



NOTA DI LAVORO

16.2013

**Constructing the FEEM
Sustainability Index:
A Choquet-integral
Application**

By **Caterina Cruciani**, Fondazione Eni
Enrico Mattei, Ca' Foscari University

Silvio Giove, Fondazione Eni Enrico
Mattei, Ca' Foscari University

Mehmet Pinar, Edge Hill University

Matteo Sostero, Fondazione Eni Enrico
Mattei, Ca' Foscari University

Climate Change and Sustainable Development

Series Editor: Carlo Carraro

Constructing the FEEM Sustainability Index:

A Choquet-integral Application

By Caterina Cruciani, Fondazione Eni Enrico Mattei, Ca' Foscari University

Silvio Giove, Fondazione Eni Enrico Mattei, Ca' Foscari University

Mehmet Pinar, Edge Hill University

Matteo Sostero, Fondazione Eni Enrico Mattei, Ca' Foscari University

Summary

This paper presents an application of a multi-attribute aggregation methodology to the construction of a sustainability index. Sustainability is a multi-faceted issue, in which synergies or conflicts may arise among the different components, thus making it a complex concept to which multi attribute methods can be employed. This paper addresses the development of the FEEM Sustainability Index (FEEM SI), a composite index including 19 different indicators grouped in the three classical pillars of sustainability – economic, social and environmental. We present the relevance of multi-attribute aggregation methodologies when dealing with such complex concepts and provide an aggregation methodology used for this case study, the Choquet-integral aggregation. First, we normalize each sustainability indicator with the use of a benchmarking procedure with a smooth target of sustainability. Furthermore, an ad-hoc questionnaire implemented to assess the importance of each sustainability indicator and their interaction with other indicators through expert elicitation. After normalizing each sustainability indicator and computing consensus importance of each sustainability indicator and their interactions for the Choquet-integral aggregation procedure, the overall sustainability index, the FEEM SI is calculated. This paper also conducts robustness analysis and discusses the main implications of the aggregation methodology used.

Keywords: Sustainability Indices, Composite Indicators, Choquet Integral, Multiattribute Value Theory

JEL Classification: C43, C44, Q01, Q56

This paper belongs to the research activities of the Climate Change and Sustainable Development Programme of the Fondazione Eni Enrico Mattei. The authors are grateful to Carlo Carraro, Fabio Eboli, Francesco Bosello, Ramiro Parrado, Elisa Lanzi and all participants at seminars in Venice and Milan in which a preliminary version was presented. The usual disclaimer applies.

Address for correspondence:

Mehmet Pinar
Business School
Edge Hill University
St Helens Road, Ormskirk
Lancashire, L39 4QP
United Kingdom
E-mail: mehmet.pinar@edgehill.ac.uk

The opinions expressed in this paper do not necessarily reflect the position of
Fondazione Eni Enrico Mattei

Corso Magenta, 63, 20123 Milano (I), web site: www.feem.it, e-mail: working.papers@feem.it

Constructing the FEEM Sustainability Index: a Choquet-integral application

Caterina Cruciani^{*°}, *Silvio Giove*^{*°}, *Mehmet Pinar*[#], *Matteo Sostero*^{*°}

^{*}Fondazione Eni Enrico Mattei, [°]Ca' Foscari University of Venice, [#] Edge Hill University

Abstract

This paper presents an application of a multi-attribute aggregation methodology to the construction of a sustainability index. Sustainability is a multi-faceted issue, in which synergies or conflicts may arise among the different components, thus making it a complex concept to which multi attribute methods can be employed. This paper addresses the development of the FEEM Sustainability Index (FEEM SI), a composite index including 19 different indicators grouped in the three classical pillars of sustainability – economic, social and environmental. We present the relevance of multi-attribute aggregation methodologies when dealing with such complex concepts and provide an aggregation methodology used for this case study, the Choquet-integral aggregation. First, we normalize each sustainability indicator with the use of a benchmarking procedure with a smooth target of sustainability. Furthermore, an ad-hoc questionnaire implemented to assess the importance of each sustainability indicator and their interaction with other indicators through expert elicitation. After normalizing each sustainability indicator and computing consensus importance of each sustainability indicator and their interactions for the Choquet-integral aggregation procedure, the overall sustainability index, the FEEM SI is calculated. This paper also conducts robustness analysis and discusses the main implications of the aggregation methodology used.

Keywords: Sustainability indices; Composite indicators; Choquet integral; Multiattribute value theory

JEL classification: C43; C44; Q01; Q56

This draft: January 2013

This paper belongs to the research activities of the Climate Change and Sustainable Development Programme of the Fondazione ENI Enrico Mattei. The authors are grateful to Carlo Carraro, Fabio Eboli, Francesco Bosello, Ramiro Parrado, Elisa Lanzi and all participants at seminars in Venice and Milan in which a preliminary version was presented. The usual disclaimer applies.

Corresponding author: Mehmet Pinar

Business School, Edge Hill University, St Helens Road, Ormskirk, Lancashire, L39 4QP, United Kingdom; e-mail:

mehmet.pinar@edgehill.ac.uk

1. Introduction

Sustainability is somewhat an elusive concept: although its main message is widely understood, it is quite hard to come by two identical descriptions of it, when it comes to spelling out its different components. The most used definition of sustainable development is given in the Brundtland report as “Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs” (WCED, 1987). Achieving a sustainable development has been one of the major concerns of modern societies, which have long been interested in understanding and governing the multi-faceted issue of development, thus making a comprehensive assessment of sustainability crucial to measure progress, identify areas to be addressed and evaluate the outcome of implemented policies. The need to find ways to measure sustainability translated into a plethora of approaches and sustainability indicators that have been aggregated in different ways to obtain composite indices. For a methodological review on the sustainability, see Bossell (1999), OECD JRC (2008), and Singh et al. (2009); and for list of sustainability indicators refer to the EU core set of indicators (EEA, 2005), and the UN Commission on Sustainable Development (2005).

This paper focuses on the methodological issues regarding the construction of a composite index of sustainability, which is an area that has been gaining interest in empirical literature due to its high policy potential. In fact, a composite index allows for a quick assessment of sustainability performance across different countries and at different times. Moreover, sustainability indices convey a straightforward message to stakeholders and policy makers, and are able to highlight best practices and weaknesses of sustainability strategies (Ness *et al.* 2007).

Such a sustainability index needs to be constructed very carefully using a procedure as transparent as possible, in order to gain trust in the policy arena. Moreover, sustainability is characterized by many different aspect that are somewhat linked one to the other, which rules out the possibility of using simple aggregation techniques, which risk to lose too much information regarding interactions across indicators.

The policy potential of an aggregate measure coupled with the complexities behind the relations across the indicators that define sustainability make it an area to employ a non-linear aggregation methodology, which is able to target specifically the interactions across the different components of sustainable development.

The contribution of this paper is threefold. First part is the normalization procedure. Since the sustainability indicators are characterised with different measurement units and therefore requires a normalisation step before aggregating all indicators into a single index. indicators composing the FEEM SI are normalised according to the *benchmarking* methodology. Since the purpose of creating a sustainability index is not only to identify best and worst practices, but also to give an appraisal of the relative distance to the sustainable target, the FEEM SI indicators are normalised according to a benchmark function, which passes through five reference levels (i.e., five stages of sustainability levels from the extremely unsustainable to fully sustainable levels for each sustainability indicator respectively having normalized values of 0 and 1). The use of a *benchmarking procedure* is very appropriate especially in the case of those indicators for which an agreed target of some kind (at EU or global level, for instance) exists. This method allows comparison through time and across countries, whilst supplying a policy-based normalisation, which is particularly suitable for the construction of the FEEM Sustainability Index. We offer further details in the section 3 on the normalization procedure.

The main contribution is to construct a composite sustainability index by applying a non-linear aggregation methodology (i.e., the Choquet integral) which accounts for the interactions among sustainability indicators. Singh et al. (2009) summarizes forty one sustainability indicators used in the literature and majority of those indices are either aggregated through equal weight assignment (e.g., Environmental Sustainability Index, Human Development Index, Sustainability Performance Index, etc.) or weights given by experts (e.g., Index of Environmental Friendliness) to each sustainability indicator. However, none of those indices allow to capture the inter-relations among different sustainability indicators. In other words, those aggregation methodologies do not account for synergies or redundancies when indicators are aggregated (e.g., see Panayotou, 1993; Grossman and Krueger, 1993; Selden and Song, 1994; Arrow *et al.*, 2004; Millennium Ecosystems Assessment, 2005; World Bank, 2010; Agliardi, 2011; Arrow *et al.*, 2012 for discussion on the relationship between economic growth and environmental quality). In the construction of the FEEM Sustainability Index (FEEM SI hereafter), we are able to address specifically the inter-relations across indicators,

thus overcoming the limitations of other aggregation methodologies.

Another contribution of the current paper is the expert elicitation and the derivation of weights for each sustainability indicator and their coalitions among each other. An ad-hoc questionnaire is implemented to assess the importance of sustainability indicators through expert elicitation. As the Brundtland report suggests, sustainable development “meets the needs of the present without compromising the ability of future generations”, therefore, the relative importance of and interaction among different indicators has to be evaluated through expert elicitation. However, the aggregated index strongly depends on the subjective relative importance of coalitions between different sustainability indicators, which may be different for each expert. Therefore, we derive a consensus measures on sustainability indicators from many expert elicitations by using a metric distance (i.e., if the evaluation of an expert is in agreement with other experts, then this expert’s valuation gets higher weight. Thus, if an expert’s valuation of sustainability indicators is extremely different than other experts, then relatively lower weight is assigned to this type of expert valuation). After obtaining the consensus measures, we employ Choquet integral to aggregate the normalized sustainability indicators to obtain the final FEEM SI outcome for the countries (macro-regions) in our analysis.

The remainder of the paper is as follows. In section 2, we review the theoretical contributions of the Choquet integral as an aggregation methodology in the sustainability context pointing out useful properties and features. In section 3, we provide the stages of aggregation framework for the FEEM SI. We first introduce the sustainability indicators considered in this study and where those indicators are located in the aggregation tree and the details how each indicator is measured. There needs a two procedure before aggregation of sustainability indicators, that is normalization of each indicator and weight assignment to each indicator. We first offer the normalization procedure for each indicator. Secondly, we offer the construction of an ad-hoc questionnaire to assess the importance of sustainability indicators through expert elicitation. We provide the derivation of the representative weights of sustainability indicators and offer the main characteristics of expert elicitations. Given the normalized indicators and representative weights for each indicator, normalized indicators aggregated through Choquet integral aggregation procedure and in section 4, we provide the FEEM SI scores for countries and macro-regions. We also conduct comparisons of sustainability outcomes between the Choquet integral and alternative aggregation operators. Furthermore, we obtain FEEM SI scores by simulating different expert elicitations from the real expert elicitations and use those in the aggregation procedure to analyze whether the results obtained through representative measures are robust. Finally, section 5 concludes and discusses the possible extension of the current study. Detailed mathematical formulations used for aggregation are gathered in an Appendix.

2. Multi attribute aggregation

Sustainability evaluation is a multi-attribute problem, since it is characterized by many different components that interact with each other. Thus, the use of a multi-criterion framework for composite indicators in general and for sustainability and well-being indexes in particular is relevant and desirable (Munda, 1997; Ulengin *et al.*, 2001; Munda, 2005). In this literature, several approaches exist to deal with multi attribute problems, each characterised by specific mathematical properties, which have very different implications. In this section, we briefly review possible aggregation options and provide elements to understand why sustainability cannot be fully addressed by some of them and offer how Choquet integral is a better choice for the aggregation of sustainability indicators.

Vincke (1999) classifies the main aggregation approaches into three categories: multi-attribute value theory (MAVT), outranking approaches, and interactive approaches. Among the three, MAVT-based techniques have been more extensively used in multi attribute problems given the great diversity of methods fall under this broad category of approaches. In MAVT, starting from the criteria evaluation a single aggregated score has to be computed, expressed in a finite scale such as $[0,1]$ where 0 represents “very bad” outcome, 1 represents “very good” outcome for any indicator which is increasing in a social ‘good’. In general, the MAVT methods use an aggregation algorithm to compute a score for each alternative (Klement *et al.*, 2000). A MAVT method is characterized by two subsequent phases; in the first one all the criteria are normalized in a common scale, usually the $[0,1]$ scale, in order to allow direct comparisons. In the second phase, the normalized values are aggregated using a suitable function, i.e., an aggregation operator. In particular, monotonicity is a strong requirement for the aggregation operator since it implies that, *ceteris paribus*, “more” is preferred to “less”.

