Silvia Terzi, Adrian Otoiu, Elena Grimaccia, Matteo Mazziotta, Adriano Pareto

OPEN ISSUES IN COMPOSITE INDICATORS

A Starting Point and a Reference on Some State-of-the-Art Issues

3 Collana
Dipartimento di Economia

RomaTrE-Press
2021

Università degli Studi Roma Tre
Dipartimento di Economia

Collana
Dipartimento di Economia

n. 1, ottobre 2020
Economics, Policy and Law
Proceedings of the Research Days Department of Economic
A cura di Alessia Naccarato, Pasquale De Muro, Saverio M. Fratini

n. 2, gennaio 2021
Research Lines of the Department of Economics
A cura di Attilio Trezzini

Università degli Studi Roma Tre
Dipartimento di Economia

Silvia Terzi, Adrian Otoiu, Elena Grimaccia, Matteo Mazziotta, Adriano Pareto

OPEN ISSUES IN COMPOSITE INDICATORS

A Starting Point and a Reference on Some State-of-the-Art Issues

3 Collana
Dipartimento di Economia

RomaTrE-Press
2021

Dipartimento di Economia
Collana
n. 3, febbraio 2021

Open Issues in Composite Indicators

A starting point and a reference on some state-of-the-art issues

Authors
Silvia Terzi, Adrian Otoiu, Elena Grimaccia, Matteo Mazziotta, Adriano Pareto

Coordinamento editoriale
Gruppo di lavoro *RomaTrE-Press*

Cura editoriale e impaginazione
teseo editore Roma teseoeditore.it

Elaborazione grafica della copertina
MOSQUITO mosquitoroma.it

Caratteri grafici utilizzati: Minion Pro, Awake (copertina e frontespizio); Garamond, Times New Roman, Cambria Math (testo).

Edizioni *RomaTrE-Press* ©
Roma, febbraio 2021
ISBN 979-12-5977-001-1
http://romatrepress.uniroma3.it

FORTE
FONDAZIONE ROMA TRE-EDUCATION

L'attività della *RomaTrE-Press* è svolta nell'ambito della Fondazione Roma Tre-Education, piazza della Repubblica 10, 00185 Roma.

Preface

The idea for this book stems from a series of seminars on composite indicators held at Roma Tre University, in Rome, by Adrian Otoiu. During and after the seminars, we, the authors of this book, found ourselves discussing the questions posed by Adrian during his presentation, open problems and our mutual points of view. It was natural to continue the discussion, exchange bibliographical references and, above all, to put together our different academic experiences and approaches to the study and construction of composite indicators.

The book is designed for those who already have a basic knowledge of both statistics and the construction of indicators, and it basically follows our debate.

We had in mind the issues encountered by practitioners, and especially those that tend to be left aside by the publications that are an established reference in the field, such as the OECD Handbook on constructing composite indicators. Our intention was to complement these references and provide an insight into what we consider to be major developments in the field, which all interested readers should be aware of, that are likely to shape the development of the field of composite indicators in the near future, especially in relation to their use in the field of well-being and human progress.

After the well-known and very notable OECD's Handbook published in 2008, other eminent scholars, published collections of articles on composite indicators.

However, the interest in the subject is far from over. On the contrary, many efforts in the construction of composite indicators have also been made in order to develop assessment tools, and rating systems, with the aim of measuring and monitoring complex and multidimensional phenomena, such as sustainability, economic and social development and well-being. These assessments provide multiple benefits, such as enabling policymakers to track progress towards achieving goals, improving transparency and accountability, raising awareness and facilitating better-informed decision-making.

The development of different frameworks aimed at monitoring the sustainability and well-being of populations – like the UN Sustainable Development Goals, the OECD Better Life Initiative and Eurostat's Quality of Life Dashboard – gave a further boost to the research devoted to composite indicators, due to the need to synthesize the huge amount of data and indicators that, without proper methodological treatment, would be of no, or scarce, utility.

In this book we want to contribute to the debate and to the understanding of some of the still unsolved issues concerning composite indicators.

Composite indicators provide a single measure for a not directly observable multidimensional phenomenon, based on indicators or variables. They are very useful because they can effectively rank countries, and regions, providing information to both policymakers and the public since the results are easily understood.

However, the idea of summarizing complex phenomena into single numbers is not straightforward. It involves both theoretical and methodological assumptions that need to be assessed carefully to avoid producing results of dubious analytic rigour. The construction of a composite indicator can be seen as an obstacle course, from the availability of data to the choice of the individual indicators to their treatment in order to compare and aggregate them. Therefore, criticisms could grow along with each step of the process.

Despite the growing interest on the construction of composite indicators, the main issues are far from being fully explored and/or solved, so along with the development of different frameworks there has been increasing attention to connected issues such as data-driven weights, subjective approaches, dichotomous variables, association sensitivity and inequality.

Our choice has been to provide a book that focuses on some of the topics that are still being debated without any definitive solutions. In the first part of the book, we illustrate methodologies that reflect the current state of knowledge, while in the second part we disentangle different recent and more critical issues. In this way, we provide a useful tool both for researchers that have limited specific knowledge on the subject and for scholars that need an updating on the most recent and advanced topics on composite indicators.

There is no part of composite indicator construction that cannot be questioned. Nevertheless, the advantages of composite indicators are clear, and they can be summarized as unidimensional measurement of the phenomenon, easy interpretation with respect to a set of many individual indicators and simplification of the data analysis.

Part one of the book deals with the theoretical framework, the measurement model, normalization, aggregation and weights, including data-driven weights.

Chapter 3 opens the second part of the book with a discussion about subjective and objective well-being, a key topic when assessing human progress. Following a general introduction about how this distinction has developed over time, a discussion on subjective well-being and the contemporary issues is carried out in the context of composite indicators. Some salient examples have been chosen to illustrate the issues related to subjective well-being in general, and in the use of composite indicators.

Chapter 4 is dedicated to composite indicators for dichotomous variables and the counting approach, an approach also suited to aggregate ordinal and continuous variables in the same composite indicator. All that is needed is to set a threshold (which could be a deprivation benchmark or, vice versa, a sustainable goal). A most valuable feature of the Alkire-Foster dual-cut-off counting approach that underlies the Multidimensional Poverty Index is that the derived composite indicator embeds information concerning the association between the different dimensions.

Chapter 5 looks in detail at some major practical issues relating to the actual construction of composite indicators: weighting, aggregation and substitutability. In the first part, a review of these topics is carried out, with the purpose of providing the main elements of the issue in a practical manner: fewer formulae and more insights. It is a review of reference works and review papers that also casts some light on aspects that deserve more focused attention and explanation for the readers, and especially for those that have consulted the previously mentioned works.

The issue of substitutability is dealt with in Chapter 6. This notion is typical of the construction of composite indicators, where a variable or indicator that composes the composite indicator may or may not be compensated by the value of another variable.

The last two chapters focus on inequality and how to include it in the composite indicator, while also distinguishing between horizontal dispersion and vertical inequality. Vertical inequality is measured across units (within each individual indicator taken separately), whereas horizontal inequality (or dispersion) is measured across dimensions (for each unit at a time). Several methodologies aimed at including the measure of inequality in the construction of composite indicators are presented.

The book is theoretically sound but is written in a simple way in order to meet the needs of a large audience: researchers, developers, scholars, university students and public officials who are interested in gaining a better understanding of composite indicators.

The book is the result of a joint effort, but single paragraphs or sometimes whole chapters are mostly written by one or two authors. Nevertheless, credit and responsibility are to be attributed to all of us.

Contents

PART 1

COMPOSITE INDICATORS: THE BUILDING BLOCKS

Chapter 1

Concepts and Measures*

1.1 Variable, Indicator and Index

Over time, literature has produced many definitions that hardly manage to converge towards a shared assumption on the difference between indicator and index, to which, for clarity, it is also appropriate to add the concept of variable. It is certainly not the task of this book to find a solution to a problem so debated over the years by illustrious exponents, however we want to arrive at common-sense definitions that can guide us in reading especially the more analytical parts so that there are no misunderstandings.

The term "variable" derives from the late Latin *variabilis*, meaning "that varies, which tends to vary". In statistics, we mean a characteristic found in one or more statistical units belonging to a population or a sample as a result of a direct survey. Nowadays this definition seems almost reductive since we live in the era of the deluge in which we are surrounded by big data or administrative sources used for statistical purposes and, therefore, not only by direct surveys. In more general terms, the variable is defined as a characteristic associated with a statistical unit such as income, age or beds in a hospital.

