



INtegrating MainSTREAM Economic Indicators with Sustainable Development Objectives

<

SEVENTH FRAMEWORK PROGRAMME
THEME FP7-ENV-2007-1

Area 6.4.2.2. Sustainable development indicators
Grant agreement number: 211759

WP6: Costs of sustainability (general equilibrium analysis)

D 6.6a: Results of the sensitivity analysis (FSI)

Francesco Bosello (*)

Elisa Portale (*)

Lorenza Campagnolo (*)

Fabio Eboli (*)

Ramiro Parrado (*)

August 2011

Table of contents

1. Introduction and background.....	2
2. Building the FSI.....	4
SELECTING INDICATORS	4
NORMALIZING INDICATORS.....	4
AGGREGATING INDICATORS	5
3. Sensitivity analysis.....	7
4. Conclusions	11
References	12

1. Introduction and background

The sustainability literature offers one particular feature that presents interesting opportunities for policy evaluations: that is the possibility to develop aggregated measures of sustainability. By compounding the different dimensions of sustainable development, indices have several positive aspects: they allow summarizing the relationship among the variables, facilitate communication to decision makers, and may serve as a basis for “early warning” (UN, 1995). Well-known examples of aggregate sustainability indices include the Human Development Index, the Environmental Sustainability Index, the Environmental Performance Index, the Ecological Footprint, the Index of Sustainable Economic Welfare and the Genuine Savings.

However few issues like the construction and use of composite and aggregate sustainability indicators raise criticisms and debate. The reason is that any step of the process - the choice of indicators to include, of the “weights” to assign to each, the aggregation procedure - are subjectivity prone, no matter the effort made. When this is the case, many criticisms can be perfectly legitimate and correct.

Against this background part of IN-STREAM methodological quantitative research aimed to: (a) explore the potential of composite indicators to provide synthetic measures of sustainability and deliver additional information compared to those conveyed by the “simple” GDP and (b) investigate if and how economic modeling tools, could support this analysis.

These issues have been addressed using a recursive-dynamic general equilibrium model for the world economic system (the Inter-temporal Computable Equilibrium System (ICES) model). The model represents 40 countries/regions, 17 industries and its simulation period is 2010-2020. With this tool a reference and an emission reduction scenarios, where the EU unilaterally cuts its GHG emissions by 20% respect to 1990 in 2020, have been analyzed. 23 sustainable development indicators belonging to the three pillars of sustainability (economic, environmental and social) have been extracted from the model output and compounded into an innovative sustainability index: the “FEEM” sustainability Index (FSI).

The major novelties in the approach followed regard firstly the weighting and aggregation procedures in building the FSI. These take into account possible complementarity or substitutability of performances among different indicators, thus they are not simply “additive”. Weights have been elicited interviewing experts in focus groups.

Secondly, the use of model projections. On the one hand this allows to assess the implications for sustainability in different countries in different *futures*. That is: sustainability can be estimated *ex ante* and not only *ex post*. On the other hand, the internal consistency of the model allows “by construction” a coherent integration of the different dimensions of sustainability inside the composite index. This will also encompass all feedbacks and interconnections among economic systems.

The reader interested in the detailed results' description is addressed directly to Bosello et al., (2011), major findings are however briefly summarized below.

Subjected to given experts' opinions it has been possible to show that:

There is a group of countries composed by more developed economic systems (but not by all of them), where economic, environmental and social sustainability move together. There is another group of countries, mostly, but not necessarily only, developing countries, where at least one dimension of sustainability diverges from the other. This seems to partly support an "environmental Kuznets curve" (EKC) idea: when a given level of economic development is reached, good economic, environmental and social performances could not be in opposition, but below that level the contrast can be stronger.

The first group of countries is also that of the top-performers while the second performs low in terms of FSI. This expresses clearly the idea of sustainability replicated by the composite sustainability indicator: the different dimensions of sustainability are complement and not substitute, therefore a bad performance in any of them greatly lowers the final score.