A broadly used MAVT-based aggregation technique is the weighted-average, which relies on the (simple) arithmetic weighted average of the (normalized) indicator values (e.g., Human Development Index and Multidimensional Poverty Index). The most common case is the one where the weights are the same for all the indicators (i.e., equal weights are given to each indicator prior to aggregation). Despite the fact that this method is simple and intuitive, the linearity of the aggregation function implies constant substitutability among the criteria which is not a very reasonable assumption, given the nature of the sustainability indicators and could lead to double counting (Nardo *et al.*, 2005). Nevertheless, the general tendency is to use the equally-weighted indices due to the lack of empirical proof on the relative importance of different sustainability indicators and is considered the most transparent way of producing aggregate indices (Environmental Sustainability Index, ESI, 2005). However, weighted average should be taken with particular care since it assumes no interactions among the indicators, which is an unlikely and quite rare situation. In fact, the substitutability very often fails to be satisfied, which means that the compensative assumption (i.e., the preferential independent axiom) is not satisfied (Marichal and Roubens, 2000) therefore, weighted-average cannot be implemented for the aggregation procedure.¹

As discussed above, specific properties of an aggregation operator are required to aggregate the sustainability indicators. In particular, aggregation operators should be easily parameterized and tuned by the decision maker, and not necessarily implying compensability, while conversely they need to be mathematically well founded and characterized. In this respect, non-additive measures (NAM) can be a solution, as they satisfy above-mentioned formal requirements. In fact, a non-additive measure (NAM) operator satisfies many theoretical requirements, and, at the same time, it is sufficiently general to cover flexible preference structures of the decision maker (DM). With NAM, one can allow for interactions among the sustainability indicators. Moreover, any preference of the decision maker can also be introduced such as the *pessimistic* or *optimistic* attitude, which indicate respectively that the satisfaction degree is high only if *all* the indicators are satisfied, or if *at least one* of the indicators is satisfied.² The linearity assumed with the weighted average can be relaxed with the NAM where the “weight” of a coalition of sub-criteria can be greater or less than the sum of the weights of each of the sub-criterion belonging to the coalition itself. In other words, the main idea consists into assigning a weight to *every possible* subset of the criteria which refer to the considered indicators. Therefore, NAM methods are nothing but an extension of the weighted average, in which instead of assigning a weight to each indicator, one can assign a weight to each possible coalition of them. However, the computation becomes exponentially more complex as the number of indicators increases. In fact, if n is the number of the indicators, a NAM requires the specification of 2^n possible coalitions among indicators, i.e., the number of all the subsets of the n indicators, while the weighted average only requires allocation of n weights.³ Nevertheless, the complexity of having too many subsets can be overcome by limiting the number of indicators in each specification to a small number.

Among the possible NAM, the Choquet integral satisfies many properties, which make it an effective tool in the analysis of multi attribute problems such as sustainability. Varying the value of the measures (i.e., the weights assigned to all possible subsets), the Choquet integral can be reduced to min or max operators (i.e., the indicators are perfect complements or perfect substitutes respectively), or ordered weighted average, or a mixture of them.⁴ Thus, according to the features of the DM preferences one can tune the measure values and obtain several interesting sub-cases. 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 the weighted average. Conversely, if for a coalition its weight is *inferior* to such a sum, a *redundant* interaction exists among the included criteria (i.e., the indicators are substitutes of some level), while if it is *greater* than the sum, a *synergic* interaction exists (i.e., the indicators in question are complements of some level). Meyer and Pothiere (2011) recently apply the Choquet integral-based multi-attribute value in well-being context and find the existence of complementarities and redundancies between standard of living dimensions, and suggest a strong heterogeneity of individual preferences. In this paper, however, we extend the Choquet integral-based multi-attribute to sustainability context and obtain sustainability outcomes and rankings of countries where individual preferences of sustainability are accounted.

¹ The compensative assumption is rarely tested in practical applications, but missing this check can induce a strong distortion in the decisional process.

² Note that both optimistic and pessimistic cases cannot be implemented with the use of weighted average aggregation method.

³ To be exact, there are $2^n - 2$ required parameters since the *border* conditions are already predetermined in which the empty set is null and the universal set is one.

⁴ The formal description of Choquet integral is offered in Appendix A.

3. Conceptualizing sustainability: the FEEM Sustainability Index

In this section we introduce the aggregation methodology developed for the FEEM SI, an aggregate sustainability index characterized by 19 indicators belonging to the three pillars of sustainability (i.e., economic, social and environmental). The indicators are constructed within a recursive-dynamic general equilibrium model ICES-SI (Carraro et al., 2012)⁵ which produces future projections of all indicators in the time frame 2011-2020 that can be used in comparative static policy analysis (see Carraro *et al.*, 2012 for further details on the construction of the FEEM SI) and provides useful policy implications (Bohringer and Loschel, 2006).

Given the set of sustainability indicators, there are two initial steps need to be taken before the aggregation procedure. Since each indicator not only is measured in different metrics but also different levels for each indicator represent different sustainability outcomes, we first normalize each indicator. The indicators are normalized using a policy-oriented benchmarking technique developed ad hoc for the FEEM SI before proceeding to the aggregation stage.⁶ After having each indicator into a common scale, 0 to 1, representing the sustainability levels of countries in which 0 being the extremely unsustainable condition and 1 being a fully sustainable level.

Second stage before aggregation is the expert elicitation. In other words, in order to aggregate normalized indicators, the importance of all possible combinations of sustainability indicators is required. Therefore, a value of importance for each subset of indicators needs to be obtained. However, as explained in the previous section, the set of questions asked to assess the importance of all possible combinations increase exponentially with the number of sustainability indicators.⁷ Therefore, sustainability indicators are allocated under a decision tree structure in which each node contains either two or three indicators to make it feasible for expert elicitation. Moreover, allocation of indicators into different theme-based pillars not only allows one to assess the final sustainability levels but also the theme-based sustainability levels (i.e., economic, social and environmental sustainability levels). For example, UN Commission on Sustainable Development (UN CSD 2006) and European Union Sustainable Development Strategy 2005 (EU SDS 2005) offer a theme based indicator set for assessing the sustainability levels. Furthermore, Environmental Performance Index (EPI) includes 25 performance indicators tracked across ten policy categories covering both environmental public health and ecosystem vitality assessing the closeness of a country to environmental policy goals (EPI, 2010). Therefore, we follow the similar approach to allocate the indicators under different theme categories to assess the sustainability level in three pillars that also make the expert elicitation feasible.

The aggregation methodology prepared for the FEEM SI builds on the capacities of the multi attribute aggregation methodology, partially following Despic and Simonovic (2000), who provides a three-step procedure to deal with the evaluation of complex concepts that are naturally composed of numerous sub-elements.⁸ Their contribution is to reduce the complexity behind the analysis of a system by proceeding with a “hierarchical analysis of its less complex components” (Despic and Simonovic, 2000). Sustainability certainly qualifies as a complex issue, thus making it a viable case to apply a methodology along the same lines as Despic and Simonovic (2000). The methodology used for the FEEM SI differs in a number of respects from the original work of Despic and Simonovic (2000), but retains their main idea and makes use of a hierarchical analysis.

The decomposition of sustainability into a set of indicators is necessary to construct a decision tree representing this hierarchical decomposition. This index deals with global sustainability, thus the subjects of this sustainability analysis are states and macro regions. The definition of the components of the FEEM SI is thus defined thorough a literature review including the work of international institutions involved in sustainability analysis.⁹ The indicators selected for the FEEM SI have been organised into a decision tree, in which partial aggregation take place at different levels, leading to the hierarchical decomposition of Figure 1.¹⁰

5 ICES-SI model is an extended version of ICES model, see Eboli et al., 2010. See also Carraro *et al.* (2012), and Annex II section of Carraro *et al.*, 2011 for the detailed construction of the ICES-SI model.

6 Each indicator is converted into a common scale between 0 and 1 which allows a full comparison among indicators, see Nardo *et al.* 2005

7 The total number of questions is $2^n - 2$ where n is the number of sustainability indicators. For example, the number of questions asked to experts to evaluate the all possible combination among the sustainability indicators would be 2, 6, 14 when n=2, n=3, and n=4 respectively.

8 The three steps proposed by Despic and Simonovic (2000) are decomposition of the system, evaluation of the lower components and aggregation.

9 see Section 1.2 of Carraro *et al.* (2011) for detailed indicator selection process, their descriptions and the affiliation area of each indicator to each sustainable development dimension.

10 The details of each indicator are given in Appendix Table A.1. Note that the sustainability indicators are given are in numbers. The nodes, which are not indicators themselves, but the aggregated outcomes of indicators, are in bold font. For example, biodiversity level of a country or macro region is an aggregated outcome of plant and animal biodiversity of that given region.

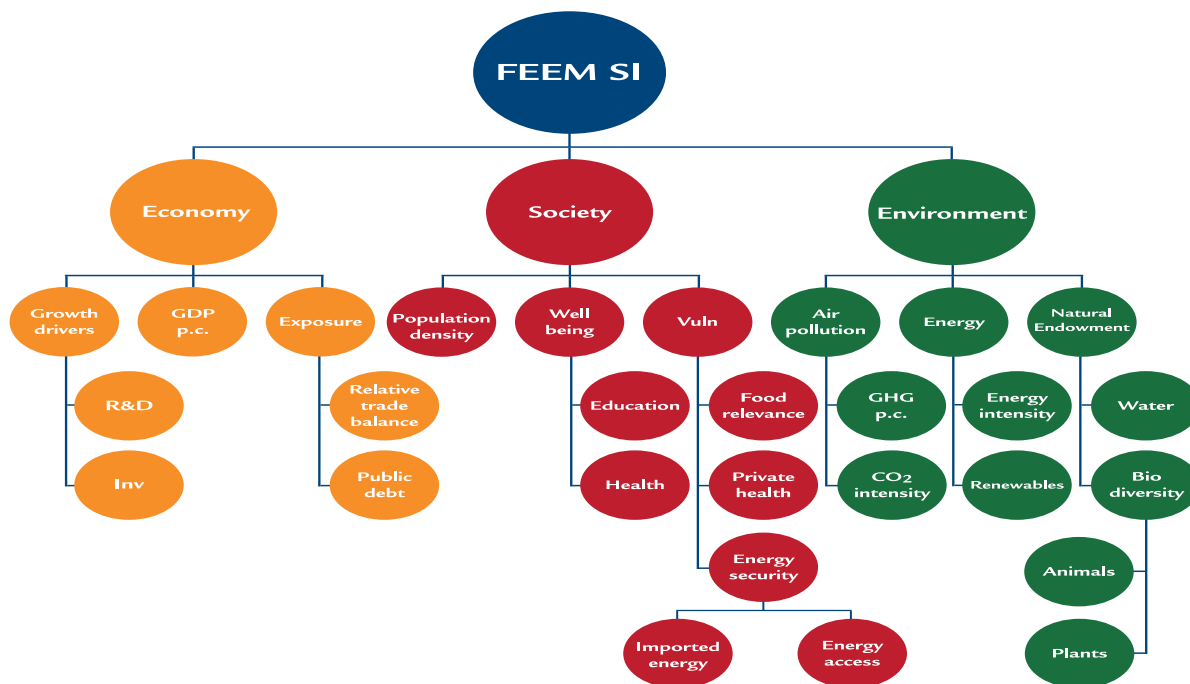


Figure 1: FEEM SI Aggregation tree

The decision tree should be read from bottom (leaves) to top (final node) and is characterised by three successive decomposition levels. The tree respects the three main pillar structure which is standard in most sustainability studies (see e.g., The United Nations Commission for Sustainable Development (UN CSD); Global Reporting Initiative (GRI) framework; Krajnc and Glavic, 2005), with the final node producing the aggregate index.¹¹

In the next sections, we offer the initial steps, a normalization procedure and expert elicitation, before the aggregation procedure. Finally combining these initial steps allows us to use Choquet integral to assess the sustainability levels of countries and macro regions.

3.1 Normalization procedure

Like any other dataset, the one created for the FEEM SI is characterised by several measurement units and needs to go through a normalisation step before aggregating all indicators into a single index. According to the OECD's *Handbook on constructing composite indicators* (OECD, 2008)¹², "normalization is required prior to any data aggregation as the indicators in a data set often have different measurement units". Several normalization techniques exist in literature. The best choice depends on the specific theoretical framework adopted for the computation of indicators.

