ISLANDS AND SMALL STATES INSTITUTE

**Occasional Papers on Islands and Small States**

# CONCEPTUAL ISSUES IN CONSTRUCTING COMPOSITE INDICES

**Nadia Farrugia**

**No: 2/2007**

**ISSN 1024-6282**

# CONCEPTUAL ISSUES IN CONSTRUCTING COMPOSITE INDICES

**Nadia Farrugia***

**ABSTRACT**

This paper presents an analysis of the conceptual issues associated with the construction of composite indices.

Composite indices, which are constructed by averaging a number of indicators or sub-indices, are multidimensional, in that they represent aggregate measures of a combination of factors. They are often used to simplify complex measurement constructs, and often have a strong political appeal due to the fact that they simplify complex matters into a single number. However, composite indices are often criticized due to their subjectivity. Indeed the methodology used to construct an index generates considerable debate on various aspects, such as the weighting method used, possible correlation among the different sub-indices, missing variables, standardisation procedures and others.

This paper will attempt to propose some desirable criteria for the construction of composite indices, including simplicity, ease of comprehension, and coverage issues and transparency. It will also discuss a number of methodological considerations including weighting. An analysis and evaluation of the different methods used by a selection of renowned composite indices, including the University of Malta's resilience index, and the effects of certain assumptions on results will also be carried out.

*Department of Economics, University of Malta

Paper prepared for the INTERNATIONAL CONFERENCE ON SMALL STATES AND ECONOMIC RESILIENCE Valletta, Malta 23 - 25 April 2007. Organised by The Islands and Small States Institute of the Foundation for International Studies at the University of Malta and the Commonwealth Secretariat, London

## 1. Introduction

Composite indices, which are constructed by averaging a number of indicators or sub-indices, are multidimensional, in that they represent aggregate measures of a combination of factors. They are often used to simplify complex measurement constructs and to measure multi-dimensional concepts which cannot be measured by a single indicator. Composite indices often have a strong political appeal due to the fact that they simplify complex matters into a single number. In the context of policy analysis, composite indices are useful in identifying trends and drawing attention to particular issues and they can also be helpful in setting policy priorities and in benchmarking or monitoring performance.

However, composite indices are often criticised due to their subjectivity as there is an existence of a wide range of methodological approaches to composite indicators. Indeed the methodology used to construct an index generates considerable debate on various aspects, such as the weighting method used, possible correlation among the different sub-indices, missing variables, standardisation procedures and others, as the results of composite indices are sensitive to different methodological choices used in their computation.

This paper will attempt to propose some desirable criteria for the construction of composite indices, including simplicity, ease of comprehension, and coverage issues and transparency. It will also discuss a number of methodological considerations including weighting. An analysis and evaluation of the different methods used by a selection of renowned composite indices and the effects of certain assumptions on results will also be carried out.

The paper is structured as follows. Section 2, which follows this introduction, will discuss the importance of composite indices and their main strengths and weaknesses. Section 3 gives an overview of the different frameworks of desirable attributes for developing statistics and composite indices. Section 4 discusses the main conceptual issues involved in constructing composite indices, focusing on the selection of variables, ways to deal with missing data, standardisation of variables, weighting and aggregation, and testing the composite index. Section 5 analyses the methodology behind the University of Malta's Resilience Index and the reasons for the methodological decisions taken. It also evaluates the methodology of three renowned composite indices, namely the Commonwealth Secretariat's Vulnerability Index, the United Nations' Human Development Index and Fraser Institute's Economic Freedom of the World Index. Section 6 concludes the study.

## 2. Composite Indices: Definition, Uses, Strengths and Weaknesses

### *2.1 Definition and Uses*

A composite index, $I_c = \sum_{j=1}^{m} w_j X_{cj}$, is a weighted (linear) aggregation of a number of variables, where $w_j$ is a weight, with $0 \leq w_j \leq 1$ and $\sum w_j = 1$; $X_{cj}$ is the variable of country $c$ in dimension $j$; and, for any country $c$ the number of policy variables are equal to j=1,…,m.

Composite indicators are increasingly being used to make cross-national comparisons of country performance in specified areas such as the economy, environment, globalisation, society and innovation/technology/information. They are popular in benchmarking exercises where countries wish to measure their performance relative to other countries and are used to identify general trends, determine performance targets and set policy priorities.

Renowned composite indices include the Economic Competitiveness Index of the International Institute for Management Development, the Commonwealth Secretariat's Economic Vulnerability Index, the Economic Freedom of the World Index of the Fraser Institute, the Environmental and Performance Index of the Universities of Yale and Columbia, the Growth Competitiveness Index and the Current Competitiveness Index of the World Economic Forum, the Human Development Index of the United Nations and the Summary Innovation Index of the European Commission. Other composite indices include stock market indices, which represent a statistical compilation of the share prices of a number of representative stocks, and retail prices indices, used to measure inflation. The most common measure of an economy's output and value added, the Gross Domestic Product (GDP), can also be considered to be a composite index.

### *2.2 Strengths and Weaknesses of Composite Indices*

There are many studies on the strengths and weaknesses of composite indices [See, for example, Saisana and Tarantola (2002), Briguglio (2003), Conway (2005), Lievesley (2005)].

#### *2.2.1 Strengths of Composite Indices*

A composite index summarises complex or multi-dimensional issues, often yielding a single-value measure of the issue under consideration. It thus facilitates the task of ranking countries on complex issues in a benchmarking exercise and can assess the progress of countries over time on complex issues, since it is easier to interpret than trying to find a trend in many separate indicators. Composite indices can help to develop a common language for discussion and are an effective tool for communicating with policy makers and the public.

Given that quantification requires pre-definition of the issue or issues, composite indices help to set the direction for policymakers and to focus the discussion. They support decision making as they help justify certain priorities and can be used to set targets, establish standards and also promote accountability. They are thus essential for empirical work on the linkages between policy and performance. Indices may help disseminate information and can be used to make the public more aware of certain problems, for communication and for alerting stakeholders about issues. Composite indices may also generate academic discussion and enhance awareness among scholars on the issues involved.

### *2.2.2 Weaknesses of Composite Indices*

Indices share a number of weaknesses, principally associated with the subjectivity in their computation, especially in the choice of variables and the weighting procedure. Indeed, composite indices may send misleading policy messages if they are poorly constructed or misinterpreted and may invite simplistic policy conclusions. They may also be misused, e.g. to support a desired policy, if the construction process is not transparent and if the methodology lacks sound statistical or conceptual principles. The selection of indicators and weights could be the target of political challenge.

Composite indices are averages of different sub-indices and the single value which they produce may conceal divergences between the individual components or sub-indices, possibly hiding useful information. Furthermore, a composite index may require some form of trade-off between the sub-indices of the composite index and averaging would conceal, for example, situations where the effect of one variable cancels out the effect of another. It may thus disguise serious falling in some dimensions and increase the difficulty of identifying proper remedial action and, may lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored. Moreover, measurement problems may arise due to absence of data for certain variables or for certain countries, different methods of statistical compilation across countries and errors in measurements of the variables.

## 3. Overview of Quality Frameworks for Developing Statistics and Composite Indices

There have been several calls for a framework for classifying and evaluating composite indices. Drewnowksi (1972: p. 77) claimed that one requires some 'ordering principles for the selection of useful indicators and rejection of ill-conceived and inapplicable ones'. Wish (1986: pp. 97-98) similarly argued that 'indicators require a systematic rationale for categorisation'.

Thus, for a composite index to be operational and to have support by policymakers, the public and the academic community, it should have certain desirable attributes. Several organizations and individual researchers have defined desirable attributes for statistics and indicators (see IMF, 2003; Eurostat; OECD, 2004; Booysen, 2002). The main features of these frameworks are described briefly in the following sections. As will be seen, they are fairly similar in content and structure. Although useful to consider when constructing composite indices, additional desirable features and considerations are required. These can be found in Briguglio (1992, 1995, 1997 and 2003), which are described briefly below. Also described are the ten steps required to construct a composite index, proposed by JRC-OECD (2005).

### *3.1 IMF (2003): Data Quality Assurance Framework*

The IMF (2003) has developed the 'Data Quality Assurance Framework' (DQAF), to assess the overall quality of statistics produced by its member countries. The DQAF assesses how the quality of statistics is affected by the legal and institutional environment and available resources and whether there exists quality awareness in managing statistics activities. The IMF assesses the overall quality of statistics produced by its member countries based on the following five dimensions, namely (1) assurance of integrity - the features which support firm adherence to objectivity in the production of statistics; (2) methodological soundness – how current practices relate to internationally agreed methodological practices for specific statistical activities; (3) accuracy and reliability – the adequacy of the source data statistical techniques, etc… to portray the reality to be captured; (4) serviceability – the way in which users' needs are met in terms of timeliness of the statistical products, frequency, consistency and revision cycle; and (5) accessibility – whether effective data and metadata are easily available to data users and whether there is assistance to users.

### *3.2 Eurostat Framework*

According to JRC-OECD (2005), the Eurostat framework is based on seven dimensions, namely: (1) relevance – whether the data are what the user expects; (2) accuracy – whether the figures are reliable; (3) comparability – whether the data are in all necessary respects comparable across countries; (4) completeness – whether the domains for which statistics are available reflect the needs expressed by users; (5) coherence – whether the data are coherent with other data; (6) timeliness and punctuality – whether the user receives the data in time and according to pre-established dates; and (7) accessibility and clarity – whether the figure is accessible and understandable.