The term "indicator" comes from the late Latin *indicator*, meaning "who indicates" or "element that indicates or signals something". In the economic field, we indicate some macroeconomic quantities (employment rate, inflation rate, GDP per capita, etc.) that are considered significant for the purpose of evaluating the performance of the economy in a given country and in a given time. The social indicator is a value, mostly empirical, with which we want to measure, in a given situation, significant variations in behaviour and social conditions (e.g.,

* This chapter is mostly written by Matteo Mazziotta and Adriano Pareto.

the number of supermarkets or telephone users per thousand inhabitants). While any normalized variable can be called an "indicator", we usually mean that the indicator is a ratio and that, therefore, it is composed of a numerator, which can be considered the variable that provides the meaning, and a denominator, which is the variable that allows its comparability in space and/or over time. Let's take the example of wanting to compare the health-care infrastructure of two European countries, Germany and the Netherlands: if we considered the variable "number of hospital beds" it would make no sense to make the comparison because, obviously, Germany will have more sleeping places than Holland. If, on the other hand, we calculate the indicator "number of beds per 1,000 inhabitants", we obtain a normalized value according to the population variable and then we can make a comparison between the two countries. In other words, the presence of a reference denominator (the population) makes it possible to transform a variable (the number of beds) into a relative measure and therefore it becomes comparable in time and over space.

The term "index" derives from the Latin *index* and means anything that serves to indicate. By index we mean the synthetic expression of the components of a given phenomenon in time and/or space, or the relationship between multiple phenomena considered at the same time. We could therefore define an index as a complex structure in which the components are linked together by mathematical operations. For example, the Gini index is a measure of the inequality in the distribution of a transferable character (e.g., income) among the N units of a population. The different formulae present in the literature for its calculation provide different solutions, for example the use of concentration quantity ratios (p_i and q_i) and the areas of the Lorenz curve. Another example is the body mass index (BMI) of a person, which is calculated as the ratio between weight and the square of height (variables that have different units of measurement). Another example is the Human Development Index (HDI), which is obtained as the geometric average of the normalized indicators of three dimensions: health, education and standard of living. These three examples show us that an index differs from an indicator because it is a more complex measure, in which operations with variables or indicators have been carried out.

As previously mentioned, both the term "indicator" and the term "index" derive from late Latin and, in literature, we find elementary (or individual) indicators for expressing the components of a complex phenomenon and composite indicators for defining a combination of elementary indicators. In this book, the term "elementary (or individual) indicator" represents a component of the complex phenomenon to be measured and "composite indicator" the measure of the phenomenon itself; in fact, the composite indicator could be defined as a mathematical combination of a set of normalized variables (or elementary indicators).

Elementary indicators, therefore, are tools capable of showing (measuring) the progress of a phenomenon that is considered representative for the analysis and they are used to monitor or evaluate the degree of success or the adequacy of the activities implemented. A composite indicator is a measure, generally expressed in quantitative form and composed of several variables, capable of summarizing the trend of the phenomenon to which it refers. The composite indicator is not the phenomenon, but it represents and summarizes the behaviour of the more complex phenomenon that we must monitor and evaluate. An example that gives the idea: the indicator is the finger reaching towards the sky... but the phenomenon is the star!

1.2 Latent factor and pillar

When we define a social, economic or environmental phenomenon with a multidimensional meaning, it means that we are willing to represent it using a number of elementary indicators as parts to be composed so that this composition is the measure of the phenomenon itself. We should imagine the figurative works of Giuseppe Arcimboldo (April 5, 1527, Milan - July 11, 1593, Milan), the famous "Composite Heads", burlesque portraits executed by combining objects or elements of the same kind (fruit and vegetables, fish, birds, books, etc.) metaphorically connected to the subject represented, in order to sublimate the portrait itself. Fruit and vegetables, considered individually, are fruit and vegetables and nothing more. If instead, as happens in the masterpieces of Arcimboldo, they are positioned in such a way as to compose a human face with a clear facial expression of joy or sadness, then, in this case, they are able to show us something that we would never have expected to see. We have a human face in front of us and we cannot see the particular fruit or vegetable. The painting depicts a human and not a basket of still life. The human is a latent factor.

The composite indicator can be generically thought of as "the search for latent variables starting from some observed variables". An observed variable is a variable that has actually been measured, while a latent variable (either hidden or underlying) is a type of variable that has not been measured, which is not even directly measurable and is therefore hypothesized and "analysed" through its effects. The links and relationships that a latent variable has with other measurable variables, and the influences it has on them, become a way to go back to this hidden variable (we consider it "underlying" both because it is not measurable and because we may not have measured it).

If, for example, we wanted to measure well-being we would have to define the phenomenon with a multidimensional approach and therefore assume that it is a latent factor since it is not immediately visible or representable through a univocal measure. Well-being is a phenomenon that exists, but we must interpret it through the involvement of many measurements. Precisely for this reason the definition of the phenomenon (or of the latent factor) must be very clear and shared by the working group that is preparing to measure it from a quantitative point of view.

After all, the purpose of scientific research is to subject the hypotheses derived from a basic theory to empirical corroboration, thus it is clear that the latent factor needs a process of operationalization, i.e. the definition of a concept and therefore its decomposition into objects that we will call, in fact, "individual indicators". When we calculate a composite indicator, we start from the original matrix constructed with the dimensions of the phenomenon under study; when we reduce the dimensions in space it means that we try to statistically "treat" the individual indicators by losing as little information as possible. Dimensions are the object to which we apply our work tools so that they are comparable to each other and aggregable to arrive at a single number.

Between the definition of the phenomenon (the latent factor) and the elementary indicators there is an intermediate structure, namely the pillar. It represents the concept of dimension and is more detailed than a latent factor. Usually, it is a component of the phenomenon that has the function of containing elementary indicators.

In the case of the Human Development Index (HDI), the latent factor is human development defined as the theoretical framework of the capability approach by Amartya Sen; the pillars are the dimensions that interpret the concept of the theoretical framework (long and healthy life, knowledge and a decent standard of living); and the elementary indicators (life expectancy at birth, education index and GDP per capita (PPP $)) are the operationalization – the single measurements – of the dimensions themselves. Another example is the Italian Equitable and Sustainable Well-being (BES). Some of the twelve pillars that comprise the latent factor are health, work, environment, education, economic well-being, etc.; within each of the twelve pillars there are the elementary indicators that measure the dimensions. If the latent factor were a tree, then the branches would be the pillars and the leaves would be the individual indicators.

The example of the HDI allows the definition of a complex indicator to be introduced, which is the aggregation of composite indicators, i.e., a higher level of aggregation. In fact, the pillar "knowledge" originally consists of two elementary indicators that are summarized in a single measure called the "education index". So, when we go to calculate the HDI composite indicator, in reality, we compose two elementary indicators and a composite indicator: the HDI is a complex indicator. This does not happen in the case of the Italian BES since a composite for each pillar is calculated, but a complex indicator is not provided by the National Institute of Statistics.

1.3 Types of variables

The choice of variables is one of the fundamental steps for measuring a multidimensional phenomenon and an error in this phase can invalidate the whole process of reading a complex reality. This choice must be carefully considered because the "nature" of the variable can influence the representation of the latent phenomenon. A fundamental classification for the construction of composite indicators is between input, output and outcome indicators. Not all of them are suitable for use as elementary indicators for the calculation of a composite indicator.

The input indicator refers to all those elements that make up the resources and the starting situation (and therefore also the needs and risk factors) that can be understood as a starting line on which to graft the design and programming. Generally, the resources (usually economic and financial) introduced into a process in order to obtain a result in a fixed time are considered inputs – let's say, the expense incurred to achieve a goal –.

The output represents the most immediate outcome of the spending programme. The outcomes express the impact that the spending programme has on the community and the environment. The output and outcome indicators are characterized by strong interdependencies and those who use them must still take into account the influence of any exogenous factors. Often, in the literature, there is a main distinction to be taken into account: on the one hand there are the product indicators (output) and on the other hand the indicators of effect or result (outcome). Objectives and indicators can be oriented to the products (output), or to the resulting effects (outcome), or to an appropriate combination of both. The product is what is directly created by the programme (for example a service carried out for a specific user). The effect can be considered more generally as a consequence of the product. Of course, between this 'effect' and the 'product' of which it is considered a consequence, there must be a plausible and close relationship, otherwise the exogenous factors may have created it and the researcher may come to incorrect conclusions that do not identify reality.

Let's take an example that could clarify the difference between the three types of elementary indicators. Let's imagine that in a socio-economic development programme you want to increase the employment rate of a specific geographical area; the goal is to equip a specific reference population with the skills necessary to find a job by attending professional training courses. The input indicator represents the expenditure invested in organizing the refresher course (classroom, teaching material, teachers, etc.), the product (output indicator) is represented by the people who successfully completed the courses, while the effect (the result or outcome indicator) is represented by the people who subsequently find a job, thereby increasing the employment rate.