The Handbook reports eight different normalisation procedures commonly used in literature. The simplest method is *ranking*: country performances are evaluated only according to the variations of their relative positions. With *standardization* and *re-scaling*, instead, indicators are converted into a common scale. In the case of *standardization* the scale is obtained by subtracting the mean and dividing by the standard deviation, whereas in the case of *re-scaling* it is a 0-1 scale. Indicators can also be ranked according to the *distance* to a reference measures, through *categorical scales*, different for each indicator, or by applying *benchmarks* which assign normalised values (normally in a 0-1 interval) according to whether the indicator values are above/below/around that threshold value. Finally, indicators can be normalised by taking the *percentage differences* over consecutive years or by applying methods for *cyclical indicators*.

Not all of these methods are relevant to the FEEM SI dataset; therefore, the focus will be only on some of them. For

¹¹ Both the UN CSD analyzes four main dimensions in their framework including the institutional (governance) dimension and future development of the FEEM SI is to include the governance dimension into the analysis of sustainability assessment.

¹² OECD (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, <http://composite-indicators.jrc.ec.europa.eu/Handbook.htm>

instance, *ranking* or using *categorical scales* would not allow keeping the quantitative aspect of indicators, and *percentage differences* do not clarify whether the variation is sustainable or not. *Standardisation* using mean and standard deviation produces normalised indicators that are not comparable across different years, but only across countries in one specific year.

Methods better suited for the construction of the FEEM SI are *re-scaling* in a 0-1 range, or applying a series of *benchmarks* that represent different levels of the indicators. The usefulness of the *re-scaling* procedure is that it translates into a widening effect of the normalised indicators whose original values were extremely close, thereby enhancing even small differences. This characteristic is very important especially for some indicators, where groups of countries tend to have very similar values. The drawback of such a methodology is that it is extremely sensitive to extreme values, which tend to distort the normalised values.

At a technical level, the values of each indicator for all countries are translated into the 0-1 scale, where 0 applies to the minimum value, 1 to the maximum value and converting the intermediate values through the following formula (OECD, 2008; previous footnote):

$$I_{qc}^t = \frac{x_{qc}^t - \min(x_q^{t_0})}{\max(x_q^{t_0}) - \min(x_q^{t_0})}$$

in which x_{qc}^t indicates the value of an indicator q for the country c at time t .

The maximum and minimum values used for this type of normalization are the lowest and highest values of a specific indicator in one year across countries. These values do not correspond necessarily to the best and worst possible values of that indicator in absolute terms, i.e. these values do not represent a value judgement. Furthermore, the reference minimum and maximum levels change over the years, making it difficult to compare sustainability changes over the years. On the other hand, if the difference between the maximum and minimum values is relatively high for one indicator and relatively low for another one, then the effect of the former on the composite index becomes somewhat lower than that of the latter (see Ravallion 1997; Noorbakhsh, 1998; Ravallion 2010 for the discussion for the Human Development Index). Finally, there exists irrational rank reversals between years even though the achievements of countries stayed the same but the only the maximum achievement being changed (Billaut *et al.*, 2010). However, the normalisation approach taken FEEM SI avoids above-mentioned problems since the benchmarks are a priori determined and do not depend on the data set.

A different reasoning holds for the other normalisation method mentioned above, which starts by individuating a best practise or target level to be used as a benchmark. The use of a *benchmarking procedure* is very appropriate especially in the case of those indicators for which an agreed target (at EU or global level, for instance) exists. The benchmarking procedure normally assigns only two values, 1 and 0, according to the correspondence to a chosen reference level. This method allows comparison through time and across countries, whilst supplying a policy-based normalisation, which is particularly suitable for the construction of the FEEM Sustainability Index. This is the reason why the indicators composing the FEEM SI are normalised according to the *benchmarking* methodology. Since the purpose of creating a sustainability index is not only to identify best and worst practices, but also to give an appraisal of the relative distance to the sustainable target, the FEEM SI indicators are normalised according to a benchmark function, which passes through five reference levels. Each of these levels corresponds to a normalized value comprised between 0 and 1 as follows in Table 1.

Normalized Value	Sustainability Level
0	Extremely unsustainable
0.25	Still not sustainable but not as severely as in the previous case
0.50	Discrete level of sustainability, but still far from target
0.75	Satisfactory level of sustainability, yet not on target
1	Fully sustainable

The normalized values correspond to specific levels of the indicator in the original measurement unit. Such levels are defined according to reliable and authoritative literature and international legislation sources, in order to increase the acceptability of the methodology. Whenever possible, the objectives outlined in the EU Sustainable Development Strategy or in the Europe 2020 (a follow up of the Lisbon Strategy) have been used to define one or more level of the benchmarking function, since the European Union legislation has already been considered a sort of best practice in the selection of the indicators outlined in chapter one. In all other cases, broader EU policy objectives and international standards from established institutions such as the Organisation for Economic Co-operation and Development, World Bank, United Nations and International Monetary Fund have been taken as primary source of information.¹³

3.2 Features of the questionnaire and expert elicitation

The second stage of the procedure requires the evaluation of the subcomponents at each one of the three levels of the decomposition, starting from the bottom node of the aggregation tree. Differently from other procedures, where a relative weight is defined for each indicator with respect to the others, the decision tree also requires to attribute weights to the coalitions of indicators at each node.

The procedure is based on the definition of a fuzzy measure, but instead of evaluating all the possible combinations among all possible levels of the indicators at each node, the procedure inspired from Despic and Simonovic (2000) allow to focus on some specific combinations of given levels of the indicators. Indicators can take in theory a wide range of values, but in this procedure, only two extreme qualitative values are identified and labelled “best” and “worst”. All the indicators of the FEEM SI are quantitative in nature, due to the features of the modelling environment in which they have been developed, yet the aggregation methodology has been constructed using “qualitative” evaluations in order to make the procedure more comprehensible.

Contrary to what Despic and Simonovic (2000) suggests, no specific numerical definitions have been given for these two levels for each indicator; given the broad scope of the exercise –global sustainability- and the relevance of including in the expert group used to evaluate the questionnaire a wide variety of backgrounds, a more neutral setting has been preferred. As a later section will detail, experts in the questionnaire have been asked to make reference to their own ideal “best” and “worst” for each indicator. Moreover, avoiding numerical evaluations of best and worst levels rules out the bias that may be caused by respondents’ judgement.

In this case study, the construction of an aggregate sustainability index requires selecting a procedure that allows for different relations across the different indicators. For instance, a sustainable policy requires economic development not to be obtained at the price of an increase in environmental pollution, while a better healthcare policy may have spill over effects on other welfare programmes. In order to account for all these aspects any aggregation methodology that is additive (e.g., weighted average) in nature cannot be chosen: a good performance in some indicators cannot compensate for a poor one in others. Thus, a methodology that allows for non-compensative aggregation, in which the different level of interaction across indicators is allowed.

In order to obtain the weights that are necessary to compute the fuzzy measure for the aggregation, a simple paper-based questionnaire has been prepared, which includes a decision matrix for each one of the twelve decomposition nodes of the aggregation tree. The questionnaire can be thought of as a list of the possible scenarios with two defined qualitative levels of the indicators, i.e. all the combinations of BEST and WORST values. If n is the number of sub-components of the considered node, the decision matrix will then have 2^n rows, thus requiring the same number of evaluations by the respondent.

The values of the measure are obtained from the elaboration of the questionnaire developed by Despic and Simonovic (2000). The questionnaire is a list of some possible scenarios, i.e. all the combinations of BEST and WORST values. Again 2^n is the number of possible questions, and the respondent needs to fulfil the last column with a number in between 0 (worst case) and 100 (best case), respecting the monotonicity constraints, i.e., the measure of a set (coalition of criteria) cannot be less than the measures of the subset forming a partition of it.

¹³ The space limitation and the scope of this current study preclude us to present the benchmarks for each sustainability indicator since the aim of the current paper is to offer normalization, expert elicitation and aggregation techniques, which allow one to assess the sustainability. See the Section 3.3 of Carraro *et al.* (2011) for the details of the benchmark selection process for each indicator.

Economic	Social	Environmental	Weights
Worst	Worst	Worst	0
Best	Worst	Worst	20
Worst	Best	Worst	50
Worst	Worst	Best	30
Best	Best	Worst	$X \geq 50$
Best	Worst	Best	$X \geq 30$
Worst	Best	Best	$X \geq 50$
Best	Best	Best	100

Figure 2: Construction of Indicator-Coalition Matrix

The respondents have to provide a numerical evaluation for each row of the decision matrix, which is provided for all twelve decomposition nodes, choosing a value between 0 and 100 for each row, except for the first and the last (where indicators are respectively all “worst” and all “best”) which are given 0 and 100 by default. Moreover, the weights given at each row of every matrix need to respect the monotonicity criterion introduced in the previous section. This implies that, if a combination where only one indicator is “best” is given a certain weight x , all combinations including that indicator in the “best” case should be given a weight at least equal to x , as the Figure 2 illustrates an example. Once the questionnaire is complete, the numerical evaluations are imputed to a software that computes the fuzzy measures and combines them with the indicator value to provide sustainability country rankings for each year of the analysis.

3.3 Decision maker behaviour

One important feature of this non-additive methodology is that in principle it allows to be extremely flexible in the definition of the degree of compensability across indicators. The level of compensation implicit in the weights assigned by the experts can be quantified by means of two complementary indices, the ANDNESS and ORNESS indices, often referred to as characterising an optimistic and a pessimistic decision maker. These indices sum to one, with each of them being given a score between 0 and 1. An ANDNESS degree close to one indicates that the expert tends to be *non-compensative*, meaning that he/she would not accept that a good performance in one sub-criteria compensates for a lower one in another. On the contrary, an ORNESS degree close to one indicates that the decision maker is satisfied even if only one sub-criterion is at “best” level. Given the nature of the problem at hand, it seems more likely that decision makers evaluating the hierarchical structure of the FEEM SI tree should be more inclined towards andness, as sustainability implicitly requires a balanced development across its different components.

Moreover, interaction among criteria (i.e., interaction indices) is also an important characteristic of Choquet integral that adds valuable information to the sustainability concept. ANDNESS and ORNESS indices offer whether a decision maker’s characteristic is pessimistic or optimistic at a given node level however, one can also analyze the interaction between two indicators at a given node. Suppose that the two indicators, i and j , are competitive (i.e. substitute) to each other, then the marginal contribution of indicator i to every combination of criteria that contains indicator j should be strictly less than the marginal contribution of i to the same coalition when j is excluded (i.e. negative interaction index for the indicators i and j). Reverse is true for the case when indicators i and j are complementary (i.e., positive interaction index for the indicators i and j). The interaction index is between -1 and 1, representing perfectly competitive (i.e. substitute) and perfectly complementary behaviour respectively. Given that, as suggested, the FEEM SI tree should be more inclined towards andness, one should also expect decision makers to have more complementary-oriented behaviour among the indicators. For example, a country having good economic and environmental conditions should be evaluated more in sustainability terms than a country having only one dimension at its best level.

The next section will present the questionnaire used to assess the sustainability concept; in particular, we asked participants to evaluate sustainability indicators at each node independently, thus enabling us to analyze both how sustainability is evaluated by the respondents and whether they evaluate this concept relying towards more andness-oriented and positive interaction among indicators or not. Detailed mathematical derivation of the andness and interaction indices are offered in Appendix A.

Pilot study – Evaluation of sustainability

The FEEM SI derives all the weights (measures) starting from a survey of experts evaluations, implemented using QUALTRICS software.¹⁴ Therefore, measures are the result of a careful reconstruction of individual preferences using a specifically built questionnaire. In the questionnaire, experts express their valuation for each indicator and their coalitions for each node separately. Figure 2 offers an example of the questionnaire in which experts are asked to express their valuation on all possible combinations of the main three pillars (economic, social and environmental pillars) of the FEEM SI node. Similarly, the indicator-coalition matrices are offered to experts at every node of FEEM SI tree, where all possible coalitions among indicators at that node are allowed. An example of the questionnaire for the final node of FEEM SI decision tree is given in Appendix Table A.2.