### *3.3 OECD (2003): Quality Framework and Guidelines for OECD Statistics*

The OECD's 'Quality Framework and Guidelines for OECD Statistics' (OECD, 2003) is built on seven dimensions, namely (1) relevance – a careful evaluation and selection of basic data have to be carried out to ensure that the right range of domains is covered in a balanced way, implying that relevance has to be evaluated considering the overall purpose of the indicator; (2) accuracy – the degree to which basic data correctly estimate or describe the quantities or characteristics that they are designed to measure; (3) credibility – refers to the fact that the data are perceived to be produced professionally in accordance with appropriate statistical standards and policies that are transparent, and implying that other things being equal, data produced by official sources are preferred to other sources; (4) timeliness – reflects the length of time period between their availability and the event or phenomenon they describe; (5) accessibility – reflects how readily the data can be located and accessed from original sources; (6) interpretability – reflects the ease with which the user may understand and properly use and analyse the data; and (7) coherence – reflects the degree to which they are logically connected and mutually consistent.

### *3.4 Booysen (2002): Dimensions for Classifying and Evaluating Development Indicators*

Booysen (2002) lists seven general dimensions for classifying and evaluating development indicators, which can be applied to other types of indicators. These are: (1) content – the aspects that the indicator measures; (2) technique and method – the method in which the indicator measures the concept, that is, quantitative (qualitative), objective (subjective), cardinal (ordinal), or uni-dimensional (multi-dimensional) manner; (3) comparative application – whether the indicator compares the concept (a) across space (cross-section) or time (time-series), and (b) in an absolute or relative manner; (4) focus – whether the indicator measures the concept in terms of input (means) or output (ends); (5) clarity and simplicity – how clear and simple the indicator is in its content, purpose, method, comparative application and focus; (6) availability – how readily available are data on the particular indicator across time and space; and (7) flexibility – how relatively flexible the indicator is in allowing for changes in content, purpose, method, comparative application and focus.

### *3.5 Briguglio (various): Desirable Characteristics for Developing Vulnerability Indices*

The desirable characteristics suggested by Briguglio (1992, 1995, 1997 and 2003) refer to composite indices, rather than individual statistics and indicators as described above. In his research on vulnerability indices, he suggested that the desirable features of a composite index are simplicity and ease of comprehension, affordability, suitability for international and temporal comparisons and transparency. He states that one of the advantages of simplicity is ease of comprehension by decision takers and other users of the index. It also permits replication by third parties for evaluation and verification. This is related to affordability, which implies that data must be relatively easy to obtain and to process. Preferably, the data should be collected as a matter of routine in line with the information required for the management of a country. Suitability for international and temporal comparisons implies that the index is based on variables which are measured in a homogenous manner, internationally and temporally. Transparency requires that the methodology used should be clearly explained by those constructing the index. This is essential for validation, evaluation and quality control purposes.

### *3.6 JRC-OECD (2005): Handbook on Constructing Composite Indicators*

The 'Handbook on Constructing Composite Indicators', developed by JRC-OECD (2005), states that the construction of a composite index involves ten steps. These are:

1. Theoretical framework: A theoretical framework should be developed to provide the basis for the selection and combination of single indicators into a meaningful composite index, clearly defining the phenomenon to be measured and its sub-components. Ideally, this process would be based on what is desirable to measure and not which indicators are available.

2. Data selection: The quality of basic data chosen to build the composite indicator strongly affects its accuracy and credibility. Indicators should be selected on the basis of their analytical soundness, measurability, country coverage, relevance to the phenomenon being measured and relationship to each other. Because there is no single definitive set of indicators for any given purpose, the selection of data to incorporate in a composite index can be quite subjective. Due to scarcity of full sets of comparable quantitative data, qualitative data from surveys or policy reviews are often used in composite indices.

3. Multivariate analysis: An exploratory analysis should investigate the overall structure of the indicators, assess the suitability of the data set, deciding whether the structure of the composite index is well-defined, if the set of the available indicators is sufficient or appropriate to describe the phenomenon, and explain the methodological choices, e.g. weighting and aggregation. Methods include: principle components analysis, which is available by using a covariance or correlation matrix; factor analysis, which is based on particular statistical models; and, the Cronbach coefficient alpha (c-alpha), which is the most common method to check for internal consistency of items in a model or survey. Cluster analysis serves as: a method of aggregation of the indicators; a diagnostic tool for exploring the impact of the methodological choices made during the construction phase of the composite index; a

method of disseminating information on the composite index; a method for selecting groups of countries to impute missing data with a view to decreasing the variance of the imputed values.

4. Imputation of missing data: The approaches for dealing with missing values include: data deletion - omitting entire records (for variables or countries) when there is a substantial number of missing data; mean substitution - substituting a variable's mean value computed from available cases to fill in missing values; regression - using regressions based on other variables to estimate the missing values; multiple imputation - using a large number of sequential regressions with indeterminate outcomes, which are run multiple times and averaged; nearest neighbour - identifying and substituting the most similar case for the one with a missing value; or ignore them - take the average index of the remaining indicators.

5. Normalisation: Normalisation of the different indicators is required to adjust the different variables on dimensions such as size/population/income and smoothened through time against cyclical variability as well as to put the different variables on a common basis. Commonly used methods for normalising indicators include: ranking of indicators across countries, standardisation (or z-scores), rescaling, distance to a reference country, categorical scales, indicators above or below the mean, method of cyclical indicators and percentage of differences over consecutive time points.

6. Weighting: No uniformly agreed methodology exists to weight individual indicators before aggregating them. Although equal weighting is a commonly used method for weighting, different weights may be assigned to component series in order to reflect their economic significance. Weights may be derived either from statistical models (principal components analysis, data envelopment analysis, regression analysis, unobserved components models) or based on public/expert opinion (budget allocation, public opinion, analytic hierarchy process, conjoint analysis). No matter which method is used, the assignment of weights involves essentially value judgments.

7. Aggregation: The literature of composite indicators offers several examples of aggregation techniques. The most commonly used are additive techniques that range from summing up country rankings in each indicator to aggregating weighted normalised indicators. Yet, additive aggregations imply requirements and properties, both of the indicators and of the associated weights, which are often not desirable and at times difficult to meet or burdensome to verify. To overcome these difficulties the literature proposes other, and less widespread, aggregation methods such as multiplicative (e.g. geometric) aggregations or non-compensatory aggregations, such as the multi-criteria analysis.

8. Sensitivity analysis: Sensitivity analysis is the study of how output variation in models such as a composite indicator can be apportioned, qualitatively or quantitatively, to different sources of

variation in the assumptions (Saltelli et al., 2004). In addition, it measures how the given composite indicator depends on the information that composes it. Meanwhile, the objective of uncertainty analysis is to quantify the overall variation in the countries' ranking resulting from the uncertainties in the model input. In the field of building composite indicators, uncertainty analysis is more often adopted than sensitivity analysis (Jamison and Sandbu, 2001) and the two types of analysis are almost often treated separately.

9. Link to other measures: The relevance and interpretability of the results can be strongly reinforced by the comparison between the composite indicator and other well known and "classical" measures of relevant phenomena. In addition, the credibility of the indicator can benefit by its capacity to produce results which are highly correlated with the reference data.

10. Visualisation: Composite indicators can be visualised or presented in a number of different ways, which can influence their interpretation. The presentation of composite indices and their visualisation affects both relevance and interpretability of the results. Composite indices should be transparent and be able to be decomposed back into their underlying indicators or values. They must be able to communicate a picture to decision-makers and other end-users quickly and accurately. This can be done using simple tabular tools or more complicated multi-dimensional graphics and interactive software. Performance can be displayed e.g., using a) absolute levels, b) absolute growth rates, e.g., in percentage points with respect to the previous year or a number of past years, c) indexed levels and d) indexed growth rates. Trends in country performance as revealed through a composite indicator can be presented through trend diagrams.

JRC-OECD (2005) argues that each phase of the composite index building process is important. The design of the theoretical framework can affect the relevance of the index; the multivariate analysis is important to increase its reliability; the imputation of missing data, as well as the normalisation and the aggregation procedure, can affect its accuracy, etc… Furthermore, JRC-OECD (2005) state that, while each step is extremely important, so is the coherence of the whole process. Choices made in one step can have important implications for other steps. Therefore, the composite indicator developer has not only to make the most appropriate methodological choices in each step, but also to identify if they fit well together.

### *3.7 Desirable Attributes of Composite Indices*

As can be seen from the previous section, the quality frameworks for collecting and constructing statistics and indicators are fairly similar. However, although there are some overlaps in the desirable attributes, some frameworks include some attributes which others omit. Therefore an assimilation of all the desirable attributes listed by the frameworks described above, and which are considered to be useful for composite index development, is carried out below in order to develop a complete list of desirable attributes for developing composite indices.

The list is made up of 8 desirable attributes as follows. They are listed in what the author believes to be a decreasing order of importance.

1. Accuracy – the degree to which the composite index correctly estimates or describes the quantities or characteristics that it is designed to measure;
2. Simplicity and ease of comprehension - the ease with which the user may understand and properly use and analyse the composite index;
3. Methodological soundness – that there is a logical connection between the different sub-indices and that that their methodology is mutually consistent and based on sound conceptual principles;
4. Suitability for international and temporal comparisons – that the index is based on variables which are measured in a homogenous manner, internationally and temporally;
5. Transparency – how available the methodology upon which the composite index was constructed is available;
6. Accessibility – how readily available the composite index is across time and space;
7. Timeliness and frequency – reflects the length of time period between the publication of the composite index and the event or phenomenon it describes and the frequency with which the composite index is published, especially if the concept it describes is not a static one;
8. Flexibility – how relatively flexible the composite index is in allowing for changes in content, purpose method, comparative application and focus.

## 4. Conceptual Issues in Constructing Composite Indices

This section will analyse the main conceptual issues in constructing composite indices and suggest ways to analyse the results of composite indices and how to communicate implications. The aspects to be considered are (1) Indicator selection; (2) Dealing with Missing Data; (3) Normalisation; (4) Weighting and Aggregation; and (5) Testing and Reviewing the Results Obtained.