Assuming, therefore, that we want to measure a complex phenomenon with a composite index, we must be extremely careful to insert an input indicator in the original matrix since it, presumably, does not interpret well the role of the component of the phenomenon; if the training course has a strong drop in participants or simply does not help learners find a job, then it means that the expenditure incurred and, therefore, the investment did not bring the desired results. Consequently, there is a risk of inserting an input indicator in the matrix whose information content is totally useless or, worse still, incorrect because the goal for which the expenditure was incurred has not been achieved. If we insert the output indicator in the original matrix – in our example, the learners who have finished the course with profit – that indicator will certainly be used to compose the complex phenomenon through the calculation of the synthetic index as it would be a relevant product of the expenditure sustained: participants have acquired a wealth of knowledge that can be spent on finding a job. The ideal would be to consider the outcome indicator, i.e., the final result of the expenditure incurred – in our case, the increase or decrease in the employment rate –.

Chapter 2

How to build composite indicators?*

2.1 Introduction

In recent years, the debate on the measurement of multidimensional phenomena has generated a renewed interest in the scientific community. It is common awareness that a number of socio-economic phenomena cannot be measured by a single descriptive indicator and that, instead, they should be represented by a multiplicity of aspects or dimensions.

Phenomena such as development, progress, well-being, quality of life, poverty, social inequality, etc., require, to be measured, a 'combination' of different dimensions, to be considered together as components of the phenomenon (Mazziotta and Pareto, 2013). The complex and multidimensional nature of these phenomena requires the definition of intermediate objectives whose achievement can be observed and measured by manifest variables (MVs) or individual indicators. The mathematical combination (or "aggregation" as it is termed) of a set of indicators that represent the different dimensions of a phenomenon to be measured is called a *composite indicator* (Saisana and Tarantola, 2002; Salzman, 2003; OECD, 2008).

Composite indicators are based on several individual (or elementary) indicators or sub indicators (pillars). These indicators or sub indicators are aggregated by analytical methods to give an overall score for each country or geographical area. The final result is usually called an "index" and is used to either create a ranking or to simply summarize the data (Bandura, 2008). Examples of well-known composite indices are the United Nations' Human Development Index (HDI) (UNDP, 1990, 2010) and Technology Achievement Index (TAI) (UNDP, 2001), and the European Commission's Regional Competitiveness Index (RCI) (Annoni and Kozovska, 2010).

---

* This chapter is mostly written by Matteo Mazziotta and Adriano Pareto.

However, there is no part of a composite indicator construction that cannot be questioned. In fact, the idea of summarizing complex phenomena into single numbers is not straightforward. It involves both theoretical and methodological assumptions that need to be assessed carefully to avoid producing results of dubious analytic rigour (Saisana et al., 2005). Constructing a composite indicator is a difficult task and full of pitfalls: from the obstacles regarding the availability of data and the choice of the individual indicators to their treatment in order to compare (normalization) and aggregate (weighting and aggregation) them. Therefore, criticisms could grow simultaneously regarding each step (Booysen, 2002).

Many scientists dispute the use of composite indicators that lead to computation of a single value for each geographic area, preferring the so-called 'dashboard', where all the indicators are shown individually (as in the case of monitoring the state of health of a vehicle: oil level, fuel, water temperature, etc.). This approach aims to provide an overall picture of a given phenomenon while avoiding weight assignment and loss of information from combining multiple indicators (UNECE, 2019). From the statistical point of view this is an incontrovertible choice, but from the standpoint of politics and media it is a heavy limitation. The easy disclosure in the media and the immediate understanding by the user are certainly the strengths of a unique indicator. Obviously, both approaches have strengths and weaknesses. The dashboard manages complexity without using composite measures so that certainly it is deficient from the communication point of view. In the case of the measurement of well-being, the question without answer will be: "Is well-being increased or decreased?" The composite indicator manages complexity by reducing the dimensions in space with an evident loss of information; however, it allows a single measure that is more communicative.

To construct a well-being composite indicator, before the theoretical and methodological aspects, we need to answer the question: "Is it possible to measure well-being with a formula?" The answer is probably "Yes" if a paradigm of work is strictly respected. In literature, several attempts to measure well-being do not respect a paradigm of work and arrive at unreliable and questionable conclusions. This is the main reason for the failure of many alternative measures to GDP.

No universal method exists for constructing composite indicators. In each case their construction is very much determined by the particular application, including formal elements, and incorporates some expert knowledge on the phenomenon. Nevertheless, the advantages of composite indicators are clear, and they can be summarized as unidimensional measurement of the phenomenon, easy interpretation with respect to a set of many individual indicators and simplification of the data analysis.

A basic rule to keep in mind is 'garbage in garbage out'; that is, if the original matrix contains garbage then the composite indicator produces garbage. If a phenomenon is poorly defined, then it will certainly be poorly measured. Despite this, the reverse is not true. If the phenomenon is well defined and the matrix is composed of individual indicators of good quality, then it is not always true that the composite indicator is valid. It depends on the statistical methodology used, which must be 'well matched' with the theoretical framework on which the phenomenon to be measured is based.

2.2 The steps for constructing a composite indicator

The construction of a composite indicator is a complex task whose steps involve several alternatives and possibilities that affect the quality and reliability of the results. The main problems, in this approach, concern the choice of the theoretical framework, the availability of the data (in space and over time), the selection of the most representative indicators and their treatment in order to compare and aggregate them (Fanchette, 1974).

The paradigm of work is based on the following basic steps (Salzman, 2003; OECD, 2008; Mazziotta and Pareto, 2012): (1) defining the phenomenon to be measured (theoretical framework); (2) selecting a group of individual indicators; (3) normalizing the individual indicators; (4) aggregating the normalized indicators; and (5) validating the composite indicator.

Other steps may be required as appropriate (e.g., preliminary exploratory data analysis and imputation of missing data).

The definition of the phenomenon

The definition of the phenomenon should give a clear sense of what is being measured by the composite indicator. It should refer to a theoretical framework, linking various subgroups and underlying individual indicators. A fundamental issue, often overlooked in composite indicator construction, is the identification of the measurement model in order to specify the relationship between the phenomenon to be measured (latent variable) and its measures (individual indicators). In this respect, a model[1] of measurement can be conceived through two different conceptual approaches: reflective or formative (Jarvis et al., 2003; Diamantopoulos and Winklhofer, 2001).

The first form of specification is the reflective model, according to which individual indicators denote effects (or manifestations) of an underlying latent variable. Therefore, causality is from the concept to the indicators and a change in the phenomenon causes variation in all its measures. In this model, the concept exists independently of awareness or interpretation by the researcher, even if it is not directly measurable (Borsboom et al., 2003).

Specifically, the latent variable R represents the common cause shared by all MVs or indicators X_i reflecting the concept, with each indicator corresponding to a linear function of the underlying variable plus a measurement error:

$$X_i = \lambda_i R + \varepsilon_i \qquad (1)$$

where X_i is the indicator i, λ_i is a coefficient (loading) capturing the effect of R on X_i and ε_i is the measurement error for the indicator i. Measurement errors are assumed to be independent and unrelated to the latent variable.

A fundamental characteristic of reflective models is that individual indicators are interchangeable (the removal of one of the indicators does not change the essential nature of the underlying concept) and correlations between indicators are explained by the measurement model (all indicators must be intercorrelated).

Another important issue concerns the polarity of the individual

[1] Only linear models are considered here.

indicators. The 'polarity' of an individual indicator is the sign of the relation between the indicator and the concept to be measured. For example, in the case of well-being, "Life expectancy" has positive polarity, whereas "Unemployment rate" has negative polarity. In a reflective model, individual indicators with equal polarities must be positively correlated, whereas individual indicators with opposite polarities must be negatively correlated. Otherwise, the model will produce inconsistent results (Mazziotta and Pareto, 2019).

A typical example of a reflective model is the measurement of the intelligence of a person. In this case, it is the 'intelligence level' that influences the answers to a questionnaire for measuring attitude, and not vice versa. Hence, if the intelligence of a person increased, this would be accompanied by an increase in the number of correct answers to all questions (Simonetto, 2012).

The second approach is the formative model, according to which individual indicators are causes of an underlying latent variable, rather than its effects. Therefore, causality is from the indicators to the concept and a change in the phenomenon does not necessarily imply variations in all its measures. In this model, the concept is defined by, or is a function of, the observed variables.

The specification of the formative model is:

$$R = \sum_i \lambda_i X_i + \zeta \tag{2}$$

where λ_i is a coefficient capturing the effect of X_i on R, and ζ is an error term.