3.4 Representative Decision Maker

Naturally, each respondent involved in the analysis has a valuation and attitude towards sustainability indicators at the different nodes. Thus, in order to derive ‘representative’ weights to assign to for each sustainability indicator and their coalitions at each node, a consensus measure among respondents has been considered. This measure is computed using the metric distance measure, which assigns higher weights to valuations in agreement (i.e., having lower distance measure) one with the other. However, if a respondent’s valuation of sustainability indicators is extremely different from other experts’ valuations (i.e., having a higher distance measure), then a relatively lower weight is assigned to this type of respondent. By doing so, a “consensus” weight for each sustainability indicator and their coalitions at every node of FEEM SI is obtained. A detailed derivation of representative decision maker from many respondents is given in Appendix A.¹⁵

Characteristics of the Decision Makers

In this section, we present the characteristics of the representative decision maker through the description of the andness and interaction indices for all sub-nodes of the FEEM SI decision tree. These indices will shed a light on the how “consensus” weights to evaluate the sustainability indicators at each sub-node. In particular, one can assess whether the representative decision maker follows a more non-compensative behaviour at a given sub-node (i.e. having an andness index greater than 0.5) and if so, among which indicators there is a complementary (competitive) behaviour. Obviously, for the case of sub-nodes that have only two indicators, both andness and interaction indices offer the similar behaviours (e.g. having an andness index that is greater than 0.5 will give a positive interaction index value among those two indicators where indicators are evaluated as being complementary). However, for the case of sub-nodes that have more than two indicators, it is necessary to analyze both andness and interaction indices separately, since one sub-node may have an andness index greater than 0.5 but may consist of indicators having negative interaction indices (i.e. competitive) and some others having positive interaction indices (i.e. complementary). Figure 3 illustrates the andness degree of each expert for the three final pillars and the FEEM SI node, showing that the evaluations of sustainability indicators at a given node do vary among different decision makers. Some experts have similar attitude towards indicators at every pillar and the FEEM SI node (e.g., expert 6 and 14 are more *non-compensative* while expert 15 have more *compensative* attitude towards all three final pillars and the FEEM SI node).

¹⁴ Qualtrics is a private research software which enables one to build web-base surveys which is easy to distribute and allow world wide participation. Qualtrics software, furthermore, allows one to set questionnaires that are study specific which can be supported with exemplary questionnaires shown either in video and/or graphic illustrations. Once the questionnaire completed, it can be imported to Microsoft Excel file and/or statistical software programming SPSS. For further details of possible implications, please refer to <http://www.qualtrics.com/>

¹⁵ Meyer and Ponthière (2011) practice elicitation exercise which reveals a strong heterogeneity of individual preferences on hypothetical societies and examined how elicited preferences can be used to cast a new light on the ranking of actual societies. In the case of sustainability, elicitation is an important process for the evaluation of sustainability indicators and expert evaluations may differ in different societies and within societies, therefore, we derive “consensus” measures from many respondents. Moreover, we implement sensitivity analysis which allows linear combination of the measures from different respondents on sustainability levels to evaluate the robustness of the rankings and sustainability levels in section 4.4.

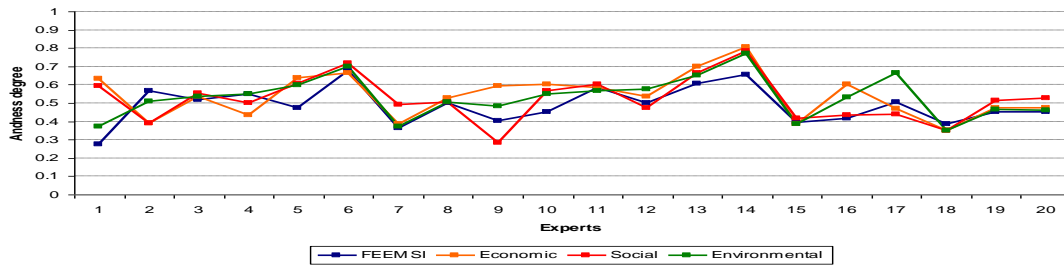


Figure 3: Andness degree of the experts at the final pillars and the FEEM SI node

As detailed in section 3.3 and confirmed in the preliminary analysis of andness degree among experts, a consensus measure has been proposed in order to derive a general “consensus” attitude towards sustainability indicators at each sub-node. Two panels of table 2 shows the andness degree and the interaction indices among indicators at each sub-node for the “representative” decision maker. The first panel of the Table 2 presents the andness degrees and interaction indices for the 3-indicator sub-nodes and the second panel offers the same information for the two-indicator sub-nodes.

Table 2 Panel A: Interaction indices and andness degree at 3-indicator sub-nodes

Node	Interaction indices			ANDNESS degree
FEEM SI	Economic	Social	Environmental	0.493
Economic	NA	-0.024	0.020	
Social		NA	-0.019	
Environmental			NA	
Economic	Growth drivers	GDP pc	Exposure	0.538
Growth drivers	NA	0.047	0.026	
GDP pc		NA	0.041	
Exposure			NA	
Social	Pop. Density	Well being	Vulnerability	0.525
Pop. Density	NA	0.016	0.041	
Well being		NA	0.020	
Vulnerability			NA	
Environmental	Air pollution	Energy	Endowments	0.532
Air pollution	NA	0.021	0.037	
Energy		NA	0.037	
Endowments			NA	
Vulnerability	Food	Private Health	Energy Security	0.528
Food	NA	0.040	0.022	
Private Health		NA	0.022	
Energy Security			NA	

Table 2 Panel B: Interaction indices and andness degree at 2-indicator sub-nodes

Node	Indicators	Interaction index	ANDNESS degree
Growth drivers	R&D, Capital Accumulation	0.058	0.529
Exposure	Relative trade, Public debt	0.187	0.5935
Well being	Education, Health	0.029	0.5145
Energy security	Imp. energy, Energy access	0.000	0.5
Air pollution	GHG p.c., CO2 Intensity	0.183	0.5915
Energy Use	Energy Intensity, Renewables	0.053	0.5265
Endowments	Biodiversity, Water	0.058	0.529
Biodiversity	Animal, Plant	0.171	0.5855

Despite the differences among experts highlighted for instance in Figure 3, the representative (consensus) decision maker has a tendency of being more andness-oriented, showing a positive interaction behaviour towards sustainability indicators in all sub-nodes with the exception of the final node of FEEM SI. The final node of FEEM SI has an andness degree of 0.493, which represents a slightly *compensative* attitude towards the final node of the FEEM SI tree.

Moreover, interaction indices between economic and social and social and environmental indicators are -0.024 and -0.019 respectively and consensus DM evaluates those interactions slightly *competitive* (or substitutes). On the other hand, the interaction index between the economic and environmental pillar is 0.020, and the representative DM evaluate those pillars as slightly complementary indicators. For the remaining sub-nodes, the representative DM features an andness index that is greater than 0.5 (i.e. more *non-compensative* attitude towards the nodes) and a positive interaction index value between two indicators at a given node (i.e. two indicator being more *complementary*).

4. Results and robustness analysis

Given the normalized sustainability indicators and the consensus weights assigned to all possible combinations of sustainability indicators, we aggregate sustainability indicators to an overall index. In this section we will present only some of the FEEM SI results in order to describe the impacts of the aggregation methodology and provide examples of the importance of such methods in evaluating policy choices; for a more complete overview of the FEEM SI results please refer to the materials available online.¹⁶

4.1 FEEM SI results using the Choquet integral as an aggregation operator

The hierarchical structure used to construct the FEEM SI allows obtaining the sustainability ranking for each year of analysis, including future projections of the sustainability levels, enlarging the scope of the analysis to policy implications. Since this paper focuses on the role of the aggregation methodology in dealing with sustainability, reported results refer only to the baseline scenario.¹⁷ Table 3 presents the sustainability rankings of countries in the year 2011.¹⁸

Table 3: FEEM SI Ranking in 2011

Rank	Country	FEEM SI	Rank	Country	FEEMSI
1	Norway	0.823	21	Russia	0.493
2	Sweden	0.774	22	RoEU	0.493
3	Switzerland	0.700	23	Mexico	0.492
4	Austria	0.691	24	Korea	0.477
5	Finland	0.661	25	Italy	0.472
6	Denmark	0.653	26	Japan	0.456
7	Canada	0.641	27	Turkey	0.453
8	France	0.630	28	Middle East	0.450
9	Ireland	0.620	29	Poland	0.430
10	New Zealand	0.609	30	South Africa	0.426
11	USA	0.554	31	Greece	0.399
12	Australia	0.553	32	RoAfrica	0.398
13	Brazil	0.546	33	RoWorld	0.385
14	UK	0.531	34	SEastAsia	0.368
15	RoEurope	0.529	35	RoFSU	0.367
16	Germany	0.525	36	North Africa	0.342
17	Portugal	0.522	37	RoAsia	0.325
18	RoLA	0.512	38	Indonesia	0.299
19	Spain	0.497	39	China	0.287
20	Benelux	0.495	40	India	0.240

Benelux: Belgium, Netherlands, and Luxembourg; **RoAfrica:** Rest of Africa; **RoAsia:** Rest of Asia; **RoEU:** Rest of European Union; **RoEurope:** Rest of Europe; **RoFSU:** Rest of Former Soviet Union; **RoLA:** Rest of Latin America; **RoWorld:** Rest of World; **SEastAsia:** Southeast Asia

Furthermore, Table 4 offers the rankings of countries and macro-regions in the economic, social and environmental pillars. The first two columns are the FEEM SI and its ranking for the countries and macro-regions and the remaining columns represent the economic, social and environmental sustainability levels and their respective rankings. One interesting aspect of this ranking is that countries that are in the higher and lower ranking positions are the ones that have a better (worse) outcome at least in two final pillars respectively. For example, Norway and Sweden not only have

¹⁶ www.feemsi.org

¹⁷ For further details and policy implications please refer to Carraro *et al.* (2012)

¹⁸ Current analysis considers individual countries (e.g., Norway) and macro-regions (e.g., Rest of Latin America) . For detailed country and macro-region classification, see Table A.3.

outstanding sustainability levels in the social pillar, but have also quite good performances both in the economic and environmental pillars. Among the lower-ranking countries, India has an extremely poor performance in the social pillar and a poor performances in the economic and social pillars. China has a moderate economic performance, but features low social and environmental sustainability outcomes. Both Rest of Asia and Indonesia have a low performance in the economic and social pillars and moderate environmental performances. On the other hand, some countries achieve good results in some pillar(s), while their remaining pillar(s) lagged behind from many countries. For example, USA and Australia have better sustainability levels in economic and social pillars, but very poor levels of environmental performance. Moreover, Korea only achieves a better economic sustainability level, but has very bad performances in social and environmental aspect.

Table 4: Sustainability pillars: Rankings in Economic, Social and Environmental Pillars in 2011

FEEM SI Rank	FEEM SI	Country	Economic	Econ. Rank	Social	Social Rank	Environ.	Envi. Rank
1	0.823	Norway	0.752	3	0.985	1	0.718	1
2	0.774	Sweden	0.728	5	0.922	2	0.664	2
3	0.700	Switzerland	0.766	1	0.668	12	0.661	3
4	0.691	Austria	0.700	7	0.755	9	0.623	5
5	0.661	Finland	0.686	8	0.799	6	0.512	10
6	0.653	Denmark	0.663	10	0.837	4	0.469	15
7	0.641	Canada	0.566	19	0.845	3	0.499	12
8	0.630	France	0.584	15	0.789	8	0.509	11
9	0.620	Ireland	0.666	9	0.683	11	0.528	8
10	0.609	New Zealand	0.591	13	0.829	5	0.411	24
11	0.554	USA	0.725	6	0.790	7	0.210	39
12	0.553	Australia	0.737	4	0.734	10	0.251	36
13	0.546	Brazil	0.446	26	0.603	17	0.597	6
14	0.531	UK	0.577	17	0.582	19	0.451	16
15	0.529	RoEurope	0.433	28	0.519	24	0.625	4
16	0.525	Germany	0.617	11	0.618	15	0.372	30
17	0.522	Portugal	0.458	23	0.646	14	0.449	17
18	0.512	RoLA	0.392	31	0.570	20	0.585	7
19	0.497	Spain	0.575	18	0.597	18	0.347	31
20	0.495	Benelux	0.611	12	0.480	29	0.396	26
21	0.493	Russia	0.586	14	0.511	25	0.393	27
22	0.493	RoEU	0.491	21	0.499	26	0.487	13
23	0.492	Mexico	0.435	27	0.656	13	0.374	29
24	0.477	Korea	0.761	2	0.330	34	0.312	33
25	0.472	Italy	0.404	30	0.559	21	0.446	19
26	0.456	Japan	0.581	16	0.351	33	0.420	22
27	0.453	Turkey	0.417	29	0.491	27	0.448	18
28	0.450	Middle East	0.558	20	0.543	22	0.283	35
29	0.430	Poland	0.463	22	0.538	23	0.304	34
30	0.426	South Africa	0.454	25	0.612	16	0.230	38
31	0.399	Greece	0.354	34	0.439	30	0.402	25
32	0.398	RoAfrica	0.279	40	0.378	32	0.523	9
33	0.385	RoWorld	0.306	37	0.405	31	0.445	20
34	0.368	SEastAsia	0.390	32	0.261	36	0.440	21
35	0.367	RoFSU	0.386	33	0.482	28	0.244	37
36	0.342	North Africa	0.350	35	0.285	35	0.385	28
37	0.325	RoAsia	0.285	39	0.185	38	0.477	14
38	0.299	Indonesia	0.331	36	0.127	39	0.419	23
39	0.287	China	0.455	24	0.260	37	0.147	40
40	0.240	India	0.301	38	0.077	40	0.328	32

The results obtained through the FEEM SI aggregation procedure suggest variability of sustainability outcomes at different pillars of sustainability. To have a deeper understanding of why some countries achieved higher (lower) levels of sustainability overall and/or in different pillars, the next section will detail the relative importance of different indicators at a given sub-node level and their global importance for the final index.