### *4.1 Indicator Selection*

#### *4.1.1 Define the Concept*

The strengths and weaknesses of composite indices largely derive from the quality of the underlying variables, which summarise complex information of value to the observer. Before one starts to select the indicators to construct the composite index, one has to start by obtaining a precise definition of the concept to be measured. Then, on the basis of that precise definition, a researcher should search for suitable indicators to measure the defined concept.

#### *4.1.2 Select Indicators which Satisfy Desirable Attributes*

Indicators should be selected according to their desirable attributes. Thus, indicators should be selected on the basis of their analytical soundness, measurability, country coverage, relevance to the phenomenon being measured and relationship to each other.

#### *4.1.3 Do Not Select Variables which Beg the Question*

It is important that when a composite index is designed to prove a hypothesis or some other relationship, it does not include among the indictors, those variables or relationships which it was designed to prove.

#### *4.1.4 Draft an Initial Indicator Set and Review Available Data*

It should be noted that while the choice of indicators must be guided by the theoretical framework for the composite index, the data selection process can be quite subjective as there may be no readily available indicators to measure the phenomenon in question. Prior to the search for indicators, it is useful to draft a tentative indicator set, i.e. an ideal set of indicators irrespective of their actual or potential availability. Every effort should be made to retain on this list indicators that are deemed important, even though the data may not be available and a researcher may have to rely on proxy variables.

#### *4.1.5 Keep the Number of Variables as Small as Possible but not Fewer than Necessary*

The number of variables making up a composite index should be as small as possible. This is due to various reasons, one of them being the fact that there is an element of trade off between the richness of information and the ease of communication. Indeed, the more comprehensive a composite index is, the weaker it may be in adequately reflecting country performance. Another reason is that combining too much information from diverse areas risks becoming meaningless. Furthermore, it can also be argued that there is a trade-off between the number of indicators and the cost of obtaining the information, with too many indicators rendering the composite index unaffordable (see Figure 1). However, this does not imply that the composite index should have fewer indicators than necessary. Paraphrasing Albert Einstein, indicator sets should be as simple as possible, but not simpler (Bossel, 1999). The composite index must be made up of a comprehensive and compact set of variables, covering all relevant aspects, suggesting that a composite index has an optimal number of indicators (see Figure 2).

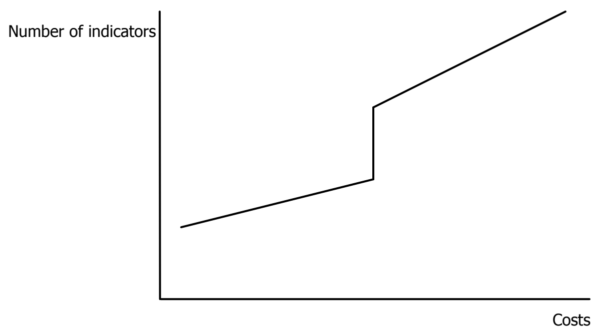


**Figure 1: Cost of Obtaining Indicators**

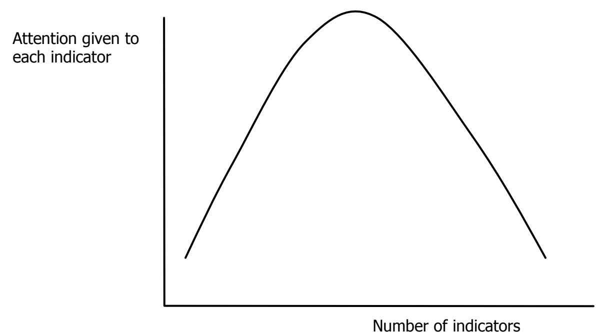


**Figure 2: Optimal Number of Indicators**

The number of variables used can be reduced by principal components analysis (PCA), a geometric method that reduces the number of variables by creating a new set of variables that are linear combination of the existing variables. It transforms correlated variables into a new set of uncorrelated variables using a covariance matrix or a correlation matrix. The objectives of PCA are: (1) dimensionality reduction; (2) the determining of linear combinations of variables; (3) feature selection: the choosing of the most useful variables; (4) visualisation of multidimensional data; (5) identification of underlying variables; (6) identification of groups of objects or of outliers (Nass, 1999). PCA cannot always reduce a large number of original variables to a small number of transformed variables. Indeed, if the original variables are uncorrelated then, the analysis is of no value. On the other hand, a significant reduction is obtained when the original variables are highly correlated – positively or negatively.

Factor analysis (FA) is also used as a tool in attempts to reduce a large set of variables into a smaller set of variables. It is similar to PCA but it is based on a particular statistical model (Spearman, 1904). FA assumes that the data is based on the underlying factors of the model and that the data variance can be decomposed into that accounted for by common and unique factors.

Similar to FA is correspondence analysis, a descriptive/exploratory technique designed to analyse simple two-way and multi-way tables containing some measure of correspondence between the rows and columns, which is, however, better suited for qualitative data. For a comprehensive description of this method, computational details and its applications, see Greenacre (1984). An extension of simple correspondence analysis to more than two variables is called multiple correspondence analysis.

*4.1.6 Check for Correlation between the Variables or Sub-Indices*

When one develops a composite index there is a risk of an element of overlap in what the different variables attempt to measure, especially if the different variables are made up of sub-indices. It is thus useful to carry out a rank correlation test to check for correlation between the different variables. If there is a high correlation between any two or more variables, it is suggested that one of the variables is discarded. This principle ties in with the principle of having a small number of variables, which helps in the operational function of the index as well as in the ease of comprehension of the index. If the composite index is made up of some highly correlated variables, this may cause the index to be biased in favour of these variables, as it implies that a higher weight is attached to these variables. Variables that are uncorrelated below a certain threshold should also discarded as redundant.

An alternative way to investigate the degree of correlation among a set of variables is the Cronbach coefficient alpha (c-alpha). If the correlation is high, then there is evidence that the indicators are measuring the same underlying construct. If the reliability coefficient increases after deleting a sub-indicator from the scale, one can assume that the sub-indicator is not highly correlated with other sub-indictors in the scale.

Cluster analysis and discriminant analysis can also be used to reduce correlation to avoid redundancy. Cluster analysis is a multivariate procedure for detecting natural groupings in data and is sometimes used to aggregate the data in a composite index. It is based upon the placing of objects into more or less homogeneous groups, in a manner such that the relationship between groups is revealed. Cluster analysis lacks an underlying body of statistical theory and is more objective than subjective (Wulder, 2005). Homogeneous and distinct groups are delineated based upon assessment of distances or in the case of Ward's method, an F-test (Davis, 1986).

Another method is discriminant analysis, which can be used either to assess the adequacy of classification given group memberships of the objects under study, or used to assign objects to one of a number of (known) groups of objects. Although discriminant analysis is relatively robust to nonnormality due to skewness, it is highly sensitive to outliers. Variables with significant outliers necessitate transformation prior to analysis. Linearity is also assumed for discriminant analysis (Wulder, 2005).

*4.1.7 Review the Indicators Selected and Seek External Advice and Opinion*

When the data set is selected, it may be useful to seek external advice and opinion on the chosen indicators.

### *4.2 Dealing with Missing Data*

When constructing a composite index comprising several different variables and a large number of countries, it is inevitable that some indicators will be unavailable for some countries. It is thus important to analyse what can be done in such a situation. Methods to deal with missing data can be split into two categories: single imputation, which substitutes a value for each missing value and multiple imputation, which replaces each missing value with a set of plausible values about the right to impute. [For an extensive literature on the analysis of missing data, see Little and Schenker (1994), Little (1997) and Little and Rubin (2002), see Yuan (2000), Rubin (1987), Lavori et al. (1995), Schafer (1997), Robsenbaum and Rubin (1983), Chantala and Suchindran (nd).]

#### *4.2.1 Exclude the country from the analysis*

One way to deal with missing data is to exclude the country from the composite index construction if it includes an unavailable observation. The argument in favour of such a rule is that a missing observation may cause the results to be biased in favour of the other available indicators and may render the composite indicator values incomparable with other countries. Another argument in favour of excluding countries with missing indicator values is that aggregating only the available indicators can also negatively affect the credibility of the composite index as an analyst would have to check which indicators are available and which are not and it would be difficult to compare the results across time and space. The disadvantage with such an approach is that the researcher ends up with a smaller sample. For this reason imputation methods are sometimes applied.

#### *4.2.2 Single Imputation Methods*

As mentioned earlier, single imputation methods substitute a value for each missing value. A brief description of the main single imputation methods, namely case deletion, mean/median/mode estimation, cold and hot deck imputation and regression imputation, raw maximum likelihood and expectation maximisation imputation, is provided below. This list is by no means exhaustive.

##### *4.2.2.1 Case Deletion*

Case deletion simply omits the missing records from the analysis. However, this approach ignores possible systematic differences between complete and incomplete samples and may produce biased estimates.

##### *4.2.2.2 Cold Deck Imputation*

Cold deck imputation recovers an observation by checking whether the observation is available for a previous year. This option is useful in composite indices whose values are not expected to change much over time. This method is a very popular missing data imputation procedure.

#### *4.2.2.3 Mean/Median Mode Estimation*

Mean/median/mode estimation replaces missing data with the mean of non missing values. The disadvantage is that the standard deviation and standard errors are underestimated.

#### *4.2.2.4 Hot Deck Imputation*

Hot deck imputation involves stratifying and sorting the data by key covariates, and then replacing missing data from another record in the same strata. More simply, it involves analysing the dataset and checking whether there is a similar country with the similar characteristics, and then replacing the missing indictor with the indicator available in the similar country. However, here again, underestimation of standard errors can be a problem.

#### *4.2.2.5 Regression Imputation*

Regression imputation imputes each independent variable on the basis of other independent variables in the model but may produce biased estimates. It is also likely to over fit the data and result in correlations to be unrealistically high. Also, for every country the missing observation is conditioned by the observations in other countries.