In this case, indicators are not interchangeable (omitting an indicator is omitting a part of the underlying concept) and correlations between indicators (r_{ij}, $i \neq j$) are not explained by the measurement model (high correlations between indicators are possible, but not generally expected). So, in a formative model, polarities and correlations are independent and individual indicators can have positive, negative or zero correlations.

It is worth noting that, because a formative model is not based on the hypothesis that the indicators are correlated, the correlation structure of the data cannot be used to determine the latent variable. So, the latent variable can be estimated by taking a weighted[2] average of the indicators that comprise the concept (Shwartz et al., 2015).

A typical example of a formative model is the measurement of the well-being of society. It depends on health, income, occupation, services, environment, etc., and not vice versa. So, if any one of these factors improved, well-being would increase (even if the other factors did not change). However, if well-being increased, this would not necessarily be accompanied by an improvement in all factors.

One of the oldest and most famous formative composite indicators is the HDI by the United Nations Development Programme. It is a composite measure of human development that includes three theoretical dimensions: Health, Education and Income. Any change in one or more of these components is likely to cause a change in a country's HDI score, but there is no reason to expect the components to be correlated. The same goes for the Canadian Index of Well-being (CIW), a composite measure of well-being based on eight domains: Living Standards, Healthy Populations, Community Vitality, Democratic Engagement, Leisure and Culture, Time Use, Education, and Environment (Michalos et al., 2011).

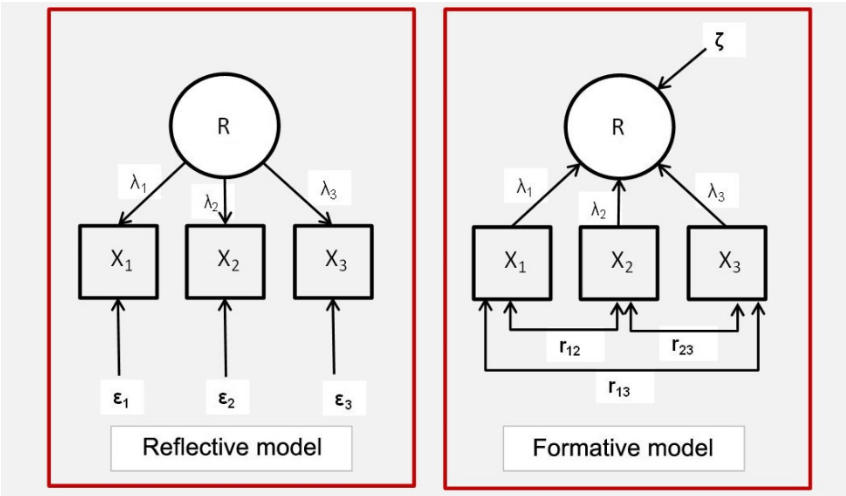
Note that Equation (1) is a system of simple regression equations where each individual indicator is the dependent variable and the latent variable is the explanatory variable, whereas (2) represents a multiple regression equation where the latent variable is the dependent variable, and the indicators are the explanatory variables. Hence, the correct interpretation of the relationships between indicators and the latent variable allows the procedure aimed at aggregating individual indicators to be correctly identified (Maggino, 2014).

In Figure 2.1, the two different approaches are graphically represented. Traditionally, the reflective model is applied in the development of scaling models for subjective measurement (e.g., attitude or satisfaction scale construction), whereas the formative model is commonly used in the construction of composite indicators based on both objective and subjective indicators (Maggino and Zumbo, 2012).

[2] Experts suggest that weights could be determined a priori, according to the theoretical contribution of the indicators to the concept (Howell et al., 2007). For Cadogan and Lee (2013), if there is no theory suggesting the contrary, individual indicators should have equal weightings.

Although the reflective view dominates the psychological and management sciences, the formative view is common in economics and sociology (Coltman et al., 2008).

Figure 2.1
Alternative measurement models



The selection of the indicators

In this step, the number and nature of the components that will make up part of the composite indicator need to be determined. Then, the specific indicators employed in estimating each of the component indicators must be selected. Individual indicators must be quantitative (discrete or continuous) and their polarity must be well defined.

The strengths and weaknesses of a composite indicator largely derive from the quality of the underlying indicators. The selection is generally based on theory, empirical analysis, pragmatism or intuitive appeal (Booysen, 2002). Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility, etc. (OECD, 2008). For example, if the composite indicator is to be calculated annually, individual indicators available on a multi-annual basis cannot be included. Similarly, if the composite indicator is to be calculated by province, individual indicators available at the regional level cannot be included. Furthermore, it should be noted that equally relevant individual indicators will have to be weighted in the same way, in the aggregation step, while indicators with different relevance will require the definition of an 'ad hoc' weighting system.

The selection step is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. A statistical approach to the choice of indicators involves calculating the correlation between potential indicators and including the ones that are less correlated in order to minimize redundancy (Salzman, 2003). However, the selection process depends on the measurement model used: in a *reflective* model, all the individual indicators must be intercorrelated, whereas in a *formative* model they can show negative or zero correlations (Diamantopoulos et al., 2008). Principal component analysis (PCA) can shed light on the correlations among individual indicators and on the consequences of including or excluding some of them.

The normalization

The normalization step is aimed at making the indicators comparable. Normalization is required before any data aggregation as the indicators in a data set are often expressed in different measurement units and can have different polarities. In particular, if a polarity is negative (the individual indicator is negatively correlated with the phenomenon to be measured), then it must be reversed, so that an increase in the normalized indicators corresponds to an increase in the composite indicator (Salzman, 2003).

Therefore, normalization has the following functions:

– bringing all the individual indicators to the same scale, transforming them into pure, dimensionless numbers;

– bringing all the individual indicators to positive polarity.

There are various normalization methods, some of which transform the variance or the range of the indicators to a common basis, and others that emphasize percentage change (Mazziotta and Pareto, 2017). The most commonly used are:

– Standardization (or *z*-scores). This is the method most commonly used by statisticians. It takes the difference between the original values and the mean, divided by the standard deviation, for each indicator. Indicators are thus converted to a common scale with a mean of 0 and standard deviation of 1. If an indicator has negative polarity, standardized values can be multiplied by -1.

– Rescaling (or Min-Max). This is the method most commonly used by sociologists. It takes the difference between the original values and the minimum, divided by the range, for each indicator. Indicators are thus converted to a common scale ranging between 0 and 1. If an indicator has negative polarity, the complement of rescaled values with respect to 1 can be calculated.

– Distance from a reference (or Indicization). This is the method most commonly used by economists. It takes the ratio between the original values and a reference value (base) for each indicator. Indicators are thus converted to a common scale where the reference is set equal to 1. If an indicator has negative polarity, it should first be transformed into its reciprocal (which will have positive polarity) and then indicized. However, indicization is recommended only for indicators with positive polarity, as the reciprocal is a non-linear transformation (Terzi and Moroni, 2004).

Each method has its advantages and disadvantages, all of which need to be evaluated. The researcher must identify the most suitable normalization method to apply to the problem at hand, taking into account its properties and robustness against possible outliers in the data.

The main pros and cons of different normalization methods are summarized in Table 2.1. Note that only standardization and rescaling normalize the 'variability' of indicators, whereas indicization saves the original coefficients of variation (only for indicators with positive polarity). On the other hand, standardization and indicization centre the variables around a common reference (the mean or the base), whereas rescaling does not centre them.

Different normalization methods will produce different results for the composite indicator. Therefore, a robustness analysis should be carried out to assess their impact on the results (Freudenberg, 2003).

The aggregation

Aggregation is the combination of all the components to form one or more composite indicators. It has the following functions:

– defining the importance of each individual indicator (weighting system);

– identifying the technique (compensatory, partially compensatory or non-compensatory) for summarizing the individual indicator values into a single number.

Table 2.1 – *Pros and cons of normalization methods*

Normalization method	Pros	Cons
Standardization (or *z*-scores)	Applicable to indicators with positive, negative and zero values. Normalized indicators are centred on the mean and have equal variances.	Not very suitable for bounded indicators[3]. Produces negative values. Sensitive to outliers.
Rescaling (or Min-Max)	Applicable to indicators with positive, negative and zero values. Normalized indicators have equal ranges.	Not very suitable for unbounded indicators. Normalized indicators are not centred on the mean. Sensitive to outliers (the range depends on extreme values).
Distance from a reference (or Indicization)	Suitable both for bounded and unbounded indicators. Normalized indicators are centred on the base (e.g., the mean) and have the same coefficient of variation as original indicators (only for indicators with positive polarity).	Not applicable to indicators with negative values (zero values are accepted only for indicators with positive polarity). Normalized indicators have different variability. Very sensitive to outliers.