4.2 Measure of relative importance of indicators

As pointed out in the previous section, the computation of the aggregated index strongly depends on the *subjective relative importance* of a coalition with respect to another one. Given the evaluations of sustainability indicators at each

node, it is possible to infer the relative importance given to every sub-node at every node of the aggregation tree. This can be achieved by computing the Shapley values of the aggregation criteria (described in Appendix A), which amounts to compute the relative importance given to a every coalition of criteria. These results, obtained from the “representative” DM evaluations used to compute the FEEM SI values, are shown in Table 5.

Table 5: Relative importance of each indicator at a given node

Node	Criterion	Shapley value
FEEMSI	Economic	0.332
	Social	0.316
	Environmental	0.352
Economic	Growth drivers	0.378
	GDP per capita	0.355
	Exposure	0.267
Social	Population Density	0.254
	Well Being	0.415
	Vulnerability	0.331
Environment	Air pollution	0.351
	Energy	0.330
	Natural Endowment	0.319
Growth Drivers	R&D	0.522
	Investment	0.478
Exposure	Relative Trade Balance	0.554
	National Debt	0.446
Well Being	Education	0.508
	Health	0.492
Vulnerability	Food relevance	0.395
	Energy Security	0.275
	Private Health	0.330
Energy Security	Energy Imported	0.500
	Energy Access	0.500
Air pollution	GHG per capita	0.520
	CO ₂ Intensity	0.480
Energy	Energy Intensity	0.458
	Renewables	0.542
Natural Endowment	Biodiversity	0.455
	Water	0.545
Biodiversity	Animals	0.516
	Plants	0.484

These values reflect the *local* relative importance of sub-nodes at every level of the aggregation tree. For instance, it is possible to see that at the highest node (*FEEM SI*), greater relative importance is given to *environmental sustainability* (0.352) than to *economic sustainability* (0.332) or *social sustainability* (0.316).

It is also possible to combine these results in a linear fashion in order to approximate how much every final node of the aggregation tree contributes towards the determination of the final FEEM SI values. By multiplying the Shapley values of every hierarchically superior node of every indicator, from the bottom of the aggregation tree to the top (e.g., contribution of health is calculated by multiplying the Shapley values of health, well being and social pillar, since health indicator is under the node of well being which is a node of social pillar), we are able to determine the overall importance weights, ranked by decreasing value and summing to one. Table 6 offers the contribution of each indicator to the final index.

Both the structure of the tree and the evaluations provided by the DM determine the contribution of indicators to the final FEEM SI values. In this instance the indicator with the largest impact on the FEEM SI values is *GDP per capita*, which is at a distance of a node from the top of the aggregation tree, is placed besides only one other sub-node (*economic pillar*, not an indicator itself). Other indicators, being further away from the final index in the aggregation tree, being part of larger nodes and receiving lower individual valuations, contribute comparatively less to the determination of the FEEM SI values. For example, indicators that are located at the lowest sub-node (i.e., attributes that measure the animal and plant biodiversity, and countries’ energy dependence and their access to the energy) are also the ones that contribute least to the final index.

Table 6: Indicator's contribution to overall index

Indicator	Contribution to overall index
GDP per capita	0.1179
Population Density	0.0803
Education	0.0666
R&D	0.0655
Health	0.0645
GHG per capita	0.0642
Renewables	0.0630
Water	0.0612
Investment	0.0600
CO2 Intensity	0.0593
Energy Intensity	0.0532
Relative Trade Balance	0.0491
Food relevance	0.0413
National Debt	0.0395
Private Health	0.0345
Animals	0.0264
Plants	0.0247
Energy Imported	0.0144
Energy Access	0.0144

4.3 FEEM SI Results using alternative aggregation operators

The results obtained so far stem from the application of the Choquet integral as an aggregation operator throughout the decision tree (see appendix for further details). In order to understand the impact of the aggregation method on the results of the final sustainability index, we apply alternative aggregation methods to the indicators and compare the results to those of the FEEM SI values. There are several alternative aggregation operators, each of which uses the available information (indicator values, tree structure and DM evaluations) in different ways.

At the simplest end of the spectrum, we can define a bottom-up equally-weighted average aggregation operator, called EWAb, which assigns the same weight to every indicator. In this instance, all 19 indicators receive a weight of $1/19 \cong 0.053$. This operator does not take into account the structure of the tree or the evaluation given by the DM and assumes compensability among indicators.

A slightly more refined version of the equally-weighted average operator involves weighting the indicators based on the structure of the tree (operator EWAt), starting from the top and giving equal weight to every *branch* of the aggregation tree until every indicator is reached and weighted accordingly. The indicators are then aggregated by a linear combination using those weights. Also in this case, the valuations of the DM are not taken into account and the indicators are assumed to be compensable.

Alternatively, one could use the *local* Shapley values of criteria described in the previous section as weights in a linear combination process. This approach would combine indicators taking into account the structure of the FEEM SI tree, and information derived from the DM evaluation, but would still be compensative in nature.

Finally, we can aggregate the indicators using the generalized mean. This aggregation operator can be applied to every node of the aggregation tree and depend on a single parameter p that represents the degree of substitutability among indicators. This approach uses the structure of the tree and, depending on p , can be non-compensative, but in its simplest case is applied uniformly to every component of a single node and does not use the evaluation provided by the DM. For the purposes of aggregation, an arbitrary value of p has been chosen to perform the aggregation throughout the tree. Given that the evaluation of the DM does not provide precise information on the *overall* substitutability of indicators (apart from the monotonicity assumption that implies that indicators tend to be complements), the chosen value p^* has been found in such a way as to minimize the sum of square deviations of the final FEEM SI value with respect to the values obtained with the Choquet integral.

The rank of countries resulting from the application of these operators for the year 2011 is reported in appendix A (Table A.4), showing also the absolute difference in the rankings when compared with the rankings obtained by using the Choquet integral aggregation. These comparisons highlight the stark differences in rankings that all of the above-mentioned aggregation operators yield with respect to the Choquet integral. All of these aggregation operators yield up

to 10-position changes in the ranking of some countries, providing conspicuous instances of rank reversal. The differences are arguably attributable to the subtlety of the aggregation under the Choquet integral, combined with the structure of the FEEM SI tree, which allows not only to assess each indicator and node on its own, but also to characterize interactions among indicators and nodes.

4.4 Robustness analysis

In a complex aggregation such as the one used for the FEEM SI, the attitude of the representative decision maker is a key component of the process. Thus, it is important to check how robust the ranking is to a change in the representative decision maker’s attitude. On the other hand, mathematical algorithms that may be built on some theoretical and empirical grounds, it is necessary to test how robust findings are therefore, sensitivity analysis is a fundamental step during the development of any composite indicator (Saltelli et al., 2004; Saisana et al., 2005).

There exist many ways to modify the weights provided by the decision makers in the hierarchical decomposition (the decision tree); a straightforward way is to consider more than one such decision maker at the time, considering each of them as a point in the weight space. Then, a robustness analysis can be performed by building a linear convex combination of the values of the weights and run a significant number of simulations, as in a Monte Carlo approach. It is easy to prove that a linear convex combination of K different non-additive monotonic measure is an additive monotonic measure too.

We employ the robustness analysis by generating 1000 sets of measures that are necessary to aggregate the indicators into the final FEEM SI. Each of these sets constitutes, for any practical purposes, an internally consistent assessment on sustainability identical to what is provided by decision makers. These sets have thus been called “artificial decision makers” (ADMs). In this particular application, each ADM represent a univocal instance of consensus among “real” decision makers, whose measures have been combined using random weights, similarly to how the representative decision maker has been constructed. The measures contained in the artificial decision makers has been used to aggregate, with the Choquet integral, the FEEM SI, using the same indicators as for the reference case. The process results in a distribution of final FEEM SI for each country considered, which can be ranked according to the relative dominance measure ρ (derivation of the measure described in the appendix). The results of this simulation, on the 2011 FEEM SI data, are show in the Figure 4.

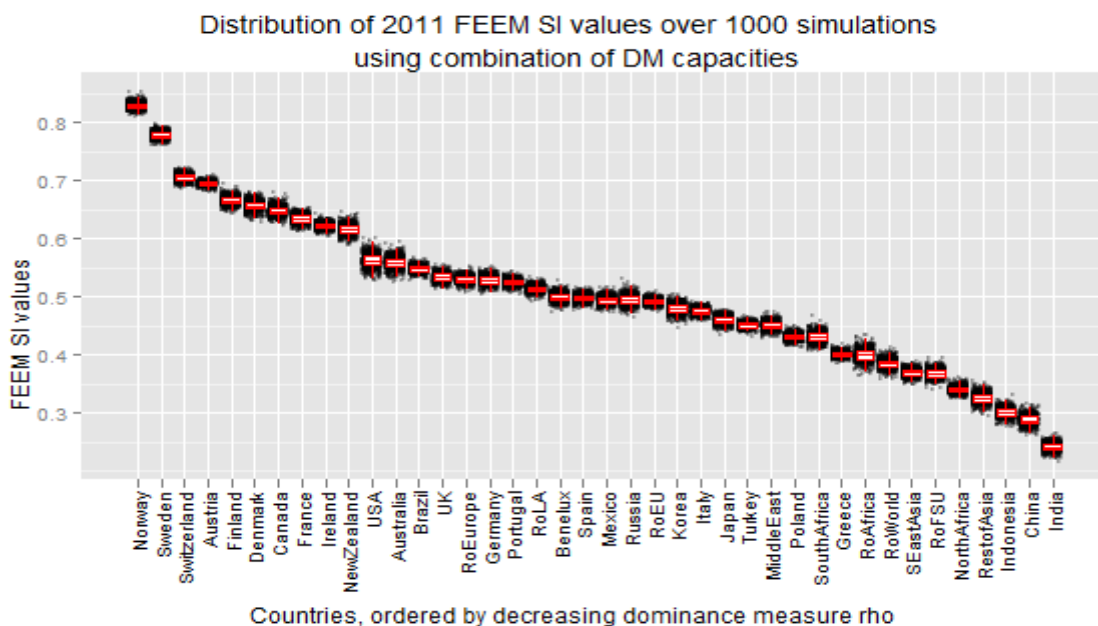


Figure 5: Distribution FEEM SI values according to 1000 artificial decision makers

The plot displays the simulated values according to every ADM. The distribution of these is summarized by box-plots for every country. It can be seen that, within the “consensus” among decision makers – reflected any ADM, which results in a point of the distribution – some countries or groups of countries clearly “dominate” others in the ranking.

One should be careful, however, whenever drawing inverse inferences from this analysis, since the distributions of simulated FEEM SI values are not independent from one another. This means that analyzing the ranking results merely by comparing the features of the distributions would not fully take into account the nature of the data and could possibly lead to misleading interpretations.

In particular, whenever any two given countries have simulated distributions that partially overlap each other, it is not necessarily the case that there is an underlying ambiguity over how these two countries are ranked according to the measures provided by ADMs. This stems from the fact that measures provided by single ADM (constructed from “real” decision makers) contributes to determine the FEEM SI values for every country. It is therefore necessary to analyze the simulation results using a measurement that takes into account the relationship between countries across simulations. The matrix, Δ , (representing average cardinal dominance) and the indices ρ^+ , ρ^- and ρ (representing, respectively, the degree to which a given country i dominates every other country, the degree to which a given country i is dominated by every other country and the relative dominance level of the country in question) have been designed for this purpose and the derivation of those are described in the appendix. These results, obtained for those indices for each country, for the 2011 FEEM SI data, are also reported in the appendix in Table A.5.

Considering the dominance analysis together with the plot of simulated distributions unveils some interesting results. For instance, it is clear that the leading countries, Norway and Sweden, are quite set apart from the rest of the group – and from each other. In fact, Norway happens to be never dominated by any other country across all simulations, a quite remarkable result given the variability introduced by the simulations. Norway and Sweden, are both followed by a group of eight countries (Switzerland, Austria, Finland, Denmark, Canada, France, Ireland and New Zealand) which constitute a faction of relatively high-scoring countries. These feature a consolidated ranking among themselves, as measured by the dominance index across simulations, which is stable by construction. These countries are followed by two somehow discontinuous cluster of countries (from USA to South Africa and from Greece to China) featuring a less dramatic discontinuity among clusters. In last position, India never dominates any other country across simulations.