#### *4.2.2.8 Drawbacks of Single Imputation Methods*

In general, single imputation results in the sample size being over estimated with the variance and standard errors being underestimated. Graham et al. (2003) referred to single imputation methods as "unacceptable methods". Multiple imputation methods were developed in order to overcome these problems.

### *4.2.3 Multiple Imputation Methods*

Contrary to single imputation methods, which substitute a value for each missing value, multiple imputation, replaces each missing value with a set of plausible values about the right to impute. The main types of multiple imputation methods are the regression method, the propensity score method and the Markov Chain Monte Carlo algorithm. A brief description of each is provided below.

*4.2.3.1 Parametric Regression Method*

In the parametric regression method, a regression model is fitted for each variable with missing values, with the previous variables as covariates. Based on the resulting model, a new regression model is then fitted and is used to impute the missing values for each variable (Rubin, 1987). This method is useful for monotone missing data patterns.

*4.2.3.2 Propensity Score Method*

The propensity score is the conditional probability of assigning to a particular treatment given a vector of observed covariants (Rosenbaum and Rubin, 1983). In the propensity score method, a propensity score is generated for each variable with missing values to indicate the probability of that observation being missing. The observations are then grouped based on these propensity scores, and an approximate Bayesian bootstrap imputation (Rubin, 1987) is applied to each group (Lavori, Dawson and Shera, 1995).

*4.2.3.4 Markov Chain Monte Carlo Algorithm*

In the Markov Chain Monte Carlo algorithm, one constructs a Markov Chain – a sequence of random variables in which the distribution of each element depends on the value of the previous one – long enough for the distribution of the elements to stabilise to a common distribution (Schafer, 1997). By repeatedly simulating steps of the chain, it simulates draws from the distribution of interest. This method is useful for an arbitrary missing data pattern.

*4.2.4 Analysis of Imputation Methods*

The imputation of missing data affects the accuracy of the composite index and its credibility (see Allison, nd). Furthermore, even if timeliness can be improved, extensive use of imputation techniques can undermine the overall quality of the indicator and its relevance. Regression coefficients for predictors with large fractions of imputed data will show substantial biases towards zero (Landerman et al., 1997).

Roth (1994), Little and Rubin (1987) and Wothke (1998) reviewed different imputation methods and concluded that case deletion and mean substituting missing data handling methods are inferior when compared with multiple imputation methods. Regression methods are somewhat better but not as good as hot deck imputation.

It should be observed that multiple imputation theories are still relatively new and are still being developed. Although at present there is still some scepticism about this methodology, it is important to state that the superiority of multiple imputation to traditional methods is based on mathematical fact, not belief or opinion (Wayman, 2003). One can calculate the efficiency of multiple imputation using a ratio developed by Rubin (1987), which analyses the relative increase in variance due to nonresponse.

### *4.2.5 Quantifying Qualitative Data*

Sometimes quantitative data may not be available for some indicators or else may be restricted to limited country coverage and only qualitative information may be obtained. For this reason, a researcher may have to transform qualitative data into a quantitative format. One way in which this can be carried out is by categorising an occurrence (in terms of intensity or frequency) along a Likert scale. The scale can be for example from 1 to 7, with 1 being the lowest possible occurrence and 7 the highest possible. The wider the spread of the scale, the more possible will it be to derive meaningful standard deviations of the averages obtained, but there is a limit to how many meaningful categories one can work with (Briguglio, 2003). This approach also permits non-linearity, such as for example, in cases where the occurrence grows or declines exponentially or when it takes a U-shaped or S-shaped pattern. It should be noted that however linear mapping is the most common procedure. The main defect of this method relates to the subjectivity of the category groupings and the choice between linear and non-linear relationships.

## ***4.3 Normalisation***

Since the indicators which make up a composite index very rarely have the same units, indicators should be standardised, i.e. converted to a similar unit, in order to render them comparable. Freudenberg (2003) and Jacobs et al (2004) list a number of normalisation methods. It should be noted that the selection of a suitable method is not trivial and deserves special attention (Ebert and Welsh, 2004). The normalisation method should take into account the data properties, as well as the objectives of the composite indicator. Different normalisation methods will yield different results and normalisation may reduce the difference between results if there are large outliers. Robustness tests might be needed to asses their impact on the outcomes. If a composite index is made up of a number of sub-indices and these sub-indices are normalised, it may be useful to re-standardise the composite index after aggregation has been carried out. The two most common methods of normalisation are rescaling and standardisation.

Rescaling is perhaps the most common normalisation method. It normalises indicators between the range (0,1) by means of the following formula: $XS_{ij} = \frac{\left(X_{ij} - MinX_j\right)}{\left(MaxX_j - MinX_j\right)}$, where $XS_{ij}$ is the value of the normalised observation for country *i* of component *j*, $X_{ij}$ is the actual value of the same observation, $MinX_j$ and $MaxX_j$ are the minimum and maximum values of the same observations for component *j*.

Standardisation (or z-scores) is very similar to the above method and converts indicators to a common scale with a mean of zero and standard deviations of one as follows: $XS_{ij} = \frac{\left(X_{ij} - \overline{X_{ij}}\right)}{\left(\sigma_j\right)}$, where $\overline{X_{ij}}$ is the average and $\sigma_j$ of the observation across countries.

Other normalisation methods, which are not recommended by the author as they are deemed inferior to the above described methods, include expressing the observations as percentage differences over consecutive years, using ratios, using rankings instead of the actual observation values, measuring the relative position of a given indicator vis-à-vis a reference point and transforming observations such that values around the mean receive 0, whereas the ones above/or below a certain threshold receive 1 and -1 respectively.

The normalisation phase is crucial both for the accuracy and the coherence of final results. An inappropriate normalisation procedure can bring about unreliable or biased results. On the other hand, the interpretability of the composite indicator heavily relies on the correctness of the approach followed in the normalisation phase.

### *4.4 Weighting and Aggregation*

One of the key issues in the construction of composite indices is the choice of the weighting and aggregation model. Almost all quality dimensions are affected by this choice, especially accuracy, coherence and interpretability (JRC-OECD, 2004). This is also one of the most criticised characteristics of composite indices and the one which generates most debate. The greatest debate lies in whether equal weights or differential weights are to be used, and if the latter is chosen, how to derive the differential weights. Aggregation/weighting questions have been extensively studied in the literature on productivity indices (Balk, 2002). This section will provide an analysis of equal and differential weights as well of the main theories that can be used to derive differential weights.

#### *4.4.1 Equal Weighting*

Most composite indices rely on equal weighting, i.e., all variables are given the same weight. This could be a result of the fact that all variables making up the composite index are deemed to be of equal importance to the concept to be measured, but it could also be a result of lack of consensus on an alternative or insufficient knowledge. When using equal weights, it may happen that, by combining variables with a high degree of correlation, one may introduce an element of double counting into the index. It is thus useful to test the indicators for statistical correlation, for example with the Pearson correlation coefficient (Manly, 1994) and choosing only indicators exhibiting a low degree of correlation or giving less weight to correlated indicators. Furthermore, minimizing the number of variables in the index using the methods described in Section 4.1.5 may be desirable on other grounds such as transparency and parsimony.

### *4.4.2 Differential Weighting*

It is often argued that equal weights are render the concept too simplistic and that instead indicators should be weighted and aggregated according to the underlying theoretical framework of the concept being measured. The OECD (2003) states that "greater weight should be given to components which are considered to be more significant in the context of the particular composite indicator". It should be noted that when equal weights are applied, if the variables are grouped into components and those are further aggregated into the composite, then applying equal weights to the variables may imply an unequal weighting of the component (JRC-OECD, 2004).

### *4.4.3 Country-Specific or Indicator-Specific Weights*

Problems arise in the determination of differential weights and whether they should be country-specific or indicator-specific. In the case of the former one can argue that country-specific weights render the composite index incomparable between different countries. On the other hand, indicator-specific weights may imply that although an indicator may have less socio-economic and/or political implications for one country compared to another, it will have to be given the same importance in the composite index according to the weight applied.

### *4.4.4 Weights Over Time: Constant or Changing?*

With regard to the time element, keeping weights unchanged across time might be justified if the researcher is willing to analyse the evolution of a certain number of variables. If instead, the objective of that analysis is that of defining best practices or that of setting priorities, then weights should necessarily change over time. In the construction of price indices, a Laspeyres index is used for constant weights, while a Paasche index is used for changing weights.

*4.4.5 Weights reflecting the Statistical Quality of the Data*

Weights may also be chosen to reflect the statistical quality of the data. Higher weights could be assigned to statistically reliable data with broad coverage. However, this method could be biased towards the readily available indicators, penalizing the information that is statistically more problematic to identify and measure.

*4.4.6 Regression Method*

In deriving weights using the regression method, one uses as a dependent variable a proxy of the composite index and this is regressed on a number of explanatory variables which represent the components of the composite index. The coefficients on the explanatory variables of the estimated equation are taken as weights for averaging the components of the index. Since this approach lets the data produce the weights, it does not require normalisation of the observations.

The procedure has a number of methodological defects, which limit the operationality and the reliability of a composite index aggregated using this method. The most important methodological defect is that if the dependent variable is considered to be a proxy for the variable to be indexed, one need not go through a cumbersome regression procedure to compute the index (Briguglio, 2003). Other defects are that the regression may produce negative coefficients, which would imply the use of negative weights; the regression may produce weights which are very small, almost significant and comparatively relatively large coefficients, which would also imply the use of weights which are very small and weights which are comparatively large in the same index. Another defect is that since the coefficients pertain to data with different units and varying distributions, it is not possible to estimate the weight of each variable in the composite index.

*4.4.7 Stochastic weights*

This technique was developed by Anders Hoffmann, ex-OECD and a co-author of the Handbook on Constructing Composite Indicators (JRC-OECD, 2004) and it generates sets of random weights each of which sums to 1. The main defect of this procedure is that the weights are generated in a procedure which assigns too much value to chance. Also, this procedure does not assign a greater weight to components which are considered to be more significant in the context of the particular composite index.