A direct consequence of this is that one "dominant" (or at least clearly dominant) sustainability component over the other cannot be identified. Thus none of them singularly taken is able to summarize all the informative content of the FSI. This applies particularly to GDP. Its country ranking is very different from that of the FSI. The case of USA is striking. They rank first as per capita GDP, but the FSI places them 21st due to their relatively high GHG emissions per capita and energy intensity .

The EU pursuit of an improvement in environmental sustainability, represented by the implementation of a unilateral 20% emission reduction policy, apparently does not originate conflicts across the different sustainability dimensions within the EU. This is potential good news for policy makers. Indeed all the three sustainability pillars improve. However, the EU policy can trigger potential conflicts with sustainability, especially environmental and social, in the non EU countries. Both are induced by the well known phenomenon of carbon leakages which on the one hand foster the economic performance of carbon intensive sectors in non EU countries, on the other worsen their environmental performance and drain resources from health and education investment.

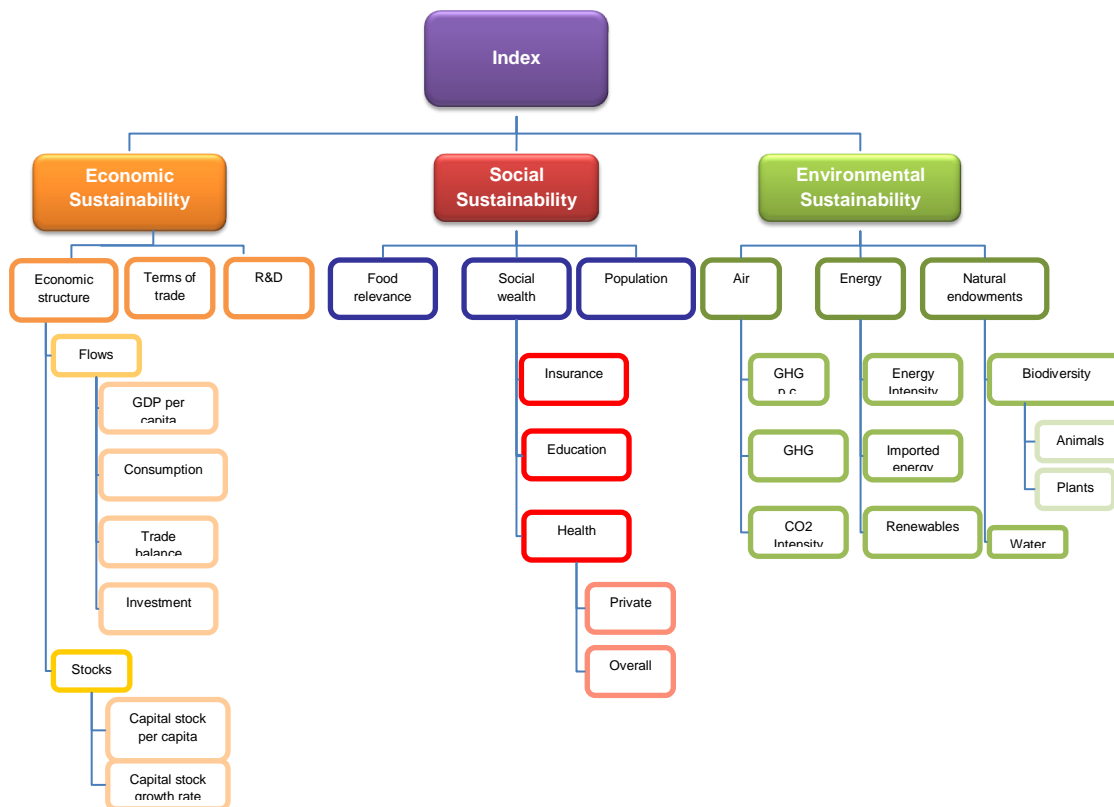
In a complex aggregation such as the one used for the FSI, the weighting procedure, reflecting the attitude of the experts, is a key component of the process. It is thus important to analyse how robust/stable the ranking and then results actually are to change in this attitude. This exercise is the aim of the following sections. Section 2 will briefly recall the FSI construction procedure (more detail in Bosello et al., 2011), section 3 will present the sensitivity analysis, section 4 will conclude.