[3] Indicators can be divided into 'bounded' and 'unbounded'. We say that an indicator is 'bounded' when it ranges between fixed values. An example of a bounded indicator is the 'Employment rate', which always ranges between 0 and 100. We say that an indicator is 'unbounded' when there are no predetermined upper or lower limits. An example of an unbounded indicator is 'Household disposable income', because there is theoretically no limit to how high the income could be.

The aim of a weighting system is that weights should reflect the relative importance (in terms of significance, reliability or other characteristics) of the individual indicators. The weights given to different indicators heavily influence the outcomes of the composite indicator. So, weights ideally should be selected according to an underlying theoretical framework for the composite indicator. The most widely used techniques for weighting individual indicators are the following: a) no weighting (equal weighting approach); b) subjective or expert weighting; and c) objective or 'data-driven' weighting.

a) If no weighting is defined, equal weights are applied to all individual indicators. This implies that all indicators in the composite indicator have equal importance, which may not be the case. However, if there are no statistical or empirical grounds for choosing different weights, this may be a valid approach in some contexts[4].

b) Subjective or expert weighting is typically set by a group of specialists who define weights for each indicator. The values determined by the specialists are then averaged. Weights are sometimes defined by policymakers or social surveys about how meaningful or important individual indicators are to people.

c) Objective or 'data-driven' weighting can be used to set weights based on the data themselves under a specific mathematical function. A typical example is the use of the coefficients of the first factor of PCA. This is an empirical option for weight selection and it has the advantage of determining the set of weights that explains the largest variation in the original indicators[5].

Different weighting systems imply different results and, given the subjectivity inherent in many of these criteria, no weighting system is above criticism. Each approach has its benefits and drawbacks, and there

[4] Note that the equal weighting approach may give extra weight to certain performance aspects if several individual indicators are in effect measuring the same attribute. As a remedy, indicators could be tested for statistical correlations, and lower weights could be given to variables strongly correlated with each other. On the other hand, correlations may merely show that unit performance on these indicators is similar (Freudenberg, 2003).

[5] Although PCA has a number of excellent mathematical properties, its use in weighting components of social indices is dubious. For example, it may lead to indicators that have little variation being assigned small weights, irrespective of their possible contextual importance (Salzman, 2003).

is no ultimate case of a clear winner or a kind of 'one-size-fits-all' solution (Greco et al., 2019). On the contrary, it is up to the indicator developer to choose the weighting system that is best fitted to the purpose of the construction, as disclosed in the theoretical framework (OECD, 2008).

Another fundamental issue concerning composite indicator construction is the degree of compensability or substitutability of the individual indicators. The components of a composite indicator are called 'substitutable' if a deficit in one component can be compensated by a surplus in another (e.g., a low value of "Proportion of people who have participated in religious or spiritual activities" can be offset by a high value of "Proportion of people who have participated in meetings of cultural or recreational associations" and vice versa). Similarly, the components of a composite indicator are called 'non-substitutable' if compensation among them is not allowed (e.g., a low value of "Life expectancy at birth" cannot be offset by a high value of "Gross national income per capita" and vice versa)[6]. Thus, we can define an aggregation approach as 'compensatory' or 'non-compensatory' depending on whether it permits compensability or not (Casadio Tarabusi and Guarini, 2013). An in-between approach based on an 'imperfect substitutability' across all components of a composite indicator is called 'partially compensatory'. Compensability is closely related to the concept of unbalance, i.e., a disequilibrium among the indicators that are used to build the composite indicator. In any composite indicator each dimension is introduced to represent a relevant aspect of the phenomenon considered, therefore a measure of unbalance among dimensions may help the overall understanding of the phenomenon. In a non-compensatory or partially compensatory approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is often used. A compensatory approach involves the use of linear functions, whereas a partially compensatory or non-compensatory

[6] Note that compensability/non-compensability does not imply dependence/independence and vice versa. For example, "Hospital beds (per 1,000 people)" and "Hospital doctors (per 1,000 people)" are two dependents (positively correlated) indicators but they are non-substitutable, because a deficit in beds cannot be offset by a surplus in doctors and vice versa (Mazziotta and Pareto, 2016).

approach requires the use of non-linear functions (Casadio Tarabusi and Guarini, 2013).

As we know, the most common aggregation function for constructing a composite indicator is the sum of weighted and normalized individual indicators (OECD, 2008). An additive aggregation function allows assessment of the marginal contribution of each individual indicator separately. These marginal contributions can then be added together to yield a total value. However, an undesirable feature of additive aggregations is the implied full compensability, such that poor performance in some indicators can be compensated for by sufficiently high values in other indicators (perfect substitutability). A widely used alternative is geometric aggregation (Zhou et al., 2010), where a multiplicative function is used. This aggregation function allows a partial compensability, so that an increase in the most deprived indicator will have a higher impact on the composite indicator (imperfect substitutability). Such a choice is advisable whenever a reasonable achievement in any of the individual indicators is considered to be crucial for overall performance (Chiappero-Martinetti and von Jacobi, 2012).

Additive and multiplicative aggregation functions can be seen as special cases of a generalized mean or power mean of order *r*. Given the normalized data matrix $Y_{n,m}=\{y_{ij}\}$, with *n* rows (statistical units) and *m* columns (normalized indicators), the power mean of order *r*, for unit *i*, is defined as follows:

$$M_i^r = \left(\sum_{j=1}^{m} y_{ij}^r w_j\right)^{\frac{1}{r}} \tag{3}$$

where w_j is the weight of indicator *j* ($0 < w_j < 1$) and $\sum_{j=1}^{m} w_j = 1$.

For $r = 1$ we have the arithmetic mean (compensatory approach), for $r \to 0$ the geometric mean (partially compensatory approach), for $r \to -\infty$ the minimum and for $r \to +\infty$ the maximum (non-compensatory approaches). In addition, we have:

$$M_i^{-\infty} \leq \ldots \leq M_i^{-1} \leq M_i^{0} \leq M_i^{1} \leq M_i^{2} \leq M_i^{3} \leq \ldots \leq M_i^{+\infty}$$

and the means are equal if and only if $y_{ij}=y_{ik}$ ($j \neq k$). So, for each value of *r*, we have a different approach and distinct features (intensity and direction) of the penalization for unbalanced values. If increasing values of the indicator correspond to an improvement of the phenomenon (e.g., socio-economic development), a downward penalization must be used ($r < 1$). On the other hand, if increasing values of the indicator correspond to a worsening of the phenomenon (e.g., poverty), an upward penalization must be used ($r > 1$). Anyhow, an unbalance among individual indicators values will have a negative effect on the value of the composite indicator.

The literature offers a wide variety of alternative aggregation methods, each with its pros and cons: for example, the Wroclaw Taxonomic Method (Harbison et al., 1970), the Mean-Min Function (Casadio Tarabusi and Guarini, 2013) and the Mazziotta-Pareto Index (Mazziotta and Pareto, 2016). Also, multivariate statistical methods such as PCA or factor analysis (FA) and data envelopment analysis (DEA) are often used.

Aggregation is the most important and most delicate step of the procedure. In this stage, the choices of the researcher assume a fundamental role, from a methodological point of view, as even minimal changes in the method applied can have a major impact on the result. Therefore, data aggregation has always been an interesting but controversial topic in composite indicator construction (Saltelli, 2007).

The validation

The validation step is aimed at assessing the robustness of the composite indicator, in terms of capacity to produce correct and stable measures, and its discriminant capacity (i.e., the ability to enhance the differences between individual scores or rankings). As seen above, the outcomes and rankings of individual units on the composite indicator may largely depend on the decisions taken at each of the preceding steps (selection of individual indicators, normalization and aggregation). For this reason, statistical analyses should be conducted to explore the robustness of rankings to the inclusion and exclusion of individual indicators and setting different decision rules to construct the composite indicator (Freudenberg, 2003).

The robustness of a composite indicator is assessed by two different methodologies: uncertainty analysis (UA) and sensitivity analysis (SA). UA focuses on how uncertainty in the input factors propagates through the structure of the composite indicator and affects the results. SA studies how much each individual source of uncertainty contributes to the output variance (Saisana et al., 2005). UA and SA can be used synergistically and iteratively during composite indicator construction to help with indicator selection, add transparency to the indicator construction process, and explore the robustness of alternative composite indicator designs and rankings (USAID, 2014).

The discriminant capacity of a composite indicator is assessed by exploring its capacity in: a) discriminating between units and/or groups; b) distributing all the units without any concentration of individual scores in a few segments of the continuum; c) showing values that are interpretable in terms of selectivity through the identification of particular reference values or *cut-points* (Maggino and Zumbo, 2012)[7].