*4.4.8 Participatory Methods*

Participatory methods can be used to assign weights, either those that incorporate various stakeholders (Moldan and Billharz, 1997) and those that make use of public opinion polls (Parker, 1991). In the budget allocation approach, experts are given a "budget" of N points, to be distributed over a number of sub-indicators, "paying" more for those indicators whose importance they want to stress (Jesinghaus in Molden and Bilharz, 1997).

### *4.4.9 Weights Based on the Precautionary Principle*

Closely related to the above, that is, based on expert opinion, is the determination of weights based on the precautionary principle. In this procedure, experts assign differential weights to the various components and a large weight is assigned to that component, which is expected to be crucial to attaining the phenomenon the composite indicator is attempting to measure.

### *4.4.10 Benefit-of-the Doubt Weighting System*

The "benefit-of-the-doubt" weighting system, proposed by Melyn and Moesen (1991), chooses the weights such that the evaluated country has a maximal composite index value. Also referred to as endogenously-weighted composite indicators, the method is based on the Data Envelopment Analysis (DEA) method (Farrel, 1957; Charnes, Cooper and Rhodes, 1978). The core idea is that a country's relatively good performance in some dimensions is indicative of the fact that this country considers the concerned policy dimensions as relatively more important (Van Pyenbroeck, 2005). This method has a high political acceptance as no other weighting scheme yields a higher composite index value. The principal is also easy to communicate: if another country, say Country B, gets a higher overall score using Country A's assigned weighting scheme, this implies that Country B is outperforming Country A. The "benefit-of-the-doubt" approach is useful when individual expert opinion is available, but when experts disagree about the right set of weights.

A possible criticism of the benefit-of-the-doubt approach is that it makes performance 'look better' than what it really is, since the selected weights can deviate from the 'true' but (unknown) priorities. The method also does not exclude extreme scenarios where all the relative weight is assigned to a single indicator, which would then completely determine the overall index value. Some restrictions can be imposed as in Cherchye et al. (2004) where they did not allow the sum of weights in each category to exceed the sum of weights in another category by more than 20 per cent.

### *4.4.11 Linear and Geometric Aggregation*

Aggregation methods also vary. While the linear aggregation method is useful when all sub-indicators have the same measurement unit, geometric aggregations are better suited if non-comparable and strictly positive sub-indicators are expressed in different ratio-scales. Furthermore, linear aggregations reward base-indicators proportionally to the weights, while geometric aggregations reward those countries with higher scores (JRC-OECD, 2004).

*4.4.12 Aggregation Methods and Weighting Systems*

In both linear and geometric aggregations, weights express trade-offs between indicators. A shortcoming in one dimension thus can be offset (compensated) by a surplus in another. This implies an inconsistency between how weights are conceived (usually they measure the importance of the associated variable) and the actual meaning when geometric or linear aggregations are used. In a linear aggregation, the compensability is constant, while with geometric aggregations compensability is lower for the composite indicators with low values.

The assumption of preference independence is essential for the existence of a linear aggregation rule Munda and Nardo (2003). Thus from a mathematical point of view, given the variables, $X_1$, $X_2$, …, $X_n$, an additive aggregation function exists if and only if these variables are mutually preferentially independent.

In terms of policy, when geometric aggregation is used, a country with low scores on one indicator will need a much higher score on the others to improve its situation, implying that in benchmarking exercises, countries with low scores prefer a linear rather than a geometric aggregation. Also, a country would be more interested in increasing those sectors/activities/alternatives with the lowest score in order to have the highest chance to improve its position in the ranking if the aggregation is geometric rather than linear.

*4.4.13 Non-Compensatory Multi-Criteria Aggregation*

If one wants to assure that weights remain a measure of importance, other aggregation methods should be used, in particular methods that do not allow compensability, such as a non-compensatory multi-criteria approach (MCA). In its basic form, this approach does not reward outliers, however this method could be computationally costly when the number of countries is high, as the number of permutations to calculate increases exponentially (Munda, 2005).

*4.4.14 Weighting and Aggregation Decisions: Subjective Choices*

The absence of an "objective" way of determining weights and aggregation methods does not necessarily lead to rejection of the validity of composite indicators, as long as the entire process is transparent. The objectives must be clearly stated at the outset, and the chosen model must be checked to see to what extent it fulfils the goals. No matter which technique is used weights are effectively value judgements.

McGillivray and Noorbakhsh (2004) argue that differential component weights, which they declare is appropriate on conceptual grounds is a rather fruitless exercise. They state that such weighting produces index values which are generally indistinguishable from values of the equally weighted index.

## *4.5 Testing and Reviewing the Results Obtained*

### *4.5.1 Uncertainty and Sensitivity Analysis*

When one constructs a composite index, it is useful to test the robustness of that composite index, as this depends on a number of factors including the amount of missing data, the choice of the imputation algorithm and the choice of weights. This is usually done by means of uncertainty and sensitivity analysis, the iterative use of which during the development of a composite index could improve its structure (Saisana et al., 2005a; Tarantola et al., 2000).

Uncertainty analysis focuses on how the sources of uncertainty propagate through the structure of the composite index and affect its values. It can be carried out by performing multiple evaluations based on the Monte Carlo approach proposed by Saisana et al. (2005a, 2005b) in order to take into account all uncertainty sources simultaneously to capture all possible effects among input factors. These include the inclusion and exclusion of sub-indicators, modelling data error based on the available information on variance estimation, using alternative editing and normalisation schemes and using different weighting and aggregation schemes. The results of the robustness analysis are generally reported as country rankings with their related uncertainty bounds. Sensitivity analysis studies how much each individual source of uncertainty contributes to the composite index value/ranking variance. The two types of analysis are almost always treated separately but uncertainty analysis is more often adopted than sensitivity analysis (Jamison and Sandbhu, 2001). The results of a sensitivity analysis are often shown as scatterplots with the values of the composite indicator for a country on the vertical axis and each input source of uncertainty on the horizontal axis.

### *4.5.2 Outliers*

It is also useful to check the index results for any outliers by either a visual inspection of the data in the table or by plotting the data in scatter diagram. A large outlier can be due to an error in inputting the data. It should be noted that large outliers can bias the results when carrying out normalisation. However, the exclusion of outliers is also subjective and may also generate an amount of bias.

### *4.5.3 Expert Opinion*

Expert opinion or knowledge may lead one to conclude that the results of the composite index do not reflect the reality of the country or the concept to be measured.  Thus the index results should be carefully reviewed to detect any inconsistencies in the results obtained and to verify that they correctly portray economic and social realities.

#### *4.5.4 Analysing the Results Obtained*

It should always be kept in mind that the methodology chosen to construct the index has important implications on the results obtained.  Thus, when one analyses the scores and/or rankings of a composite index, it is important to analyse not just the final results obtained but also the results of the sub-indices or the sub-components of the index, to assess whether the results obtained are a result of similar performance in all areas, or whether one area is a strong determinant of the country performance.  It also important to be aware of the methodological choices made in developing a composite index, so that one is aware of any inconsistencies and methodological flaws when one interprets index results.

## 5. An Analysis of the Methodology of Renowned Composite Indices

This section outlines the methodology used in constructing the University of Malta's Economic Resilience index and provides justification for the use of such methodology. It also presents an analysis of three renowned composite indices, namely the Commonwealth Vulnerability Index, the United Nation's Human Development Index and Fraser Institute's Economic Freedom of the World Index. The way the analysis will be carried out is three fold: (1) the objective of the composite index will be highlighted; (2) a brief outline of the methodology used will be given; and, (3) a critical analysis of the methodology employed will be carried out, focussing on weighting and aggregation, normalisation and imputation of missing data.

### *5.1 University of Malta Economic Resilience Index*

#### *5.1.1 Objective of the Index*

The University of Malta's Economic Resilience Index (Briguglio et al., 2006) is a composite index which attempts to measure economic resilience, where economic resilience is defined as the "nurtured" ability of an economy to recover from or adjust to the effects of adverse shocks to which it may be inherently exposed.

#### *5.1.2 Outline of the Methodology*

The resilience index is made up of 4 variables, which are linearly aggregated using equal weights. The 4 variables are: (1) macroeconomic stability (measured by the average of the standardised values of – the fiscal deficit to GDP ratio, the sum of the unemployment and inflation rates, and the external debt to GDP ratio); (2) microeconomic market efficiency (measured by the "regulation of credit, labour and business" component of the Economic Freedom of the World Index); (3) good governance (measured by the "legal structure and security of property rights" component of the Economic Freedom of the World Index); and (4) social development (measured by the education and health indicators of the UNDP Human Development Index).

#### *5.1.3 Critical Analysis of the Methodology*

The choice of variables which compose the index is somewhat subjective. It can be argued that the indicators chosen do not really measure the variables they were chosen to portray. It should be however, noted that care was taken to base the choice on a set of desirable criteria related to appropriate coverage, simplicity and ease of comprehension, affordability, suitability for international comparisons and transparency.

The imputation method chosen was a cold deck imputation, that is, if data was not available for a particular year for a particular country, then a previous year was utilised. Also, alternative sources were also consulted. Indeed, for the macroeconomic stability component, the sources listed include the IMF, World Bank and National Statistics Offices. This could negatively impact on the inter-country comparability of the statistics as the definitions of the indicators may differ (slightly) across countries. Also, it was decided that should one of the indicators making up the composite index be missing, even within the sub-components, then the resilience index was not computed for the country in question. This was done in order to ensure that there no country's score suffers from any sort of bias. It, however, did result in many countries having to be excluded from the analysis. Indeed, a considerable number of small states had to be excluded, either because the Economic Freedom of the World did not compute an index for the country (for the microeconomic market efficiency and good governance components) or because some statistics making up the macroeconomic stability component were missing. Had a more sophisticated imputation method been used, more countries could possibly have been included in the resilience index.

