: a new tool to assess the strategy”, *International Journal of Innovation and Regional Development* 4, 365-385.

[15] Pasimeni P., 2013, “The Europe 2020 Index”, *Social Indicators Research* 110, 613-635.

[16] Pasimeni F., Pasimeni P., 2016, “An Institutional Analysis of the Europe 2020 Strategy”, *Social Indicators Research* 127, 1021-1038

[17] Rappai G., 2016, “Europe En Route to 2020: A New Way of Evaluating the Overall Fulfillment of the Europe 2020 Strategic Goals”, *Social Indicators Research* 129, 77-93.

[18] Rogge N., 2018, “On aggregating Benefit of the Doubt composite indicators”, *European Journal of Operational Research* 264, 364-369.

[19] Saltelli A., D’Hombres B., Jesinghaus J., Manca A., Mascherini M., Nardo M. Saisana M., 2011, “Indicators for EU policies. Business as usual?”, *Social Indicators Research* 102, 197-207.

[20] Walheer B., 2017, “Decomposing the Europe 2020 index”, *Social Indicators Research* accepted.

[21] Walheer B., 2018, “Aggregation of metafrontier technology gap ratios: the case of European sectors in 1995-2015”, *European Journal of Operational Research* 269, 1013-1026.

Appendix

Table 1: Abbreviations

Abbreviation	Country
'B'	Belgium
'D'	Germany
'DK'	Denmark
'E'	Spain
'F'	France
'FIN'	Finland
'GR'	Greece
'I'	Italy
'IRL'	Ireland
'L'	Luxembourg
'NL'	Netherlands
'P'	Portugal
'S'	Sweden
'UK'	United Kingdom
'A'	Austria
'BG'	Bulgaria
'CY'	Cyprus
'CZ'	Czech Republic
'EE'	Estonia
'H'	Hungary
'HR'	Croatia
'LT'	Latvia
'LV'	Lithuania
'MT'	Malta
'PL'	Poland
'RO'	Romania
'SI'	Slovenia
'SK'	Slovakia