2.3 Best practices

As we have seen above, there does not exist a composite indicator that is universally valid for all areas of application, since its validity depends on the strategic objectives of the research. The main factors to take into account in the choice of the method to be adopted for summarizing a set of individual indicators are as follows (Mazziotta and Pareto, 2013):
– type of model (reflective/formative);
– type of indicators (substitutable/partially substitutable/non-substitutable);
– type of aggregation (simple/complex);
– type of comparisons (absolute/relative);
– type of weights (objective/subjective).

There is not always a 'well-established' solution, and sometimes it may be necessary to waive some requirements to satisfy others.

Type of model

The choice of measurement model is closely related to the selection of the individual indicators. If individual indicators are seen as the

[7] Point (a) can be verified by applying the traditional approaches of statistical hypothesis testing, whereas specific coefficients were proposed for evaluating (b) (Guilford, 1954). Receiver operating characteristic (ROC) analysis allows discriminant *cut-points* to be identified in (c).

'effect' of the latent variable, a reflective model must be adopted. If individual indicators are seen as the 'cause' of the latent variable, a formative model will have to be adopted. Note that the measurement model may affect the choice of the aggregation method. For example, FA is the most suitable method for the aggregation of reflective indicators (interchangeable and highly correlated with each other), but it is not suitable for the aggregation of formative indicators (not interchangeable and/or not related to each other).

Type of indicators

This is one of the main factors that affect the choice of aggregation method. If the individual indicators are fully substitutable, then a compensatory approach with a linear aggregation function is indicated (e.g., the arithmetic mean). If the individual indicators are partially substitutable (or non-substitutable), then a partially compensatory approach (or non-compensatory approach) with a non-linear aggregation function is required (e.g., the geometric mean or the minimum).

Type of aggregation

The choice of aggregation method also depends on the aim of the work and on the type of 'users' (researchers or common people). Generally, an aggregation method can be considered 'simple' or 'complex'. We say that an aggregation method is 'simple' when an easily understandable mathematical function is used (e.g., the HDI). On the other hand, an aggregation method is said to be 'complex' if a sophisticated model or multivariate method is used (e.g., PCA).

Type of comparisons

Data normalization firstly depends on the type of comparisons required. All the normalization methods allow for space comparisons (comparisons between values of different units, at the same time), whereas time comparisons (comparisons between values of the same unit or of different units, at different times) may be difficult to make or to interpret. Comparisons over time may be 'absolute' or 'relative'. We say that a time comparison is 'relative' when the composite indicator values, at time *t*, depend on one or more endogenous parameters that change with time (e.g., mean and variance of the individual indicators at time t). Similarly, we say that a time comparison is 'absolute' when the composite indicator values, at time t, depend on one or more exogenous parameters that do not change over time (e.g., minimum and maximum of the individual indicators fixed by the researcher). Standardization allows only for relative comparisons since it is based exclusively on values of the individual indicators at time t. Other methods, such as rescaling and indicization, require that the minimum and maximum (e.g., the 'goalposts' of the HDI) or the base of normalized indicators are independent from time t, in order to perform comparisons in absolute terms (Tarantola, 2008).

Type of weights

The question of the choice of a weighting system in order to weight the individual indicators, according to their different importance in expressing the considered phenomenon, necessarily involves the introduction of an arbitrary component.

In particular, it should be noted that implicit weights can be introduced during normalization, while explicit weights – such as equal weighting – may be used during aggregation (e.g., the weights w_j of formula 3). In fact, equal weighting of the normalized indicators is not necessarily equivalent to equal weighting of the original indicators. For example, equal weighting of standardized (or rescaled) indicators means assigning weights inversely proportional to the standard deviation (or range) of the original indicators. Therefore, the variability of each original indicator acts as an implicit weight during the aggregation.

Finally, it is good to keep in mind that the first criterion to be followed in the construction of a composite indicator (as in any statistical model) is the *principle of parsimony* (Mazziotta and Pareto, 2020). This principle states that the composite indicator must be as simple as possible to allow an easy interpretation of results, both in space and time. In order to construct a composite indicator that is as simple as possible, the processing to be performed on the data must be reduced to the minimum necessary. Therefore, only one normalization method must be applied to the data matrix and no further transformation of the obtained scores should be carried out, as they are already normalized.

In conclusion, the construction of a composite indicator must follow a precise work paradigm and international literature is unanimous in this sense. Methodological shortcuts or even fanciful approaches, such as normalizing data several times, are absolutely to be avoided since a composite indicator has a great responsibility: measuring multidimensional phenomena to better understand the reality.

References

Annoni, P., & Kozovska, K. (2010). *EU regional competitiveness index 2010* (JRC Scientific and Technical Reports). Luxembourg: Publications Office of the European Union.

Bandura, R. (2008). *A survey of composite indices measuring country performance: 2008 Update*. A UNDP/ODS Working Paper, New York.

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

Borsboom, D., Mellenbergh, G.J., & Heerden, J.V. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203-219.

Cadogan, J.W., & Lee, N. (2013). Improper use of endogenous formative variables. *Journal of Business Research*, *66*, 233-241.

Casadio Tarabusi, E., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research*, *112*, 19-45.

Chiappero-Martinetti, E., & von Jacobi, N. (2012). Light and shade of multidimensional indexes. How methodological choices impact on empirical results. In F. Maggino, & G. Nuvolati (Eds.), *Quality of life in Italy, research and reflections* (pp. 69-103). Social Indicators Research Series, 48. Cham: Springer.

Coltman, T., Devinney, T.M., Midgley, D.F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*, 1250-1262.

Diamantopoulos, A., & Winklhofer, H.M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*, 269-277.

Diamantopoulos, A., Riefler, P., & Roth, K.P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*, 1203-1218.

Fanchette, S. (1974). Social indicators: Problems of methodology and selection. In UNESCO, *Social indicators: Problems of definition and of selection*. Reports and papers in the social sciences, 30, Paris.

Freudenberg, M. (2003). *Composite indicators of country performance: Acritical assessment*. OECD, DSTI/IND, 5, Paris.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, *141*, 61-94.

Guilford, J.P. (1954). *Psychometric methods*. New York-London: McGraw-Hill.

Harbison, F.H., Maruhnic, J., & Resnick, J.R. (1970). *Quantitative analyses of modernization and development*. New Jersey: Princeton University Press.

Howell, R.D., Breivik, E., & Wilcox, J.B. (2007). Reconsidering formative measurement. *Psychological Methods*, *12*, 205-218.

Jarvis, C.B., Mackenzie, S.B., & Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*, 199-218.

Maggino, F, & Zumbo, B.D. (2012). Measuring the quality of life and the construction of social indicators. In K.C. Land et al. (Eds.), *Handbook of social indicators and quality of life research* (pp. 201-238). New York: Springer.

Maggino, F. (2014). Indicator development and construction. In: A.C. Michalos, (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 3190-3197). Dordrecht: Springer.

Mazziotta, M., & Pareto, A. (2012). A non-compensatory approach for the measurement of the quality of life. In F. Maggino, & G. Nuvolati (Eds.), *Quality of life in Italy: Research and reflections* (pp. 27-40). New York: Springer.

Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one? *Rivista Italiana di Economia Demografia e Statistica*, *LXVII* (2), 67-80.

Mazziotta, M., & Pareto, A. (2016). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social*

Indicators Research, *127*, 983-1003.
Mazziotta, M., & Pareto, A. (2017). Synthesis of indicators: The composite indicators approach. In F. Maggino (Ed.), *Complexity in society: From indicators construction to their synthesis* (pp. 159-191). New York: Springer.
Mazziotta, M., & Pareto, A. (2019). Use and misuse of PCA for measuring well-being, *Social Indicators Research*, *142*, 451-476.
Mazziotta, M., & Pareto, A. (2020). *Gli indici sintetici*. Giappichelli, Torino.
Michalos, A.C., Smale, B., Labonté, R., Muharjarine, N., Scott, K., Moore, K., et al. (2011). *The Canadian Index of Well-being*. Technical report 1.0. Waterloo, ON: Canadian Index of Wellbeing and University of Waterloo.
OECD (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD Publications.
Saisana, M., & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. European Commission-JRC, EUR 20408 EN, Ispra.
Saisana, M., Tarantola, S., & Saltelli, A. (2005). Uncertainty and sensitivity techniques as tools for the analysis and validation of composite indicators. *Journal of the Royal Statistical Society A*, *168*, 307-323.
Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, *81*, 65-77.
Salzman, J. (2003). *Methodological choices encountered in the construction of composite indices of economic and social well-being*. Technical Report, Center for the Study of Living Standards, Ottawa.
Shwartz, M., Restuccia, J.D., & Rosen, A.K. (2015). Composite measures of health care provider performance: A description of approaches. *The Milbank Quarterly*, *93*, 788-825.
Simonetto, A. (2012). Formative and reflective models: State of the art. *Electronic Journal of Applied Statistical Analysis*, *5*, 452-457.
Tarantola, S. (2008). *European Innovation Scoreboard: Strategies to measure country progress over time*. JRC Scientific and Technical Reports, EUR 23526 EN, Luxembourg.
Terzi, S., & Moroni, L. (2004). Graduatorie della qualità della vita e loro sensibilità al pretrattamento delle variabili che la definiscono: alcune

critiche al dossier de Il Sole 24 Ore, *Quaderni di Statistica*, *6*, 105-127.