2. Building the FSI

Selecting indicators

The indicators included in the FSI (see Figure 1) are an enlarged set compared to that considered by INSTREAM D6.2 (Bosello et al., 2010) on their turn a subset of those resulting from the thorough selection process described by INSTREAM D4.1 (Best et al. 2009) defining the scope of quantitative assessment within the project. The strategy adopted by INSTREAM was first of all to focus on a limited number of key indicators on which to perform quantitative assessment. This safeguards transparency and tractability. Then the choice has been made among those indicators which are more commonly used to measure sustainability, explicitly referring to the lists set by the Lisbon Strategy and the EU Sustainable Development Strategy.

Figure 1. The structure of the FSI



Normalizing indicators

Normalisation is required prior to any data aggregation when, as usual, indicators in a data set have different measurement units (see Nardo et al., 2008).

The methods that appeared better-suited for the construction of the FSI is re-scaling applying a series of benchmarks that represent different levels of the indicators¹(for detail see INSTEAM D6.5 (Bosello et al., 2011)).

Each of these levels corresponds to a **normalised value** comprised between **0 and 1** as follows:

Table 1: Normalisation grid

0	Extremely unsustainable situation
0.25	indicator is still not sustainable but not as severely as in the previous case
0.50	a discrete level of sustainability, but still far from target
0.75	satisfactory level in the sustainability, yet not on target
1	target level, fully sustainable

The normalised value corresponds to a specific level of the indicator, defined according to reliable and authoritative literature and international legislation sources, in order to increase the acceptability of the methodology. A different benchmark has been designed for each indicator according to different European and international sources, both at a technical and at a policy level. Whenever possible, the objectives outlined in the EU Sustainable Development Strategy or in the Lisbon Strategy have been used to define one or more level of the benchmarking function. In all other cases broader EU policy objectives and international standards from established institutions such as the OECD or the World Bank have been taken as primary source of information.

The five benchmark values for each indicator define four closed intervals and two open intervals that form a step function, whose form is different according to the values chosen (linear increasing, quadratic, etc). In order to avoid the discontinuity of a step function, each step has been “linearised”, taking the mean values of two subsequent intervals and interpolating, thereby creating a continuous step function.

Aggregating indicators

To aggregate indicators, it is necessary to assign a weight, to determine their relative importance to the final composite index. A broadly used aggregation technique is the **Equally Weighted Average (EWA)**, in which all indicators are given the same weight. It assumes a perfect substitutability among indicators and does not capture the relative importance of different sustainability, nonetheless it is considered the most transparent methodology (Yale Center for Environmental Law and Policy, 2005).

¹ Ranking or using categorical scales would not allow to keep the quantitative aspect of the indicators, and percentages of variations do not clarify whether the variation is sustainable or not. Standardisation using mean and standard deviation has also been excluded as it produces normalised indicators that are not comparable across different years, but only across countries in one specific year, finally rescaling in 0 -1 mostly shows the relative evolution of the value of the indicator in time across countries, rather than providing relevant boundaries. Furthermore, the reference minimum and maximum levels change over the years, making it difficult intertemporal comparison of sustainability.

In fact, perfect substitutability very often fails to be satisfied. In these circumstances the **Non-Additive Measures** approach (**NAM**) can be more appropriate as it allows for the modeling of many types of interactions going from a “compensative” to a “substitutive” attitude of the subject(s) expressing the weights (Decision Maker DM). The first indicates that her/his satisfaction is high only if all the criteria are satisfied (this can be represented by the logical conjunction operator “AND” and reflects a pessimistic or conservative attitude regards sustainability). The second that her/his satisfaction is high if at least one of them is high (this can be represented by the logical disjunction operator OR and reflects an optimistic or compensative attitude regards sustainability).