The index may also be criticised as being strongly dependent on the Economic Freedom of the World Index. Indeed, the microeconomic market efficiency and the good governance sub-index are both measured using components of the above mentioned index. However, it should be stated that these components were chosen because they spanned a large number of countries. An extensive search of available indicators measuring microeconomic market efficiency and attempts at producing a sub-index measuring microeconomic market efficiency for a wide range of countries was unsuccessful. An alternative measure of good governance is presented by the World Bank (Kaufmann et al., 2006). However, a Pearson correlation test of the World Bank governance indicators and the Economic Freedom of the World's "legal structure and security of property rights" component yielded a value of 0.92. Thus, both indices are measuring a similar phenomenon as countries are ranked similarly using both indicators and it was concluded that the component chosen to measure good governance was satisfactory. The economic resilience index can also be criticised as "carrying" the problems present in the Economic Freedom of the World Index (analysed in Section 5.4). However, unless new indicators are developed by the authors to measure microeconomic market efficiency and good governance, then the criticism that the resilience index carries along with it the problems of another index, will remain. Since the resilience index is calculated for a large number of countries, alternative indicators are hard to come by.

The index uses equal weights across the four components. It can be argued that since there is some correlation between the microeconomic sub-index and the good governance sub-index, and between the social development sub-index and the good governance sub-index, there can be some bias towards these indices in the resilience index. However, the authors deemed such correlation to be not unduly high and thus the four sub-indices were kept in the resilience index.

### *5.2 Commonwealth Vulnerability Index*

*5.2.1 Objective of the Index*

The Commonwealth Secretariat's Vulnerability Index for Developing Countries – the Position of Small States (Easter et al., 2000) attempts to measure the vulnerability of a country and identify vulnerable states and is used to determine whether small states should be given differential treatment by the international community.

*5.2.2 Outline of the Methodology*

The Commonwealth Vulnerability Index is based on the observation that income volatility, measured as the standard deviation of annual rates of growth of GDP per capita at constant prices between 1980 and 1992, is the most obvious evidence of vulnerability and that this is mainly determined by: (1) lack of diversification (UNCTAD diversification index of merchandise exports); (2) extent of export dependence (the average exports of goods and non-factor services as a percentage of GDP); and, (3) impact of natural disasters (proportion of the population affected by such events as estimated over a relatively long period of time).

The methodology adopted follows a two stage procedure. First, a linear regression was carried out which explained output volatility in terms of the three variables mentioned above. Second, the regression results were used to predict individual vulnerability scores for all countries for which data were available. Thus, the variable coefficients derived from the regression were used as indicator weights.

It was hypothesised that the variables affect output volatility in a different manner for small and large countries. For this reason, a dummy variable was used in connection with the variable measuring susceptibility to natural disasters and was assigned a value of one if the country had a population of 1.5 million or less and zero otherwise.

Data corresponding to 11 countries of the initial 111 developing countries was excluded in order to carry out the regression in order to avoid a few countries influencing the entire model procedure and biasing the results. The countries not included in the fitting of the model were still included in the composite index. The 11 countries excluded from the analysis consisted of 5 small states – Bahrain, Kiribati, Maldives, Malta and Vanuatu – and six large states – Chad, Mexico, Myanmar, Iran, Rwanda and Singapore.

*5.2.3 Critical Analysis of the Methodology*

The regression method used to define the differential weights can be criticised on a number of fronts. If historical volatility of income is used as the underlying measure of external shocks, it can be argued that there is no need to construct a vulnerability index as one can use volatility of income as a proxy. This points has also been made by Briguglio (2003) who states that the most important methodological defect of weights being determined through the regression method is that if the dependent variable is considered to be a proxy for the variable to be indexed, one need not go through a cumbersome regression procedure to compute the index.

When one carries out a regression in order to determine the weights of the indicators making up the composite index, one runs the risk of ending up with unreasonable weights. Since the regression was carried out using 'raw' data, the coefficients of the variables were based on different units and different scales and so are not comparable at face value. Thus, it is almost impossible to determine, for instance, whether economic exposure has a higher weight in the vulnerability index than lack of export diversification.

Another aspect is that fact that vulnerability to natural disasters was only included for those countries which were classified as small states even though there are many larger countries which have a high vulnerability to natural disasters, resulting in the vulnerability score for large countries to be lower than that for small countries, questioning the comparability of the vulnerability scores across countries.

The omission from the analysis of a number of countries for variety of reasons has also been criticised (see Crowards, 2000). He states that the justifications provided for omitting the eleven countries is far from compelling and they limit the scope for the exercise to be repeated in the future. Furthermore, he argues that the omission of particular countries from the analysis determining the structure of the underlying model represents a crude solution to a criticism that is wrongly applied to the process of normalisation used in other studies.

Regarding missing data, it appears that the imputation procedure carried out was cold-deck imputation. Given that vulnerability is an inherent feature and not policy induced, the factors which determine vulnerability are not likely to change over time, especially in the short term. Thus, such an imputation method is a reasonable one. Regarding the normalisation procedure, since 'raw' data was used for the regression, no normalisation was carried before aggregation.

### *5.3 Human Development Index*

#### *5.3.1 Objective of the Index*

The Human Development Index (HDI) measures the average achievements in a country in three basic dimensions of human development: (1) a long and healthy life; (2) knowledge; and (3) a decent standard of living.

*5.3.2 Outline of the Methodology*

As outlined above, the HDI is a summary measure of human development and, using an equal weighting system, measures the average achievements in a country in 3 basic dimensions of human development: (1) a long and healthy life (measured by life expectancy at birth); (2) knowledge (measured by the adult literacy rate and the combined primary, secondary and tertiary gross enrolment ratio); and (3) a decent standard of living (measured by GDP per capita in purchasing power parity – PPP – terms in US dollars).

Standardisation was carried out using the re-scaling procedure. The formula used for the life expectancy and education indices was:

$$\text{Dimension index} = \frac{\text{actual value - minimum value}}{\text{maximum value-minimum value}}$$

The formula used for the GDP index was:

$$\text{Dimension index} = \frac{\log(\text{actual value}) - \log(\text{minimum value})}{\log(\text{maximum value}) - \log(\text{minimum value})}$$

The maximum value and minimum values are assigned in advanced, that is, not obtained from the data available. The maximum and minimum values assigned for each indicator are the following:

**Table 1: HDI – Assigned Maximum and Minimum Values**

| Indicator | Maximum Value | Minimum Value |
|---|---|---|
| Life expectancy at birth (years) | 85 | 25 |
| Adult literacy rate (%) | 100 | 0 |
| Combined gross enrolment ratio (%) | 100 | 0 |
| GDP per capita (PPS US$) | 40,000 | 100 |

In response to the desire of countries to be included in the HDI table, and in line with the goal of including as many UN member countries as possible, estimates from other international, regional or national sources were used when the primary international data agencies lack data for one or two HDI components for a country. In a few cases, estimates were produced, but were clearly documented.

*5.3.3 Critical Analysis of the Methodology*

The three components that make up the HDI have equal weights. The knowledge variable, however, is made up of two indicators – adult literacy and the combined primary, secondary and tertiary gross enrolment ratio. The first indicator has a weight of two-thirds, while the second indicator has a weight of one-third. Thus, the weights assigned to each indicator are the following:

**Table 2: HDI Variable Weights**

| | |
|---|---|
| *Long and Healthy Life* | |
| Life expectancy at birth | 1/3 |
| *Knowledge* | |
| Adult literacy rate | 2/9 |
| Combined primary, secondary and tertiary enrolment ratio | 1/9 |
| *Decent Standard of Living* | |
| GDP per capita in purchasing power standards | 1/3 |

As indicated in the previous section, the values for the maximum and minimum used in the normalisation formula were assigned values, not the actual maximum and minimum values. This does not change the countries' ranking but it does not allow the best performing country to obtain a value of 1 and the worst performing country to obtain a value of 0. Sometimes, analysts may find it useful to look over standardised values between 0 and 1, because at a glance they can tell which the best and worst performing countries are. The maximum and minimum values obtained from the standardisation formula used in the compilation of the HDI are the following:

**Table 3: HDI – Actual Maximum and Minimum Values**

| **Index** | **Maximum Value** | **Minimum Value** |
|---|---|---|
| Life expectancy at birth | 0.95 | 0.10 |
| Education | 0.99 | 0.23 |
| GDP per capita | 1.00 | 0.29 |

As can be seen the assigned maximum and minimum values in the standardisation formula affected the minimum index values rather than the maximum index values, which in the three cases are very close to or equal to 1. The pre-assignment of maximum and minimum values may also affect the distribution of the standardised values and may in some cases restrict some indicators to being smaller than others.

Since data covering a large number of countries is required to compute the HDI, it is very likely that some indicators will not be available for some countries. Thus, some form of data imputation was carried out, with the result that the data has varying quality and reliability.

Data on life expectancy at birth was largely available, except for five countries where it was imputed by means of cold deck imputation. It is likely that the cold deck imputation procedure here may bias the results slightly downwards, as it is expected that life expectancy improves over the years, thus using 'old' data may give the country a lower rank than it would otherwise have achieved.

Over half of the countries required some form of missing data imputation in order to obtain values for the adult literacy rate. Again, the most common method of imputing the data was the cold deck imputation procedure. The downward bias expected for life expectancy data is also expected in the case of the adult literacy rate. Around one-third of the countries, mainly high-income countries, no longer collect basic literacy statistics because they have already attained high levels of literacy, and were assigned a literacy rate of 99.0%. Cold deck imputation was also the main method used to impute the gross enrolment ratio variable.

Different methods were used to impute data on GDP per capita (PPP US$). For a large number of countries, over 45, data was imputed by means of the regression procedure, while in a few other cases cold deck imputation was carried out. It should be observed that regression imputation may produce biased estimates and is also likely to over fit the data.