UNDP (1990). *Human Development Report 1990: Concept and Measurement of Human Development*. Oxford: Oxford University Press.

UNDP (2001). *Human Development Report 2001*. New York: Oxford University Press.

UNDP (2010). *Human Development Report 2010*. New York: Palgrave Macmillan.

UNECE (2019). *Guidelines on Producing Leading, Composite and Sentiment Indicators*. United Nations, Geneva.

USAID (2014). *Design and Use of Composite Indices in Assessments of Climate Change Vulnerability and Resilience*. U.S. Agency for International Development, Washington.

Zhou, P., Ang, B.W., & Zhou, D.Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research*, *96*, 169-181.

Chapter 3

The use of latent variables in composite indicators*

3.1 Introduction

We usually call a *composite indicator* (CI) a not directly observable multidimensional concept/phenomenon (*latent variable*) that can be computed (constructed) by means of a mathematical combination of manifest variables (or rather *elementary indicators*) (Lauro et al., 2018; Trinchera et al., 2008) as illustrated in Chapter 2. Composite indicators should ideally measure multidimensional concepts that cannot be captured by a single indicator, e.g., competitiveness, industrialization, sustainability, quality of life, well-being, etc. (Nardo and Saisana, 2008; Saisana and Tarantola, 2002). CIs are based on elementary indicators or on *subindices* (*pillars*), i.e., aggregations of elementary indicators. These elementary indicators (EIs) or subindices are aggregated by analytical methods to give an overall score for each individual unit, usually country or geographical area. The results are used to either create a ranking or to simply summarize the data (Bandura, 2008).

The statistical approaches we will recall in this chapter are all designed for the study of latent variables, so the language we will use might be slightly different from the usual CI glossary. In fact, we will refer to a CI as a latent variable or a latent construct that – depending on the measurement model – affects (in the formative model) or is affected by (in the reflective model) manifest variables (MVs) or by MVs, also called elementary indicators (EI).

As already outlined, in order to derive a CI, the first step is the definition and identification of the concept to analyse and consequently of the manifest variables (including checking the consistency of the MVs and the latent concept to be measured); the second step is concerned with the normalization (which can include tail trimming, or even

* This chapter is mostly written by Silvia Terzi.

smoothing of some MVs); the third step consists in the definition of suitable weights and in the choice of the aggregation function.

Let us focus on the first step, the theoretical framework, an important but often disregarded issue. In the identification stage, the nature of the measurement model must be clearly defined, i.e., the relationships between the measures (either MVs or EIs) and the underlying latent variables.

First of all, it is necessary to specify the direction of causality: from the MVs (or EIs) to the latent concept or vice versa. The first case defines a formative measurement model. The model assumes that the MVs are defining essential characteristics of the latent concept, and consequently changes in the manifest variables will cause changes in the CI. Moreover, omitting a manifest variable may alter the conceptual domain of the latent construct. So, the choice of manifest variables is an aspect of considerable relevance. On the other hand, in the reflective framework, each MV (or each EI) reflects CI variations: changes in the CI cause changes in all the MVs. In this case the MVs are assumed manifestations of the same construct, thus they should be highly correlated and interchangeable. Again, the choice of manifest variables is far from trivial: since they are assumed to reflect the same underlying construct, as well as being exchangeable, they are supposed to have the same antecedents and consequences. Before proceeding to the construction of a reflective measurement model, the internal consistency of the latent concept is usually verified by means of the Cronbach's alpha index. This index ranges between 0 and 1: a value of around 0.7 validates the assumption that the CI is a consistent representation of the MVs.

Once the theoretical framework is clear and the manifest variables have been normalized (so they are now elementary indicators) the focus is on the choice of the aggregating function and on the weights.

The most widely used are additive methods: weighted sum of the elementary indicators. The simplest choice is equal weights to normalized MVs. But other choices are expert weighting or weights derived from an objective function.

Examples of this last criterion are multivariate techniques such as principal component analysis (PCA) and (exploratory) factor analysis (FA), but we could also quote the Benefit of the Doubt (Cherchye et al., 2007) and Unobserved Components methods (Kaufmann et al., 1999).

In particular, PCA and exploratory FA deserve to be mentioned since they are widely used to build latent variables and/or as data reduction methods. Both these methods allow a CI to be built– either as a principal component or as a unique factor – deriving weights from an objective function. The desired weights must preserve the maximum possible proportion of the total variation in the original data set. However, in order to retain an adequate proportion of variance, these techniques might suggest the computation of more than one CI.

For this reason, it can be said that these are methods aimed at shedding light on the nature of the multidimensional phenomenon through a synthesis of its most relevant aspects, i.e., of its latent dimensions, or put more succinctly that they are data reduction techniques.

3.2. Principal components and factor analysis

A principal component (PC) is an unobserved variable Y, linear combination of p variables X_j $(j=1,\dots,p)$ with weights a_j:

$$Y = a_1X_1 + \dots + a_pX_p$$

For the linear combination to make sense, the variables must all have the same unit of measurement or be pure numbers; if not, the variables will be normalized.

Let a be the column vector of weights: $a' = (a_1, a_2, \dots, a_p)'$; let X be the matrix nxp whose columns are the observations on the possibly normalized variables: $X = [x_1, x_2, \dots, x_p]$ where x_j $(j=1,\dots,p)$ is the n-dimensional vector of observations on the j-th variable; let Σ be the variance-covariance matrix of the variables X_j. The column vector $y=\Sigma_j a_j x_j$ can also be written as $y = Xa$, and it can be shown that $\sigma^2_y = a'\Sigma a$ (where σ^2_y indicates the variance of Y). In order to have a finite and uniquely determined solution for the weights a_j, the constraint $a'a = 1$ is set.

The problem of finding the maximum of $a'\Sigma a$ subject to the constraint $a'a = 1$ is equivalent to maximizing the Lagrange function: $a'\Sigma a - \lambda\ (a'a - 1)$, where λ is a Lagrange multiplier.

Differentiating with respect to the vector a, and equating to the null vector, produces the equation $\Sigma a - \lambda a = 0$, whose solutions are the eigenvalues λ of the matrix Σ, and the corresponding unit-norm eigenvectors a. Moreover, pre-multiplying both sides of the equivalent equality $\Sigma a = \lambda a$ *by* a' we obtain:

$$a'\Sigma a=\lambda a'a=\lambda$$

This implies that the eigenvalues of the matrix Σ are the variances of the linear combinations defined by the corresponding eigenvectors a.

Depending on the rank of the symmetric matrix Σ we have q distinct solutions ($q \leq p$), i.e., the eigenvalues λ_h, the associated eigenvectors ah and the principal components $y_h = Xa_h$, $h=1,2,...,q$. Moreover, due to the pairwise orthogonality of the eigenvectors of a symmetrical matrix, the principal components are pairwise uncorrelated.

Another useful property of any square matrix is that the sum of its eigenvalues is equal to its trace (that is, the sum of its diagonal elements), so for what concerns the covariance matrix Σ is concerned, the sum of its eigenvalues ($\Sigma_h \lambda_h$), i.e., the sum of the variances of the principal components Y_h, equals the sum of the variances of the variables X_j, $\Sigma_j var(X_j)$.

Within the composite indicator construction framework, the problem now becomes: how to choose one set of weights among the q eigenvectors?

Since $\lambda_h = var(Y_h)\ \forall h=1,...,q$ and assuming the eigenvalues are arranged in decreasing order, we are interested in the largest eigenvalue, λ_1 (and corresponding eigenvector a_1), so that the first PC $y_1 = Xa_1$ has the maximum possible variance ($\sigma^2_1 = \lambda_1$). Whenever σ^2_1 accounts for at least 75% of the sum of the variances of the variables X_j, the first PC is considered a satisfactory synthesis of the original variables, and thus an adequate CI.

Assume the PCs Y_h are arranged in decreasing order of variance. Usually, in order to decide what the minimum number of latent variables Y_h is that provide a good synthesis of the multidimensional phenomenon under study, a threshold value is set for the proportion of variability pertaining to the first s principal components, that is, the ratio $(\lambda_1+\ldots+\lambda_s)/(\lambda_1+\ldots+\lambda_q)$. It is usually 0.80 or 0.75. This reference value represents the desired/required summary power/capacity of the most explanatory latent variables. The smallest number of PCs whose summary power reaches the set threshold is chosen.