By nature of the dominance analysis, these results tend to produce a robust ranking and illustrate the extent to which a change in “consensus” among decision makers can result in variability in the score of countries, thereby adding a valuable complement to the “representative” decision maker constructed for the reference aggregation.

5. Conclusions

This paper aimed at proposing an application of non-linear aggregation methods to sustainability literature, extending the current work in this field to address the intrinsic complexity underlying the sustainability concept. The aggregation approach was inspired by two considerations: firstly, the non-compensative nature of the sustainability concept, fraught with inter-linkages and synergies across its different components. Secondly, the clear policy relevance of any sustainability analysis, which requires to involve the subjective judgements of policy makers and relevant stakeholders in order to define a feasible plan for the implementation of a new definition of world progress. This requires to manage the subjective character of the decision support tool. Combining the non-additive measure algorithm -a novelty in the field of sustainability analysis- with sensitivity analysis- a well-known approach for simulation- the scoring system for sustainability assessment has been improved with respect to other similar ones. Despite the unavoidable partial uncertainty of any scoring system, the method proposed fulfils two requirements that are necessary for a rational sustainability analysis: the monotonicity and the non-compensability assumptions. Robust options are enhanced by numerical simulation, as soon as some pillars are defined as basic measures with respect to such requirements. It is quite important that these properties be fully understood and accepted.

This paper also has interesting policy-making potential. In fact, using the method proposed in this paper, a complete sustainability ranking of the regions of the world has been proposed, both for the current and future years, exploiting the features of the ICES computable general equilibrium model. Thus, comparative static analysis both across countries and through time has been made possible- a novelty in the field of sustainability assessment that may have important policy-making applications.

The analysis has been completed by three further investigations: through the computation of the Shapley index, it has been possible to address the relative importance of different indicators, which could also be used in the future to refine the current sustainability tree. Secondly, ANDNESS and interaction indices highlight that the representative DM

evaluates majority of the sustainability indicators as being more *complementary* and therefore, for a country to have a higher sustainability level, it needs to perform well in all indicators rather than simply having a satisfactory performance in only one of those. Finally, a robustness analysis has provided a measure of the subjectivity implied by the artificial decision makers developed, confirming the validity of the overall method in evaluating overall sustainability. Despite the importance of extending the current pool of decision makers involved in the analysis, the method proposed is already able to capture important information about sustainability, economizing on computational time without sacrificing too much information –another important feature for policy-making applications.

Lastly, the comparison of the rankings obtained through the Choquet integral and other aggregation operators - some of which are much more frequently used in sustainability indices such as the EWA –has shown the ability of Choquet integral method to address subtle inter-linkages and connections across components, supporting the use of such methods for sustainability assessment.

As future development, it will be important to extend the pool of decision makers involved in the determination of the measures used in the aggregation operator. Extension of the pool of decision makers will not only allow for obtaining a more representative ‘consensus’ measures, but also for the evaluation of sustainability perceptions from different parts of the world. Given the heterogeneity of the current conditions (economic, social and environmental condition in general) in different countries (macro-regions), the need for future generations will vary and therefore importance given to sustainability indicators and their interactions may differ. If this is the case, a tool like the FEEM SI can offer different policy implications in different regions considered in the analysis.

References

- Agliardi, E. (2011). Sustainability in uncertain economies, *Environmental and Resource Economics* 48, 71-82.
- Arrow, K.J., Dasgupta, P.S., Goulder, L.H., Daily, G., Ehrlich, P.R., Heal, G.M., Levin, S., Mäler, K.-G., Schneider, S., Starett, D.A., Walker, B. (2004). Are we consuming too much? *Journal of Economic Perspectives* 18, 147-172.
- Arrow, K.J., Dasgupta, P.S., Goulder, L.H., Mumford, K.J., Oleson, K. (2012). Sustainability and the measurement of wealth. *Environmental and Development Economics*, 17, 317-353.
- Bohringer, C., Loschel, A. (2006). Computable general equilibrium models for sustainability impact assessment: Status quo and prospects. *Ecological Economics*, 60(1), 49-64.
- Billaut, J.C., Bouyssou, D., Vincke, P. (2010). Should you believe in Shanghai ranking? *Scientometrics*, 84(1), 237-263.
- Bossel, H. (1999). *Indicators for Sustainable Development: Theory, Method, Applications*. IISD International Institute for Sustainable Development. Winnipeg, Manitoba.
- Carraro, C., Campagnolo, L., Eboli, F., Lanzi, E. Parrado, R., Portale, E. (2012). Quantifying sustainability: A new approach and world rankings. FEEM Working Papers 94.2012.
- Despic, O. and Simonovic, S. P. (2000). Aggregation operators for decision making in water resources, *Fuzzy Sets and Systems*, 115(1), 11-33.
- Eboli, F., Parrado, R., Roson, R. (2010). Climate Change Feedback on Economic Growth: Explorations with a Dynamic General Equilibrium Model. *Environment and Development Economics*, 15(5), 515-533.
- EPI (2010) *Environmental Performance Index*, Summary for policymakers, Yale and Columbia Universities
- ESI (2005). Environmental Sustainability Index. Yale Center for Environmental Law & Policy, New Haven.
- European Commission (2005), “Sustainable Development Indicators to monitor the implementation of the EU Sustainable Development Strategy”, Communication from Mr. Almunia to the Members of the Commission, (SEC(2005) 161 final).
- European Environment Agency (2005), *EEA core set of indicators guide*, Technical Report No. 1.
- European Union Sustainable Development Strategy (EU SDS). (2006). *Review of the European Union Sustainable Development Strategy – Renewed Strategy*.
- Grabisch, M. (1995). Fuzzy integral in multicriteria decision making. *Fuzzy Sets & Systems* 69, 279-298.
- Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European J. of Operational Research*, 89,

445-456.

Grabisch, M. (1997). k -Order additive discrete fuzzy measure and their representation. *Fuzzy Sets and Systems*, 92, 131–295.

Grabisch, M., Labreuche, C. and Vansnick, J.C. (2003). On the extension of pseudo-Boolean functions for the aggregation of interacting criteria. *European Journal of Operational Research*, 148(1), 28-47.

Grabish M., Marichal J.-L., Mesiar R., Pap E. (2009). *Aggregation Function*, Cambridge University Press, Cambridge, ISBN 978-0-521-51926-7.

GRI – Global Reporting Initiative. Sustainability Reporting Guidelines, version 3.1, Global Reporting Initiative, Amsterdam, The Netherlands.

GRI – Global Reporting Initiative. Global Reporting Initiative Sustainability Report 2009/2010, Global Reporting Initiative, Amsterdam, The Netherlands.

Grossman G. and Krueger A. (1993). Environmental impacts of a North American free trade agreement. In *The U.S.-Mexico Free Trade Agreement*, Garber, P (ed). MIT Press: Cambridge; 13-56.

Kacprzyk J., Fedrizzi M. and Nurmi H. (1992). Group decision making and consensus under fuzzy preferences and fuzzy majority, *Fuzzy Sets and Systems*, 49, 21-31.

Klement E.P., Mesiar R. and Pap E. (2000). *Triangular norms*, Kluwer Academic Publishers, Netherlands.

Krajnc, D., Glavic, P. (2005) A model for integrated assessment of sustainable development. *Resource Conservation Recycling* 43, 189–208.

Marichal J.-L. (1998). Aggregation Operators for Multicriteria Decision Aid, PhD thesis, Institute of Mathematics, University of Liège, Liège, Belgium.

Marichal J.-L. and Roubens, M. (2000). Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, 124(3), 641-650.

Meyer P., and Ponthière, G. (2011). Eliciting Preferences on Multiattribute Societies with a Choquet Integral, *Computational Economics*, 37(2), 133-168.

Millennium Ecosystem Assessment. (2005). *Ecosystems and Human Well-Being: Synthesis*. Island Press: Washington, DC.

Munda, G. (1997). Environmental economics, ecological economics, and the concept of sustainable development. *Environmental Values*, 6(2), 213-233.

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

Murofushi T., Sugeno M., and Machida M. (1994). Non-monotonic fuzzy measures and Choquet integral, *Fuzzy Sets and Systems*, 64, 73-86.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S. (2005). Tools for Composite Indicators Building. European Commission, Ispra.

Noorbakhsh, F. (1998). A modified human development index. *World Development*, 26(3), 517-528.

Ness, B., Urbel-Piirsalu, E., Anderberg, S., Olsson, L. (2007). Categorising tools for sustainability assessment. *Ecological Economics* 60, 498–508.

OECD, European Commission, Joint Research Centre. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD, Paris.

Panayotou, T. (1993). "Empirical Tests and Policy Analysis of Environmental Degradation at Different Stages of Economic Development." Working Paper WP238, Technology and Employment Programme, International Labor Office, Geneva

Ravallion, M. (1997). Good and bad growth: The human development reports. *World Development*, 25(5), 631-638.

Ravallion, M. (2010). Troubling tradeoffs in the Human Development Index. Policy Research Working Paper 5484, World Bank, Washington DC.

Saisana M., Tarantola S. and Saltelli A. (2005), “Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators”, *Journal of the Royal Statistical Society A*, 168(2): 307-323.

Saltelli A. Tarantola S., Campolongo F. and Ratto M. (2004), *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*.

John Wiley & Sons publishers, New York.

Selden, T.M., and Song, D. (1994). Environmental Quality and Development: Is there a Kuznets Curve for Air Pollution Emissions? *Journal of Environmental Economics and Management*, 27, 147-162.

Singh, R. K., Murty, H., Gupta, S., and Dikshit, A. (2009). An overview of sustainability assessment methodologies. *Ecological Indicators*, 9(2), 189-212.

Ulengin, F., Topuc, Y.I., Sahin, S.O. (2001). An integrated decision aid system for Bosphorus water-crossing problem. *European Journal of Operational Research*, 134, 179-192.

UN CSD (2005). *Indicators of sustainable development – CSD theme indicator framework*. UN Commission on Sustainable Development, NY, USA.

Vincke, P. (1999). Outranking approach. In: T. Gal, T. J. Stewart, T. Hanne (eds.), *Multicriteria decision making. Advances in MCDM models, algorithms, theory and applications*. Boston, MA: Kluwer.

Von Altrock C. (1995). *Fuzzy logic and neurofuzzy applications explained*, Prentice Hall PTR, Upper Saddle River, NJ.

WCED – World Commission on Environment and Development. (1987). *Our Common Future*. Oxford Univ. Press, Oxford.

World Bank. (2010). *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium*. World Bank: Washington, DC.

Yager R. R. (1993). Families of OWA operators, *Fuzzy Sets and Systems*, 59, 1 25-148.

APPENDIX A

A.1 Non-Additive Measures and the Choquet Integral

In this section, we present the formal definition of NAM and the Choquet integral. To this purpose, let us consider the following Definition:

Definition 1. Let $N = \{1, 2, 3, \dots, n\}$ be the set of attribute for a given node in the tree. A *non additive (monotonic) measure* is a set function $m : S \subseteq N \rightarrow [0, 1]$, which satisfies:

$$m(\emptyset) = 0, m(N) = 1$$

$$\forall S, T \subseteq N : S \subseteq T \Rightarrow m(S) \leq m(T),$$

The two first constraints are two border condition, while the second represents a monotonicity constraint, a rational property. A NAM is *additive* if $m(S \cup T) = m(S) + m(T)$, $S \cap T = \emptyset$, while if $m(S \cup T) < m(S) + m(T)$, $S \cap T = \emptyset$, the measure is called *sub-additive*, implying a *redundancy* effect, and if $m(S \cup T) > m(S) + m(T)$, $S \cap T = \emptyset$ it is *super-additive* (a *synergic* effect).

The values of the criteria will be aggregated using the Choquet integral or similar methods like the *multi-linear* approach (Grabish 1995, 1996). Again, if the measure of all the subsets with the same cardinality is the same, the OWA is obtained (Ordered Weighted Averaging (Yager 1993)), and as particular cases the Min and the Max operators, the k-order statistics, their combination, and other ones. Let now (x_1, \dots, x_n) be the values of the (normalized) criteria, obtained from the benchmark filtering. Let $(x_{(1)}, \dots, x_{(n)})$ the ordered vector of the vector (x_1, \dots, x_n) , obtained by a suitable permutation of indices, so that $0 = x_{(0)} \leq x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ and $x_{(i)} \in (x_{(0)}, \dots, x_{(n)})$, $\forall i = 1, \dots, n$.