## ***5.4 Freedom of the World Index***

### *5.4.1 Objective of the Index*

The index published in Economic Freedom of the World (EFW), published by the Fraser Institute, measures the degree to which the policies and institutions of countries are supportive of economic freedom in five areas: (1) size of government (expenditure, taxes and enterprises); (2) legal structure and security of property rights; (3) access to sound money; (4) freedom to trade internationally; and (5) regulation of credit, labour and business. The index analysed in this section will be the most recent one published, that is, the one published in the 2006 Annual Report and which refers to the year 2004.

### *5.4.2 Outline of the Methodology*

The EFW index has been constructed with a number of methodological goals in mind. First, the index needs to cover a relatively large number of countries over as lengthy a period of time as possible. Second, all data used to construct index ratings are from third-party international sources such as the IMF, World Bank, World Economic Forum and so on. Data provided directly from a source within a country are used only rarely. Third, the report aims to be as transparent as possible about the data sources, the methodology for transforming raw data into index ratings and for constructing area and summary ratings, and so on.

Within the five major areas mentioned above, 21 components are incorporated into the index, but many of those components are themselves made up of several sub-components. Counting the various sub-components, the EFW index comprises 38 distinct pieces of data. Each component and sub-component is placed on a scale from 0 to 10 that reflects the distribution of the underlying data. The component ratings within each area are averaged to derive ratings for each of the five areas. In turn, the summary rating is the average of the five area ratings.

Over the years, a number of different weighting methods ranging from the subjective views of "experts" to principal component analysis have been tried. In most cases, the choice of weighting method exerts little impact on the rating and ranking of countries. So, in an effort to keep the procedure simple and transparent, a simple average is used to combine the components into area ratings and the area ratings into summary ratings. However, this does not mean to imply that all components and areas of economic freedom are equally important. For some purposes, clearly some of the components are more important than others.

Some data from the various business surveys (18 subcomponents in total) are not available for all of the countries covered by the EFW index. Two of the areas, Size of Government: Expenditures, Taxes, and Enterprises (Area 1) and Access to Sound Money (Area 3), are unaffected by the omitted variables. The omissions, however, could be important in Legal Structure and Security of Property Rights (Area 2) and Regulation of Credit, Labor, and Business (Area 5) and, to a lesser extent, in Freedom to Trade Internationally (Area 4). In these three areas, a regression was run among the countries for which complete data was available, in order to find out how the omission of the survey data affects the area rating. The dependent variable was the area rating with the survey data and the independent variable was the area rating excluding the survey data. The regression estimates were used to adjust the area ratings for the countries without survey data and, thereby, make them more comparable with the ratings of the countries for which the survey data were available. The same adjustments were performed in all years. While these statistical adjustments enhance the overall comparability among the countries, comparisons between the nations that have the survey data and the nations that do not should be made with a degree of caution.

*5.4.3 Critical Analysis of the Methodology*

The EFW uses a methodology based on equal weights. However, it should be observed that since some the variables are grouped into components and these are further aggregated into the composite, then applying equal weights to the variables may imply an unequal weighting of the component. The table below shows the effective weighting system applied to the components.

## Table 4: Areas and Components of the EFW Index

| The Areas and Components of the EFW Index | Weights* |
|---|---|
| | |
| **1: Size of Government: Expenditures, Taxes, and Enterprises** | |
| A. General government consumption spending as a percentage of total consumption | 0.05 |
| B. Transfers and subsidies as a percentage of GDP | 0.05 |
| C. Government enterprises and investment as a percentage of total investment | 0.05 |
| D. Top marginal tax rate (and income threshold to which it applies) | |
| i. Top marginal income tax rate (and income threshold at which it applies) | 0.03 |
| ii. Top marginal income and payroll tax rate (and income threshold at which the top marginal rate applies) | 0.03 |
| **2: Legal Structure and Security of Property Rights** | |
| A. Judicial independence: the judiciary is independent and not subject to interference by the government or parties in disputes | 0.04 |
| B. Impartial courts: A trusted legal framework exists for private businesses to challenge the legality of government actions or regulation | 0.04 |
| C. Protection of intellectual property | 0.04 |
| D. Military interference in rule of law and the political process | 0.04 |
| E. Integrity of the legal system | 0.04 |
| **3: Access to Sound Money** | |
| A. Average annual growth of the money supply in the last five years minus average annual growth of real GDP in the last ten years | 0.05 |
| B. Standard inflation variability in the last five years | 0.05 |
| C. Recent inflation rate | 0.05 |
| D. Freedom to own foreign currency bank accounts domestically and abroad | 0.05 |
| **4: Freedom to Trade Internationally** | |
| A. Taxes on international trade | |
| i. Revenue from taxes on international trade as a percentage of exports plus imports | 0.01 |
| ii. Mean tariff rate | 0.01 |
| iii. Standard deviation of tariff rates | 0.01 |
| B. Regulatory trade barriers | |
| i. Non-tariff trade barriers | 0.02 |
| ii. Compliance cost of importing and exporting | 0.02 |
| C. Actual size of trade sector compared to expected size | 0.04 |
| D. Difference between official exchange rate and black market rate | 0.04 |
| E. International capital market controls | |
| i. Foreign ownership/investment restrictions | 0.02 |
| ii. Restrictions on the freedom of citizens to engage in capital market exchange with foreigners—index of capital controls among 13 IMF categories | 0.02 |
| **5: Regulation of Credit, Labor, and Business** | |
| A. Credit Market Regulations | |
| i. Ownership of banks: percentage of deposits held in privately owned banks | 0.01 |
| ii. Competition: domestic banks face competition from foreign banks | 0.01 |
| iii. Extension of credit: percentage of credit extended to private sector | 0.01 |
| iv. Avoidance of interest rate controls and regulations that lead to negative real interest rates | 0.01 |
| v. Interest rate controls: interest rate controls on bank deposits and/or loans are freely determined by the market | 0.01 |
| B. Labor Market Regulations | |
| i. Impact of minimum wages | 0.01 |
| ii. Hiring and firing practices: hiring and firing practices of companies are determined by private contract | 0.01 |
| iii. Share of labor force whose wages are set by centralized collective bargaining | 0.01 |
| iv. Unemployment Benefits: the unemployment benefits system preserves the incentive to work | 0.01 |
| v. Use of conscripts to obtain military personnel | 0.01 |
| C. Business Regulations | |
| i. Price controls: extent to which businesses are free to set their own prices | 0.01 |
| ii. Burden of regulation | 0.01 |
| iii. Time with government bureaucracy: senior management spends a substantial amount of time dealing with government bureaucracy | 0.01 |

| | |
|---|---|
| iv. Starting a new business: starting a new business is generally easy | 0.01 |
| v. Irregular payments: irregular, additional payments connected with import and export permits, business licenses, exchange controls, tax assessments, police protection, or loan applications are very rare | 0.01 |

**Weights have been rounded up to 2 decimal places. These weights apply in cases where all data is available.*

Thus, as can be seen from the table above, a certain element of unequal weights has been carried out implicitly and which is important to note, especially when one analyses the results of the index. For example, the indicator transfers and subsidies as a percentage of GDP has a weight which is five times as high as the weight for the mean tariff rate.

Although much detail has been given on the missing data imputation procedure used for Areas 2, 4 and 5, there appears to be no information on the imputation procedure used for areas 1 and 3. An analysis of Area 1 indicates that there are many countries which have missing data. Indeed out of 130 countries, just 58 have a complete data set. The way the EFW dealt with this missing data problem was by means of the case deletion procedure, that is missing records were simply omitted from the analysis. Say if a country had 3 of the 5 indicators available, then the average was carried out on the basis of 3 values rather than 5. This is a rather simplistic way of imputing missing data is considered to a low quality imputation procedure, as it ignores possible systematic different between complete and incomplete samples and may produce biased estimates. All countries had complete data sets for Area 3.

An analysis of Area 2, in which 23 countries had incomplete data sets, shows that all countries, except 3 had area averages that were lower than they would have been had the case deletion imputation procedure been carried out. The other 3 countries had the same area average as they would have obtained had the case deletion procedure been carried out. Thus, for this area, there is little risk that the averages are biased upwards.

Area 4 is made up of 5 sub-indices and the procedure used to deal with missing data within each sub-index, where there was at least one indicator available, was case deletion, with the exception of Areas 4C and 4D, which were made up of just one indicator and thus no imputation could be carried out. Imputation by means of regression estimation was carried out in order to derive the area score. In this case, all 21 countries that required missing data imputation obtained scores which were higher had the case deletion procedure been carried out, implying that there could be some element of upward bias in the scores obtained.

Area 5 is made up of 3 sub-indices, and like Area 4, the procedure used to deal with missing data within each sub-index, all of which were made up of more than one variable, was case deletion. However, no overall score for Area 5B was issued for those countries which did not have any data on unemployment insurance (Component 5Biv). Similarly for Area 5C, no overall score was issued for those countries whose available information was restricted to solely to component 5Ci (price controls) and/or component

5Civ (starting a new business). Like Area 4, the regression imputation was carried out in order to derive the overall score for Area 5. When the results were compared to the overall score obtained by case deletion, it transpired that 18 countries achieved lower scores under the regression procedure and 3 higher. Thus, we can conclude that there is no danger of any upward bias taking place. It should be noted that regression imputation is likely to over fit the data and result in correlations to be unrealistically high, however, this aspect was not tested for this index. Furthermore, comparisons between the nations that have the complete data set and the nations that do not should be made with a degree of caution. Also, for every country the missing observation is conditioned by the observations in other countries.