These attitudes can be measured by an index, depending solely on the measure values, the OR-ness index, together with the AND-ness index. The first one measures the tendency to optimism, while the second one, its complement to 1, measures the tendency to pessimism. If OR-ness =1 the DM is optimistic, implicitly using the maximum operator (logical disjunction), if OR-ness =0, the DM is pessimistic, corresponding to the minimum operator (logical conjunction), if OR-ness =0.5 the DM is additive, and no interaction exist among the criteria (for the detailed computation of the OR-ness index refer to INSTREAM D6.5 Bosello et al., (2011)).

How to obtain these weights? They can be obtained as usual by means of questionnaires. In practice an exponentially increasing numerical complexity is determined by the number of weights involved. Comparing NAM and EWA approaches: if n is the number of the criteria or indicators, the **EWA** approach, only needs to determine n weights, a **NAM** approach requires the specification of 2^n weights, i.e. one for every possible combination of the n criteria (see INSTREAM D6.5 (Bosello et al., 2011) for further detail).

This higher complexity allows the more realistic “violation” of the linearity implicitly assumed in EWA. In fact, the “weight” of a coalition of sub-criteria (group of indicators or nodes) can be greater or lower than the sum of the weights of each of the sub-criterion belonging to the coalition itself².

Subsequently, a simple algorithm, the so-called Choquet Integral (De Waegenaere, 2001; Murofushi, 1994), computes the weighted average, taking the coalitions of criteria into account, instead of considering the single criteria.

The results of the weighting procedure is reported by table 2.

² Naturally, if for every coalition the weight (or the *importance*) of each coalition is formed by the sum of the weights of each sub-set of its criteria forming a partition, we obtain again the EWA. Conversely, if for a coalition its weight is *inferior* to such a sum, a redundant interaction exists among the included criteria, while if it is *greater* than the sum, a synergic interaction exists.

Table 2: Contribution of single indicators to the FSI

Indicator	Contribution
Food relevance	12.22%
Population	9.72%
R&D	9.21%
Water	8.26%
Terms of trade	5.67%
Energy intensity	4.91%
Renewables	4.26%
Education	4.18%
Imported energy	3.93%
CO2 intensity	3.83%
Plants	3.55%
GHG intensity	3.32%
Capital stock per capita	3.23%
Animals	3.21%
Overall health	3.08%
GHG per capita	3.07%
GDP p.c.	2.69%
Insurance	2.47%
Consumption	2.29%
Capital stock growth rate	2.15%
Investment	1.91%
Private health	1.66%
Relative trade balance	1.18%
Sum	100%

Note: different colors correspond to the three areas of sustainability: economic (yellow), social (pink), environmental (green)

According to respondents' judgments the most important indicator is *Food relevance*, followed by *Population* and the ratio of *R&D expenditure*. The indicator with the lowest marginal contribution is *Relative trade balance*. A noteworthy feature of the contribution of indicators to the final index is that they are influenced both by the evaluation of decision makers, and by the structure of the tree, in which some indicators contribute more directly than others to the final index.

The average OR-ness of this set of weight is 0.42, therefore, quite conservative.