Definition 2. The Choquet integral of the vector (x_1, \dots, x_n) with $x_i \in [0, 1]$ with respect to the (non additive) measure $m : S \subseteq N \rightarrow [0, 1]$ is given by:

$$C_m(x_1, \dots, x_n) = \sum_{i=1}^n (x_{(i)} - x_{(i-1)}) \cdot m(A_{(i)})$$

being $A_{(i)} = \{i+1, i+2, \dots, n\}$, and $A_{(n+1)} = \emptyset$, $x_{(0)} = 0$. The same integral can be similarly written as:

$$C_m(x_1, \dots, x_n) = \sum_{i=1}^n x_{(i)} [m(A_{(i)}) - \mu(A_{(i+1)})]$$

showing that it extends the WA approach, given that it is the linear combination of the marginal gains between the ordered criteria, permitting to represent many aggregation operators, suitably tuning the values of the capacity. The Choquet integral is mathematically characterized by a set of properties and requirements, and can be alternatively computed using the Möbius transform α_m of the non additive measure, see (Grabisch et al., 2009) as follows:

$$C_m(x_1, x_2, \dots, x_n) = \sum_{T \subseteq N} \alpha_m(T) \min_{i \in T} x_i$$

where:

$$\alpha_m(T) = \sum_{S \subseteq T} (-1)^{|S|-|T|} m(S) \quad \forall S \subseteq N$$

are the Möbius coefficients associated to the capacity m .

Being that the *minimum* operator is the upper bound of a class of conjunctive operators, namely, the T-norm (*triangular* norm), which are monotonic operators that extend the logical conjunction to real values (*pessimistic* operator, in the sense that it computes the minimum of its arguments, independently on the values of the other ones; in this sense, it is a totally *non compensative* operator). In the computation of the Choquet integral using the Möbius representation, substituting the T-norm minimum with an other one, a different aggregation operator is obtained, the *multi linear* aggregation operator, similar to the Choquet integral for what it concerns the majority of characterizing properties. In particular, the T-norm *product* can be alternatively used, since, with respect to the minimum T-norm, is differentiable, and smoother, see Marichal (1998).

A.2 Shapley Value

The Shapley value characterizes the “relative importance” of each criterion and can be derived directly by the NAM values. The Shapley value can be computed for each criterion at every node of the hierarchy tree. It is obtained by averaging all the marginal gains obtained by adding the criterion to every coalition not including itself (Grabisch, 1995 and 1996).

For the *i*-th criterion, the Shapley value is calculated as follows:

$$v(i) = \sum_{T \subseteq N \setminus i} \frac{(n-t-1)!t!}{n!} [m(T \cup i) - m(T)] \quad \text{where } t = \text{card}(T)$$

These values have the property that $\sum_{i=1}^n v(i) = 1$. It is possible to verify that the Shapley values vary between 0 and 1, higher value representing higher importance of that criterion. It is also convenient to scale these values by a factor n , therefore, a value greater than 1 indicates an attribute more important than the average.

This value can be written alternatively as:

$$v(i) = \frac{1}{n} \sum_{T \subseteq N \setminus i} \frac{(n-t-1)!t!}{(n-1)!} [m(T \cup i) - m(T)] \quad \text{where } t = \text{card}(T)$$

$\frac{(n-t-1)!t!}{(n-1)!}$ being the inverse of the number of subsets with cardinality t . It follows that it can be interpreted as the average of the marginal average gain of each coalition with the same cardinality.

A.3 ORNESS and ANDNESS Indices

It is clear that as soon as the measure values, $m(A_{(i)})$, are close to $(0,1,1,\dots,1)$, i.e. the maximum operator, the DM behaviour tends to be optimistic, and the contrary, as the measure values, $m(A_{(i)})$, are close to $(0,0,0,\dots,0,1)$, i.e. the minimum operator, the DM behaviour tends to be pessimistic.

To characterize whether DM follows more pessimistic or optimistic behaviour, it is possible to compute an index, depending solely on the measure values, the ORNESS index, together with the ANDNESS index. The former one measures the tendency to optimism, while the second one, measures the tendency to pessimism. Using the Möbius values of the measure, the ORNESS is computed as follows¹⁹:

¹⁹ The ANDNESS index can be computed also using the measure values, but the computation is more complicated, and it is not here reported.

$$ORNESS_m(i) = \frac{1}{n-1} \sum_{T \subseteq N} \frac{n-t}{t+1} a(T)$$

If ORNESS=1, then the DM is fully optimistic, implicitly using the maximum operator (logical disjunction), if ORNESS=0 (i.e. ANDNESS=1 since ORNESS+ANDNESS=1), then the DM is extremely pessimistic, corresponding to the minimum operator (logical conjunction), if ORNESS=0.5 the DM is additive in average.

A.4 Interaction Index

The intuition behind the interaction index is very similar to Shapley index but considering two indicators' contribution together rather than only one indicator's. Let's consider two indicators, i and j , and if the $m(i,j) > m(i) + m(j)$, then it shows a complementary effect between i and j . Similarly, $m(i,j) < m(i) + m(j)$ suggests that i and j interact in a redundant (substitutive) way. Finally, if $m(i,j) = m(i) + m(j)$, it can be considered that the indicators i and j do not interact, i.e., that they have independent roles.

To measure the all interaction of two indicators, the average interaction between two indicators i and j is calculated with the following interaction index (see Murofushi and Soneda, 1993):

$$I_m(ij) = \sum_{T \subseteq N \setminus ij} \frac{(n-t-2)!t!}{(n-1)!} [m(T \cup ij) - m(T \cup i) - m(T \cup j) + m(T)] \quad \text{where } t = \text{card}(T)$$

The outcome of $I_m(ij)$ can be considered as a measure of the average marginal interaction between i and j . One of the important property of the interaction index is that $I_m(ij) \in [-1,1]$ for all i and j . The interaction index being 1 (respectively -1) represents to full complementarity (resp. substitutivity) between i and j (see Grabisch, 1997).

A.5 Aggregation of Decision Makers

Let v_{ki} is the valuation (i.e. judgement) of k -th DM for i -th coalition at a given sub-node. Each sub-node have $2^n - 2$ possible coalitions that Decision Makers may have different judgements where n is the number of indicators at a given sub-node. One can calculate the distance measure of every DM's valuations to all other DMs' valuations at a given node.

Let us denote D_k , as the total absolute distance of k -th DM's judgements to all other DMs' judgements. D_k can be calculated as follows:

$$D_k = \sum_{i=1}^{i=2^n} \sum_{l=1}^{l=m} |v_{ki} - v_{li}|$$

where $l \neq k$, $l = 1, 2, \dots, m$ are the DMs, i is the valuation of a criterion or any possible combinations of criteria and n is the number of indicators at a given sub-node.

After calculating the absolute distance measure for each DM, we can find the sum of absolute distances of all DMs as:

$$\bar{D} = \sum_{k=1}^{k=m} D_k$$

Since absolute distance measures for all DMs and also the sum of absolute distance of all DMs are obtained for a given sub-node, (consensus) weights of each DM can be calculated. Each DM's weight is inversely related to the ratio of

DM's absolute distance to the sum of all absolute distances. For instance, if a DM has the lowest absolute distance, that DM's valuations should be weighted more than the other DMs.

$$W_k = \left(\frac{D_k}{D} \right)^{-1}$$

Weights given to each DM's evaluation can be further normalized so that the weights are bounded between 0-1.

$$w_k = \frac{W_k}{\sum_{k=1}^{k=m} W_k} \quad \text{where} \quad \sum_{k=1}^{k=m} w_k = 1$$

Since we have the normalized weights for each DM, one can obtain the "representative" DM valuations, v_i^r , for all possible coalitions at a given sub-node by weighted average of coalitions as:

$$v_i^r = \sum_{k=1}^{k=m} w_k v_{ki} \quad \text{for } \forall i \text{ where } i = 1, 2, \dots, 2^n$$

After obtaining representative measures for each coalition at each node of FEEM SI tree, the Choquet integral is used to aggregate all indicators to an overall index where the aggregation takes place at different stages starting from bottom nodes and ending at the final node.

A.6 Dominance analysis

As described in section 4.4, the analysis of the simulation results should take into account the fact that the distributions of simulated FEEM SI values are not independent from one another, since the data provided by single ADM contributes to determine the FEEM SI values for every country.

In order to describe more accurately the simulation results, the following measures have been implemented to compare any two countries i and j included in the ranking:

$$\Delta(i, j) = \frac{1}{N} \sum_{k=1}^k F[R_k(i) - R_k(j)]$$

where N is the number of countries included in the ranking, $R(i)$ and $R(j)$ are the FEEM SI values for the i^{th} and j^{th} country respectively. k is the number of simulations and $F(x)$ takes the form:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

Constructed in this way, $\Delta(i, j)$ represents the "average cardinal dominance" of country i on country j . That is, the measurement expresses by how much, on average, the i^{th} country dominates the j^{th} across simulations. The overall dominance measure of country i on every other country is given by:

$$\rho^+(i) = \frac{1}{N-1} \sum_{j=1}^N \Delta(i, j)$$

Whereas the degree to which country i is dominated by every other country is given by

$$\rho^-(i) = \frac{1}{N-1} \sum_{j=1}^N \Delta(j, i)$$

We can thus construct the following measure:

$$\rho(i) = \frac{\rho^+(i)}{\rho^+(i) + \rho^-(i)}$$

which indicates the extent of relative dominance of the i^{th} country. This measures 1 if the country in question dominates any other across all simulations and measures 0 if country i is being dominated by all other countries. Being within the $[0,1]$ range, its interpretation is quite straightforward.

Table A.1: Indicators and their description

DIMENSION	NODE	INDICATORS		DESCRIPTION
Economic Dimension	Growth drivers	1.R&D		R&D Expenditure / GDP
		2.Investment		Net investment / Capital stock
	GDP per capita	3.GDP per capita		GDP / Population
	Exposure	4.Relative trade balance		Net exports/ (Exports + Imports)
		5.Public debt		Government debt / GDP
Social Dimension	Population density	6.Population density		Total population / Total area of the country (in kilometre square)
	Wellbeing	7.Education		Expenditure on education / GDP
		8.Health		Expenditure on health / GDP
	Vulnerability	9.Food relevance		Total food expenditure / Total expenditure
		10.Private health		Private health expenditure / GDP
		Energy Security	11.Energy imported	Energy imported / Total energy consumption
			12.Energy Access	Population that has access to electricity / Total population
Environmental Dimension	Air pollution	13.GHG per capita		Total GHG emissions / Total population
		14.Co2 intensity		Total Co2 emissions / Total primary energy consumption
	Energy	15.Energy intensity		Total primary energy supply / GDP
		16.Renewables		Renewable energy consumption / Total primary energy consumption
	Natural endowment	17.Water		Total water use / Total renewable water resources available
		Biodiversity	18.Animals	Endangered Species / Total Species
			19.Plants	Endangered Species / Total Species

Table A.2: Sample questionnaire of the FEEM SI final node



Overall sustainability

You will have to rate the **sustainability** of some hypothetical scenarios.

In these scenarios, the following indicators can have a **GOOD** or **BAD** performance:

- **ECON**: Economic sustainability
- **SOC**: Social sustainability
- **ENV**: Environmental sustainability

Please **rate on a 0-100 scale** all of the following hypothetical situations.