The normalisation procedure used for quantitative factors was the rescaling formula: $I_F = \frac{(V_i - V_{\min})}{(V_{\max} - V_{\min})} \times 10$, where $I_F$ is the sub-index for the relevant factor, $V_i$ is the value of the factor for the country in question, $V_{max}$ is the maximum value for the parameter in question and $V_{min}$ is the minimum. For qualitative factors, scores were assigned a value of 0 to 10, where 0 represents the lowest score and 10 the highest. These normalisation procedures are simple, transparent and easy to understand.

## 6. Conclusion

As we have seen above, composite indices have their pros and cons. Saisana et al., (2002) state that in practice, it is hard to imagine that debate on the use of composite indicators will ever be settled. All things considered, composite indicators should be identified for what they are – simplistic presentations and comparisons of performance in given areas to be used as starting points for further analysis.

However, the importance of composite indices should not be undermined. If an index is built according to the desirable attributes described in this paper and if it is based on sound methodological choices and there is transparency in the documentation of an index, then this will make an index a valuable measure and one that can be relied upon to portray a complex phenomenon, that is unable to be portrayed using a single indicator, in a simplistic manner.

## References


Allison, P., (nd). *Multiple Imputation for Missing Data: A Cautionary Tale*, University of Pennsylvania.

Balk, B., (2002). *The Residual. On Monitoring and Benchmarking Firms, Industries and Economies with respect to Productivity*, EIA-2002-07-MKT, Erasmus Research Institute of Management.

Booysen, F. (2002). *An Overview and Evaluation of Composite Indices of Development*, Social Indicators Research, 59(2): 115–51.

Bossel, H. (1999). *Indicators for Sustainable Development: Theory, Method, Applications – A Report to the Balaton Group*, International Institute for Sustainable Development, Canada.

Briguglio, L. (1992). *Preliminary Study on the Construction of an Index for Ranking Countries According to their Economic Vulnerability*, UNCTAD/LDC.Misc.4.

Briguglio, L. (1995). *Small Island States and their Economic Vulnerabilities*, World Development, 23:1615-1632.

Briguglio, L. (1997). *Alternative Economic Vulnerability Indices for Developing Countries*, Report prepared for the United Nations Department of Economic and Social Affairs.

Briguglio, L. (2003). *The Vulnerability Index and Small Island Developing States: A Review of Conceptual and Methodological Issues*, Paper Prepared for the AIMS Regional Preparatory Meeting on the BPoA+10 Review, 1-5 September 2003, Praia, Cape Verde.

Briguglio, L., Cordina, G., Farrugia, N., and Vella, S., (2006). *Conceptualising and Measuring Economic Resilience*, in Building the Economic Resilience of Small States, Briguglio L., Cordina G., and Kisanga E., ed.

Chantala, K., and Suchindran, C., (nd). *Multiple Imputation for Missing Data*, Lecture Notes.

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

Cherchye, L., Moesen W.and Van Puyenbroeck T. (2004). *Legitimately Diverse, Yet Comparable: on Synthesising Social Inclusion Performance in the EU*, Journal of Common Market Studies, 42, 919-955.

Conway, P. (2005). *OECD Composite Indicators: Pros and Cons, Weighting, Sensitivity, Updating Issues*, Paper presented at the Workshop on European Indices and Scoreboards, 24-25 October 2005, Brussels.

Crowards, T., (2000). *A Critique of the Commonwealth Vulnerability Index*, Caribbean Development Bank, Staff Working Paper No. 4/00.

Davis, J. (1986). *Statistics and Data Analysis,* in Geology , John Wiley & Sons, Toronto, 646p.

Drewnowski, J. (1972). *Social indicators and welfare measurement: remarks on methodology*, Journal of Development Studies 8, pp. 77–90.

Easter, C.D., Atkins, J.P. and Mazzi, S., (2000). *A Commonwealth Vulnerability Index for Developing Countries: The Position of Small States*, Economic Paper 40, Commonwealth Secretariat, London.

Ebert, U. and Welsh, H., (2004). *Meaningful environmental indices: a social choice approach*, Journal of Environmental Economics and Management 47, 270-283.

Farrell, M.J. (1957). *The Measurement of Productive Efficiency*, Journal of the Royal Statistical Society 120(3):253-290.

Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). *Methods for handling missing data*. In J. A. Schinka & W. F. Velicer (Eds.). Research Methods in Psychology (pp. 87-114). Volume 2 of Handbook of Psychology (I. B. Weiner, Editor-in-Chief). New York: John Wiley & Sons.

Greenacre, M. J. (1994). *Multiple and joint correspondence analysis*. In Correspondence Analysis in the Social Sciences - Recent Developments and Applications (Edited by M. Greenacre and J. Blasius). Academic Press.

IMF (2003). *Data Quality Assessment Framework and Data Quality Programme*, Fifth Review of the Fund's Data Standards and Initiatives, Prepared by the Statistics Department (in consultation with other departments).

Jacobs, R. P. Smith and M. Goddard, (2004). *Measuring performance: an examination of composite performance indicators*, Centre for Health Economics, Technical Paper Series 29.

Jamison, D.T. and Sandbu, M.E. (2001). *WHO Ranking of Health System Performance*, Science, 293, (31 August 2001), 1595-6.

JRC-OECD (2005). *Handbook on Constructing Composite Indices: Methodology and User Guide*, OECD Statistics Working Paper.

Kaufmann D., Kraay A., and Mastruzzi M. (2006). *Governance Matters V: Governance Indicators for 1996-2005*, World Bank.

Landerman, Lawrence R., Kenneth C. Land and Carl F. Pieper (1997). *An Empirical Evaluation of the Predictive Mean Matching Method for Imputing Missing Values*, Sociological Methods & Research 26: 3-33.

Lavori PW, Dawson R, Shera D (1995): *A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data*. Stat Med 14:1913-1925.

Lievesley, D. (2005). *The Politics of Performance Indicators*, Paper presented at the Workshop on European Indices and Scoreboards, 24-25 October 2005, Brussels.

Little R.J.A. and Schenker N. (1994). *Missing Data*, in Handbook for Statistical Modeling in the Social and Behavioral Sciences (G. Arminger, C.C Clogg, and M.E. Sobel eds.) pp.39-75, New York: Plenum.

Little R.J.A (1997). *Biostatistical Analysis with Missing Data*, in Encyclopedia of Biostatistics (p. Armitage and T. Colton eds.) London: Wiley.

Little R.J.A. and Rubin D.B. (2002). *Statistical Analysis with Missing Data*, Wiley Interscience, J. Wiley &Sons, Hoboken, New Jersey.

Manly B. (1994). *Multivariate statistical methods*, Chapman & Hall, UK.

McGillivray M. and Noorbakhsh F. (2004). *Composite Indices of Human Well-Being*, World Institute for Development Economics Research (WIDER), Research paper No. 2004/03.

Melyn W., and Moesen W., (1991). *Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available*, Public Economic research Paper 17, CES, KU Leuven.

Moldan, S. Billharz, Eds. (1997). *Indicators of Sustainable Development*, 420 pp. John Wiley, Chichester.

Munda, G., and M. Nardo, (2003). *On the Methodological Foundations of Composite Indicators Used for Ranking Countries*. Universitat Autonoma de Barcelona, Dept. of Economics and Economic History, Barcelona, Spain.

Munda, G., (2005). *Aggregation Issues: Non-Compensatory Multi Criteria Analysis*, Paper prepared for the Workshop on European Composite Indicators and Scoreboards, Brussels, 24-25 October 2005.

Nass, P., (1999). *Multivariate Analysis Methods*, Lecture Notes.

OECD (2004). *Quality Framework and Guidelines for OECD Statistical Activities*, www.oecd.org/statistics.

Parker J. (1991). *Environmental reporting and environmental indices*, PhD Dissertation, Cambridge, UK.

Rosenbaum, P. R., and Rubin, D. B., (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, Biometrika 70, 41-55.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: John Wiley & Sons.

Saisana, M. and Tarantola, S. (2002). *State-of-the-Art Report on Current Methodologies and Practices for Composite Indicator Development*, EUR 20408 EN.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., (2004). *Sensitivity Analysis in Practice*, Wiley, New York.

Saisana M., Tarantola S. and Saltelli A. (2005a). *Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators*, Journal of the Royal Statistical Society A, 168(2), 1-17.

Saisana M., Nardo M. and Saltelli A. (2005b). *Uncertainty and Sensitivity Analysis of the 2005 Environmental Sustainability Index*, in Esty D. Levy M., Srebotnjak T. and de Sherbinin A. 2005 Environmental Sustainability Index: Benchmarking National Environmental Stewardship. New Have: Yale Center for Environmental Law and Policy, p.75-78.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman and Hall.

Spearman, C. (1904). *General intelligence, objectively determined and measured*. American Journal of Psychology, 15, 201-293.

Tarantola S., Jesinghaus J. and Puolamaa M. (2000). *Global sensitivity analysis: a quality assurance tool in environmental policy modelling. In Sensitivity Analysis* (eds Saltelli A., Chan K., Scott M.) pp. 385-397. New York: John Wiley & Sons.

Van Puyenbroeck, T., (2005). *Composite Indicators and Data Envelopment Analysis*, Paper presented at the Workshop on European Indices and Scoreboards, Brussels 24-25 October 2005.

Wayman, J.C., (2003). *Multiple Imputation for Missing Data: What is it and How can I use it*? Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, IL.

Wish, N.B. (1986). *Are we really measuring the quality of life?*, American Journal of Economics and Sociology 45, pp. 93–99.

Wothke, W. (1998). *Longitudinal and multi-group modeling with missing data*. In T. D. Little, K. U. Schnabel, and J. Baumert [Eds.]. Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples.. Mahwah, NJ: Lawrence Erlbaum Publishers.

Wulder, M., (2005). *A Practical Guide to the Use of Selected Multivariate Statistics*, Canadian Forest Services, Canada.

Yuan, Y., (2000). *Multiple Imputation for Missing Data: Concepts and New Development*, SAS Institute Inc., Rockville, MD.