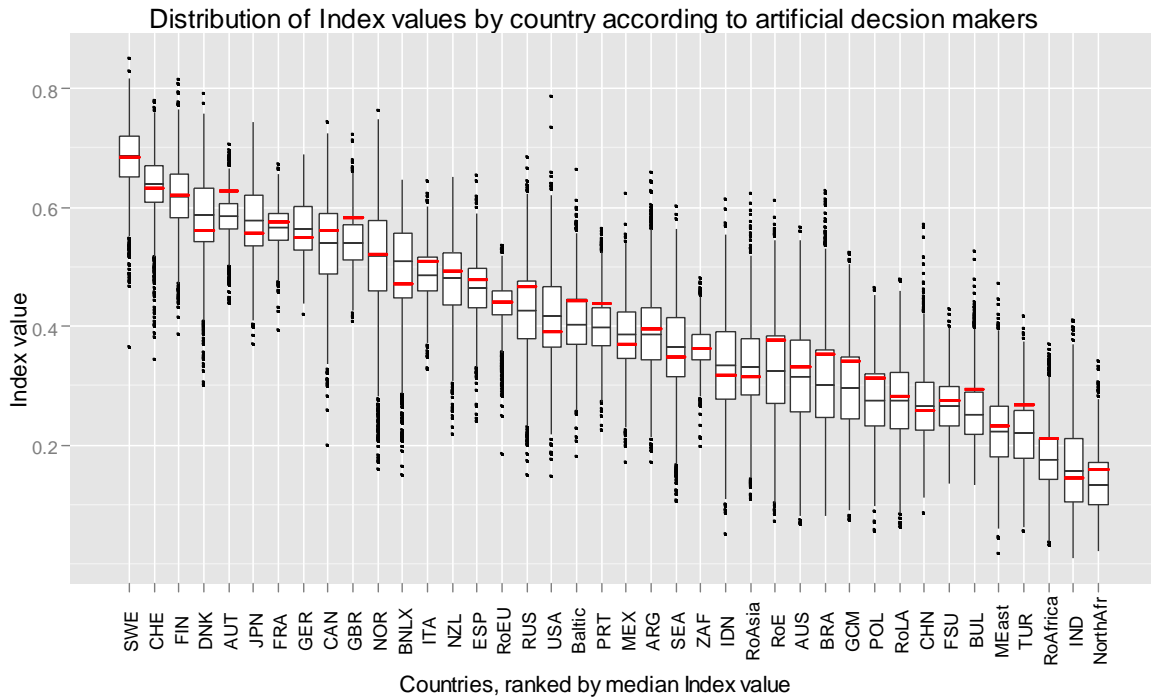
3. Sensitivity analysis

To introduce variability in the determination of subjective weights, a set of 2000 artificial decision makers was generated. Each artificial decision maker provided a random, monotonic set of weights for every node of the tree. Since each set of decision maker weights produces a different FSI value, every country is associated to 2000 alternative FSI values depending on which set of weights was used. These sets were constructed varying within an interval of +/- 10% the "optimism" or the OR-ness the of the original set.

The sensitivity analysis is then conducted with three different approaches: analyzing the distributions of FSI scores, analyzing the "agreement in ranking" and comparing the reference FSI with the median of the distribution of the generated FSI values.

The distribution of FSI scores is depicted in figure 2. The reference values of the index, for each country, are displayed as red lines in the graph.

Figure 2: Sensitivity analysis



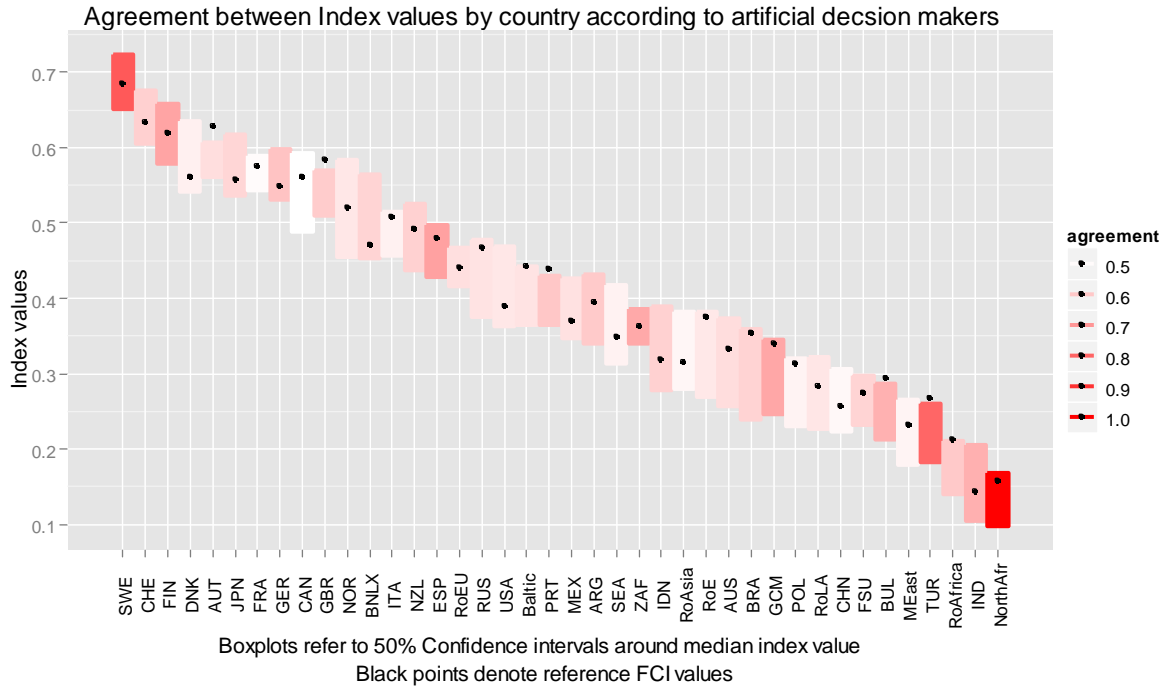
It is immediately appreciable that some countries, such as South Africa (ZAF), Indonesia (IDN), France or the Rest of Europe (RoEU) have a FSI index more robust to changes in decision maker weights than others, such as Norway or Australia (see also Appendix II for details). Nonetheless on a first approximation, it would seem that countries with neighboring positions in the ranking are quite likely to switch positions, since the distribution of FSI values often tend to overlap between consecutive countries.