Remember: a scenario where TWO indicators have a "GOOD PERFORMANCE" must be rated at least as the maximum score assigned to a scenario where only ONE of these TWO indicators have a "GOOD PERFORMANCE"

drag to determine your evaluation

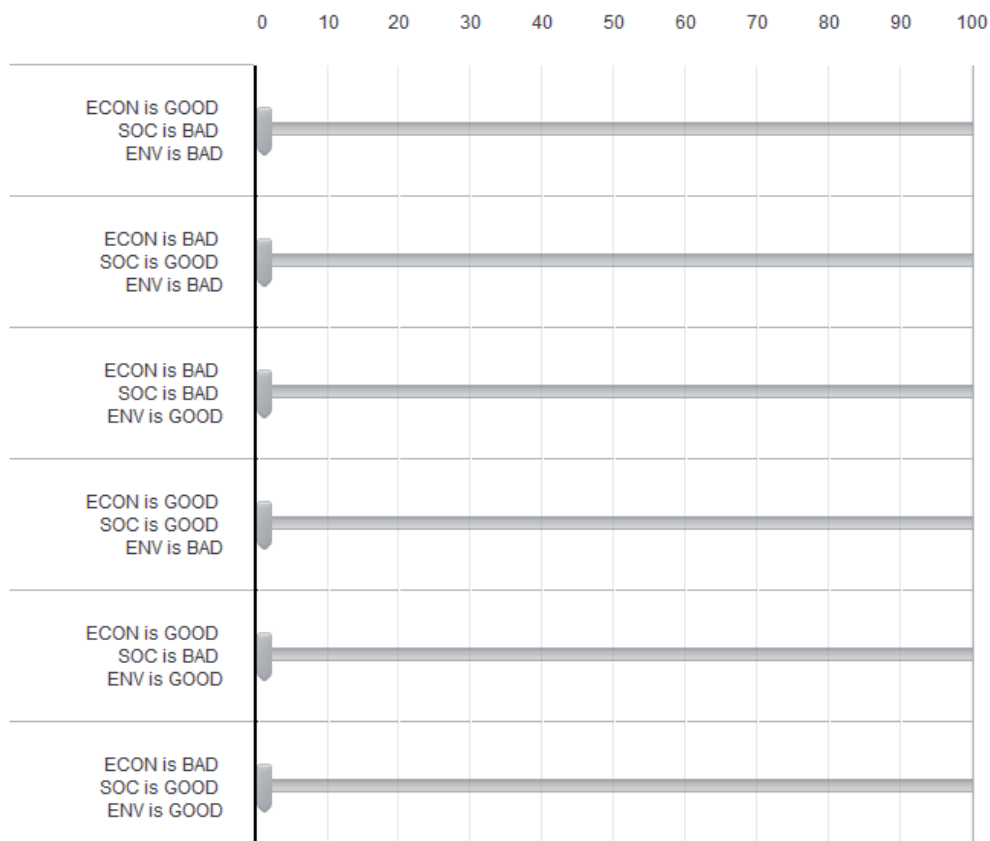


Table A.3: List of countries and macro-regions

No.	Macro-Regions	Countries
1	Australia	Australia
2	New Zealand	New Zealand
3	Japan	Japan
4	Korea	Korea
5	China	China, Hong Kong, Taiwan
6	India	Indonesia
7	Indonesia	India
8	SEastAsia	Malaysia, Philippines, Singapore, Thailand, Vietnam
9	RoAsia	Afghanistan, Bangladesh, Bhutan, Brunei Darassalam, Cambodia, Democratic Republic of Korea, Lao People's Democratic Republic, Macau, Maldives, Mongolia, Myanmar, Nepal, Pakistan, Sri Lanka, Timor Leste
10	USA	USA
11	Canada	Canada
12	Mexico	Mexico
13	Brazil	Brazil
14	RoLA	Argentina, Bolivia, Chile, Colombia, Ecuador, Paraguay, Peru, Uruguay, Venezuela, Falkland Islands (Malvinas), French Guiana, Guyana, Suriname, Costa Rica, Guatemala, Nicaragua, Panama, Belize, El Salvador, Honduras, Saint Vincent and the Grenadines, Trinidad and Tobago, Turks and Caicos, Anguilla, Antigua & Barbuda, Aruba, Bahamas, Barbados, Cayman Islands, Cuba, Dominica, Dominican Republic, Grenada, Guadeloupe, Haiti, Jamaica, Martinique, Montserrat, Netherlands Antilles, Puerto Rico, Saint Kitts and Nevis, Saint Lucia, Virgin Islands (British), Virgin Islands (U.S.)
15	Austria	Austria
16	Benelux	Belgium, Luxembourg, Netherlands
17	Denmark	Denmark
18	Finland	Finland
19	France	France
20	Germany	Germany
21	Greece	Greece
22	Ireland	Ireland
23	Italy	Italy
24	Poland	Poland
25	Portugal	Portugal
26	Spain	Spain
27	Sweden	Sweden
28	UK	UK
29	RoEU	Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Malta, Slovakia, Slovenia, Bulgaria, Romania
30	Switzerland	Switzerland
31	Norway	Norway
32	RoEurope	Albania, Andorra, Bosnia and Herzegovina, Croatia, Faroe Islands, Gibraltar, Iceland, Liechtenstein, Macedonia, the former Yugoslav Republic of, Monaco, San Marino, Serbia and Montenegro
33	Russia	Russia
34	RoFSU	Belarus, Ukraine, Moldova, Republic of, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan, Armenia, Azerbaijan, Georgia
35	Turkey	Turkey
36	MiddleEast	Bahrain, Islamic Republic of Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Occupied Palestinian Territory, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, United Arab Emirates, Yemen
37	North Africa	Algeria, Egypt, Libyan Arab Jamahiriya, Morocco, Tunisia
38	RoAfrica	Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Congo, Democratic Republic of the Congo, Cote d'Ivoire, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mayotte, Mozambique, Niger, Nigeria, Reunion, Rwanda, Saint Helena, Sao Tome and Principe, Senegal, Seychelles, Sierra Leone, Somalia, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, Zimbabwe
39	South Africa	South Africa
40	RoWorld	American Samoa, Cook Islands, Fiji, French Polynesia, Guam, Kiribati, Marshall Islands, Micronesia, Federated States of, Nauru, New Caledonia, Norfolk Island, Northern Mariana Islands, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tokelau, Tonga, Tuvalu, Vanuatu, Island of Wallis and Futuna, Bermuda, Greenland, Saint Pierre and Miquelon

Table A.4: Comparison of alternative aggregation operators for the year 2011

Country	Rank FEEMSI	Rank EWAb	Δ ranking wrt Choquet	Rank EWAt	Δ ranking wrt Choquet	Rank GM(p*)	Δ ranking wrt Choquet
Australia	12	13	-1	11	+1	25	-13
Austria	4	4	=	3	+1	3	+1
Benelux	20	14	+6	23	-3	24	-4
Brazil	13	16	-3	13	=	14	-1
Canada	7	8	-1	7	=	10	-3
China	39	38	+1	39	=	37	+2
Denmark	6	6	=	6	=	7	-1
Finland	5	5	=	5	=	5	=
France	8	7	+1	10	-2	8	=
Germany	16	15	+1	18	-2	15	+1
Greece	31	35	-4	33	-2	30	+1
India	40	40	=	40	=	40	=
Indonesia	38	39	-1	38	=	39	-1
Ireland	9	9	=	8	+1	9	=
Italy	25	27	-2	28	-3	20	+5
Japan	26	28	-2	30	-4	29	-3
Korea	24	20	+4	26	-2	28	-4
Mexico	23	25	-2	22	+1	22	+1
MiddleEast	28	26	+2	24	+4	21	+7
NewZealand	10	10	=	9	+1	13	-3
NorthAfrica	36	33	+3	35	+1	32	+4
Norway	1	1	=	1	=	1	=
Poland	29	29	=	29	=	23	+6
Portugal	17	21	-4	20	-3	16	+1
RoAfrica	32	32	=	31	+1	36	-4
RoAsia	37	37	=	37	=	38	-1
RoEU	22	18	+4	19	+3	17	+5
RoEurope	15	11	+4	15	=	6	+9
RoFSU	35	34	+1	34	+1	34	+1
RoLA	18	19	-1	16	+2	11	+7
RoWorld	33	31	+2	32	+1	35	-2
Russia	21	17	+4	14	+7	27	-6
SEastAsia	34	36	-2	36	-2	31	+3
SouthAfrica	30	30	=	27	+3	26	+4
Spain	19	23	-4	21	-2	19	=
Sweden	2	2	=	2	=	2	=
Switzerland	3	3	=	4	-1	4	-1
Turkey	27	24	+3	25	+2	12	+15
UK	14	12	+2	17	-3	18	-4

Table A.5: Ranking of countries according to average dominance index across simulated values, ρ , for 2011 FEEM SI

Country	ρ^+	ρ^-	ρ
Norway	8.272517	0	1
Sweden	7.017498	0.032180	0.995435
Switzerland	5.241969	0.125629	0.976595
Austria	4.993397	0.145784	0.971633
Finland	4.322133	0.220431	0.951474
Denmark	4.124075	0.249059	0.943048
Canada	3.932928	0.282739	0.932931
France	3.632407	0.346158	0.912994
Ireland	3.381697	0.409414	0.892007
NewZealand	3.264846	0.443318	0.880448
USA	2.244132	0.783920	0.741114
Australia	2.171497	0.811651	0.727921
Brazil	1.979238	0.893875	0.688883
UK	1.738591	1.010101	0.632516
RoEurope	1.671711	1.049046	0.614429
Germany	1.658449	1.056061	0.610957
Portugal	1.592388	1.097545	0.591980
RoLA	1.427004	1.218559	0.539395
Benelux	1.221855	1.392268	0.467405
Spain	1.206080	1.406183	0.461699
Mexico	1.163242	1.451047	0.444955
Russia	1.163253	1.453589	0.444526
RoEU	1.130448	1.487507	0.431806
Korea	0.994921	1.667540	0.373685
Italy	0.950801	1.733224	0.354245
Japan	0.801836	1.979846	0.288256
Turkey	0.725650	2.124143	0.254632
MiddleEast	0.722531	2.130927	0.253212
Poland	0.572879	2.479087	0.187708
SouthAfrica	0.568334	2.491081	0.185766
Greece	0.383163	3.049703	0.111616
RoAfrica	0.372474	3.087105	0.107665
RoWorld	0.294316	3.390436	0.079874
SEastAsia	0.226390	3.716297	0.057420
RoFSU	0.221337	3.745096	0.055803
NorthAfrica	0.132058	4.360702	0.029393
RestofAsia	0.094577	4.698028	0.019734
Indonesia	0.046588	5.290119	0.008730
China	0.029825	5.608623	0.005290
India	0	6.771014	0

NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI

Fondazione Eni Enrico Mattei Working Paper Series

Our Note di Lavoro are available on the Internet at the following addresses:

<http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1>
http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalbrowse&journal_id=266659
<http://ideas.repec.org/s/fem/femwpa.html>
<http://www.econis.eu/LNG=EN/FAM?PPN=505954494>
<http://ageconsearch.umn.edu/handle/35978>
<http://www.bepress.com/feem/>

NOTE DI LAVORO PUBLISHED IN 2013

CCSD	1.2013	Mikel Bedayo, Ana Mauleon and Vincent Vannetelbosch: Bargaining and Delay in Trading Networks
CCSD	2.2013	Emiliya Lazarova and Dinko Dimitrov: Paths to Stability in Two-sided Matching with Uncertainty
CCSD	3.2013	Luca Di Corato and Natalia Montinari: Flexible Waste Management under Uncertainty
CCSD	4.2013	Sergio Currarini, Elena Fumagalli and Fabrizio Panebianco: Games on Networks: Direct Complements and Indirect Substitutes
ES	5.2013	Mirco Tonin and Michael Vlassopoulos: Social Incentives Matter: Evidence from an Online Real Effort Experiment
CCSD	6.2013	Mare Sarr and Tim Swanson: Corruption and the Curse: The Dictator's Choice
CCSD	7.2013	Michael Hoel and Aart de Zeeuw: Technology Agreements with Heterogeneous Countries
CCSD	8.2013	Robert Pietzcker, Thomas Longden, Wenying Chen, Sha Fu, Elmar Kriegler, Page Kyle and Gunnar Luderer: Long-term Transport Energy Demand and Climate Policy: Alternative Visions on Transport Decarbonization in Energy Economy Models
CCSD	9.2013	Walid Oueslati: Short and Long-term Effects of Environmental Tax Reform
CCSD	10.2013	Lorenza Campagnolo, Carlo Carraro, Marinella Davide, Fabio Eboli, Elisa Lanzi and Ramiro Parrado: Can Climate Policy Enhance Sustainability?
CCSD	11.2013	William A. Brock, Anastasios Xepapadeas and Athanasios N. Yannacopoulos: Robust Control of a Spatially Distributed Commercial Fishery
ERM	12.2013	Simone Tagliapietra: Towards a New Eastern Mediterranean Energy Corridor? Natural Gas Developments Between Market Opportunities and Geopolitical Risks
CCSD	13.2013	Alice Favero and Emanuele Massetti: Trade of Woody Biomass for Electricity Generation under Climate Mitigation Policy
CCSD	14.2013	Alexandros Maziotis, David S. Saal and Emmanuel Thanassoulis: A Methodology to Propose the X-Factor in the Regulated English and Welsh Water And Sewerage Companies
CCSD	15.2013	Alexandros Maziotis, David S. Saal and Emmanuel Thanassoulis: Profit, Productivity, Price and Quality Performance Changes in the English and Welsh Water and Sewerage Companies
CCSD	16.2013	Caterina Cruciani, Silvio Giove, Mehmet Pinar and Matteo Sostero: Constructing the FEEM Sustainability Index: A Choquet-integral Application