This perception, however, ignores the fact that decision-maker weights tend to have a somewhat homogeneous effect across countries on the FSI score. It can be shown that, on average, the weights provided by a more “optimistic” decision maker (i.e., one giving set of weights with a higher-than-average OR-ness score to all the nodes in the tree) tend to produce higher FSI score for most countries.

Taking this into account is possible to determine a measure of agreement of decision makers with respect to the median ranking. This amounts to comparing successive countries in the ranking pair wise and counting how many artificial decision makers, out of 2000, provided measures such that the former countries has a higher index than the latter. The measure of agreement among artificial decision makers can then be summarized in a single number for each country. This measure would take a value of 1 if there is perfect agreement among all 2000 artificial decision makers that a given

country is more sustainable than the next one in the ranking; would take a value of 0.5 if only 50% of them agree etc... These measures are displayed in the figure 3 below, where the shade of their box plot indicates the extent of agreement among artificial decision makers.

Figure 3: FSI sensitivity



According to this measure of agreement, some positions, particularly those at the top and at the bottom of the ranking are more stable than others. When individual countries are concerned nearly 82% of artificial decision makers agree that Sweden has a higher FSI value than Switzerland and 79% seem to prefer Turkey to Rest of Africa. These countries are likely to be ranked in the same way with respect to one another irrespective of the measures provided by decision makers. By contrast, only 52% of artificial decision makers provide measures such that Italy is preferred to New Zealand. In general an agreement lower than the 60% concerns the position of France, Canada, South East Asia (SEA), Rest of Asia (RoAsia), Poland, China and Middle East. Therefore this part of the ranking is more brittle. Table 3 finally compares the reference with the simulated FSI values

Table 31: Reference and simulated median FSI values by country for the year 2010.

		FSI 2010		Median simulated FSI 2010		
1	SWE	0.684	=	0.687	SWE	1
2	CHE	0.633	=	0.640	CHE	2
3	AUT	0.629	-2	0.619	FIN	3
4	FIN	0.620	1	0.588	DNK	4
5	GBR	0.583	-5	0.584	AUT	5
6	FRA	0.575	-1	0.577	JPN	6
7	DNK	0.561	3	0.566	FRA	7
8	CAN	0.560	-1	0.564	GER	8
9	JPN	0.557	3	0.541	CAN	9
10	GER	0.550	2	0.540	GBR	10
11	NOR	0.520	=	0.519	NOR	11
12	ITA	0.509	-1	0.509	BNLX	12
13	NZL	0.492	-1	0.485	ITA	13
14	ESP	0.480	-1	0.481	NZL	14
15	BNLX	0.470	3	0.464	ESP	15
16	RUS	0.467	-1	0.441	RoEU	16
17	Baltic	0.443	-2	0.427	RUS	17
18	RoEU	0.440	2	0.417	USA	18
19	PRT	0.438	-1	0.404	Baltic	19
20	ARG	0.395	-2	0.397	PRT	20
21	USA	0.390	3	0.386	MEX	21
22	RoE	0.376	-5	0.386	ARG	22
23	MEX	0.371	2	0.365	SEA	23
24	ZAF	0.362	=	0.363	ZAF	24
25	BRA	0.354	-4	0.334	IDN	25
26	SEA	0.349	3	0.331	RoAsia	26
27	GCM	0.340	-3	0.326	RoE	27
28	AUS	0.332	=	0.315	AUS	28
29	IDN	0.318	4	0.300	BRA	29
30	RoAsia	0.316	4	0.295	GCM	30
31	POL	0.314	=	0.276	POL	31
32	BUL	0.294	-3	0.275	RoLA	32
33	RoLA	0.283	1	0.265	CHN	33
34	FSU	0.274	=	0.265	FSU	34
35	TUR	0.267	-2	0.250	BUL	35
36	CHN	0.257	3	0.222	MEast	36
37	MEast	0.231	1	0.221	TUR	37
38	RoAfrica	0.212	=	0.175	RoAfrica	38
39	NorthAfr	0.159	-1	0.155	IND	39
40	IND	0.144	1	0.133	NorthAfr	40

It should be noted that there is generally close agreement between the reference FSI values and the central interval of the simulated FSI distribution. For most countries, the reference value of the FSI tends to be relatively close to the median value of the simulation.

It is therefore misleading to assume that overlapping distributions of FSI values necessarily imply high variability in ranking positions, although these do vary to some extent in practice.

4. Conclusions

Few issues like the construction and use of composite and aggregate sustainability indicators raise criticisms and debate. The reason is that any step of the process - the choice of indicators to include, of the “weights” to assign to each, the aggregation procedure - are subjectivity prone, no matter the effort made. When this is the case, many criticisms can be perfectly legitimate and correct.

As shown by the present exercise, it has to be accepted that, notwithstanding the technical feasibility, it is neither possible to un-controversially summarize sustainability in just one figure, nor that subjectivity can be ruled out of composite indicators. In fact, we have shown that the country ranking proposed by the complex FSI, demonstrates a rather good degree of robustness, especially concerning the positions at the top and at the bottom. Nonetheless this robustness is far from offering full objectivity and invariance.

Nonetheless, there are very good reasons in favor of the use of composite indicators. As shown by IN-STREAM research, they can be invaluable communication devices to explicit the preference structure and value judgments originating a given synthetic sustainability assessment. They can also offer the opportunity to investigate in depth if and how this assessment can change when those preferences and values change. In this respect, sensitivity analyses, coupled with the transparency of construction are key features to apply to composite indicators. All the information gathered can then be very interesting for policy decision makers and, potentially, are as, if not, more important than the synthesis provided.

References

- Best, A, Bosello, F. and F. Ciampalini (2009), "Full set of baseline indicators", INSTREAM D4.1.
- Bosello, F., Ciampalini, F. and F. Eboli (2010), "Report on sectoral effects for different sustainability indicators", INSTREAM D6.3
- Bosello, F., Portale, E., Campagnolo, E., Eboli, F. and R. Parrado (2011), "Comparison of informative capacity of aggregated sustainability index with disaggregated indicators", INSTREAM D6.5.
- De Waegenaere, A. & Wakker, P., (2001). "Nonmonotonic Choquet integrals," *Journal of Mathematical Economics*, Elsevier, vol. 36(1), pages 45-60
- Murofushi T., Sugeno M., Machida M. (1994), Non-monotonic fuzzy measures and Choquet integral, *Fuzzy Sets and Systems*, 64, 73-86, 1994.
- Nardo et al., 2008, Handbook on constructing composite indicators: methodology and user guide, OECD Statistic Working paper, 2008