One undesirable feature of principal component analysis (PCA) is that if we change the normalisation normalization transformation (or introduce a normalization criterion), the covariance structure changes – the matrix Σ – as do the eigenvalues and eigenvectors of the new covariance matrix. Consequently, if we use PCA to build a composite indicator, its interpretation, its loadings and its summary power depend on the normalization chosen for the variables X_j.

PCA could in principle refer either to a formative or to a reflective model; however, when building a well-being composite indicator, the variables (or the elementary indicators) might be highly correlated – as in reflective measurement models – but are scarcely interchangeable; consequently, within our general framework a formative model seems more suitable.

Factor analysis (FA) is designed only for reflective models. It is aimed at explaining the existing interdependence among a set of p variables $X_1, X_2,..., X_p$ in terms of a smaller number (q, in general; hopefully $q=1$ when resorting to FA to define one composite indicator) of common non-observable underlying factors $f_1, f_2, \ldots, f_q$.

In a way, factor analysis represents a step ahead with respect to PCA since it consists, rather than in the weighted aggregation of the observed variables, in the estimation of a model that reproduces their covariance structure.

However, in contrast to PCA, the FA model assumes that the data are based on the underlying factors of the model, and that the data variance can be decomposed into that accounted for by common factors and that due to unique factors.

There are two types of factor analysis, exploratory and confirmatory. Exploratory factor analysis is a method to explore the underlying structure of a set of observed variables; confirmatory factor analysis is a method to verify a factor structure that has already been defined. Obviously within CI building our interest is turned to exploratory methods.

Exploratory FA is based on the common factor model. In this model, manifest variables are expressed as a function of common factors and unique factors. Each unique factor influences only one manifest variable and does not explain correlations between manifest variables. Common factors influence more than one manifest variable and "factor loadings" are measures of the influence of a common factor on a manifest variable. For the exploratory FA procedure, we are mainly interested in identifying the common factors and the related manifest variables.

Let X be a p-dimensional random vector with an expected value μ and covariance matrix Σ. Let Λ be a pxq matrix of factor loadings λ_{jh}, F a q-dimensional vector of unobservable common factors $(f_1,f_2,\dots,f_q)'$,and U a p-dimensional vector of unobservable unique factors $(u_1,u_2,\dots,u_p)'$.

We can write the factorial model as:

$$X = \mu + \Lambda F + U$$

The usual assumptions are that common factors have mean zero, unit variance and are uncorrelated, and that unique factors have mean zero, variances Ψ_j, $j=1,\dots,p$ and that they are pairwise uncorrelated and uncorrelated with common factors f_h:

$$\begin{gathered} E(F) = 0 \\ Cov(F) = E(FF') = I_q \\ E(U) = 0 \\ Cov(U) = E(UU') = \Psi = diag(\psi_1, \psi_2, \dots, \psi_p) \\ E(FU') = O \end{gathered}$$

For the j-th observable variable X_j the model is:

$$X_j - \mu_j = \Sigma_h \lambda_{jh} f_h + U_j,\ j=1,2,\dots,p$$

and

$$var(X_j) = \Sigma_h \lambda^2_{jh} + \psi_j\ ;\ cov(X_j, X_t) = \Sigma_h \lambda_{jh} \lambda_{th}$$

$\Sigma_h \lambda^2_{jh}$ is called communality of the j-th variable and denoted by c_j. It represents the proportion of the variance of X_j due to the common factors. The interrelation (covariance) among MVs is entirely due to common factors.

In matrix notation:

$$X - \mu = \Lambda F + U$$

and

$$var\text{-}cov(X) = \Sigma = \Lambda \Lambda' + \Psi$$
$$cov(X,F) = \Lambda$$

Within the FA framework the focus is on the estimation of the factor loadings. Unfortunately, the matrix Λ is not uniquely defined, or rather it is defined only up to an orthogonal transformation. This means that once we have obtained a set of q common factors, we can orthogonally rotate them to obtain a new factor loadings matrix Λ^*, which gives rise to the same decomposition: $\Sigma = \Lambda^* \Lambda^{*\prime} + \Psi$. However, for a given number q of common factors, this characteristic can be seen as an opportunity: it allows the possibility of performing a rotation of the factors so as to enhance the interpretability of the results.

Various rotational strategies have been proposed. The goal of all of these strategies is to obtain a clear pattern of the loadings. However, different rotations imply different loadings, and thus different meanings/interpretations of common factors – a problem that can be considered a drawback to the method.

Before tackling the problem of the interpretation of the common factors, the analyst is faced with the problem of estimating the factor loadings and deciding how may common factors to retain.

The most common procedure for FA is to resort to PCA to extract the first principal components from the (sample) variance-covariance matrix Σ, neglecting the others.

Once the number of common factors has been identified, a rotation can help to make the output more understandable, by seeking a so-called "simple structure". For example, within a varimax rotation each factor will tend to have either large or small loadings on any particular variable. A varimax solution yields results that make it as easy as possible to identify each variable with a single factor. This is the most common rotation option.

Within the CI framework, if the first factor accounts for at least 65% of the total variance, the latent concept is considered unidimensional and the first factor is assumed to be the CI (i.e., its loadings are the desired weights).

However, FA could also yield more than one CI, and in fact, like PCA it is designed to compute all q common factors and then suggest how many to retain, whether one or more. Both these methods fall in the so-called "data-driven approach", an exploratory approach according to which (Benzecri, 1980) the models have to follow the data and not vice versa. What could happen is that when looking for one CI, defined as the latent variable the CI should measure/represent, PCA or FA could suggest that the multidimensional concept should be broken down into more than one latent construct (more than one dimension).

It would be desirable to clarify whether (or rather when, or under what circumstances) data should drive the theory or whether the theory should define the whole framework.

We believe that constructing a composite indicator is not a mere question of reducing data dimensionality or of extracting the maximum amount of information. In PCA, as in FA, the loadings (weights) have a clear mathematical meaning, but often lack socio-economic interpretta-tion. They are in fact data reduction techniques. They might even provide CIs that are inconsistent with preliminary assumptions: the PCA or the exploratory FA might suggest that one or more manifest variable is irrelevant, or that the multidimensional concept should be broken down into pillars (latent dimensions).

It is not totally clear what a CI is, but we assume it is something that can be defined – as far as its meaning and the variables it is based upon are concerned – but not directly observed: it is a latent variable, but what most characterizes a composite indicator is its multidimensionality. Consequently, CI building requires a theoretical framework that includes the definition of the multidimensional concept to be measured, the variables it is connected to, the definition of the measurement model and the selection of the most representative indicators.

A FA model is by its very nature reflective: causality goes from the LVs to the MVs, and the MVs are interchangeable. Within FA the latent nature of the CI prevails on its multidimensionality.

Within this framework, PCA could be used as a preliminary exploratory technique, to detect redundant manifest variables or to discover that in fact the multidimensional latent concept (the desired CI) should be broken down into more than one (uncorrelated) dimension, i.e., into more than one composite indicator (or latent variable).

In fact, when the latent multidimensional phenomenon can be broken down into latent dimensions, both PCA and FA will suggest two or more uncorrelated principal components/common factors. And this is precisely the case in which multidimensional techniques appear to be more suitable than aggregating functions based on predetermined weights. However, once the preliminary exploration is accomplished, the most convenient approach is not based on PCA, since it is preferable not to restrict the latent dimensions to being uncorrelated.

It seems, in fact, more appropriate to define a more flexible theoretical model linking two or more latent concepts and to compute a complex indicator, i.e., a synthesis of composite indicators. A useful methodological framework is provided by structural equation modelling.

3.3 Partial least squares path models

Structural equation models (SEMs) can be used to build systems of multidimensional latent variables (each measured by means of a set of manifest variables) connected by causal relations.

Each SEM model involves two levels of relations: the first one is the measurement (or outer) model that takes into account the relations between the MVs and the corresponding LVs; the second level considers the causal relations among the LVs (structural or inner model). Thus, the LVs can be seen not only as composite indicators, due to their relations with the corresponding MVs, but also as latent dimensions that give rise to complex indicators, as a consequence of the causal relations among each other.

The estimation of an SEM provides two sets of weights: one set measuring the impact of each MV on the corresponding LV, the other measuring relations among the LVs in the system.

Several methods have been developed to estimate SEM parameters, among them the partial least squares (PLS) path modelling approach (Wold, 1982).

A partial least squares path model (PLS-PM) is designed to estimate a network of causal relations, defined according to a theoretical model linking two or more latent concepts, each measured through a number of observable variables, and is a so-called "component-based estimation method", because of the key role that is played by the estimation of the

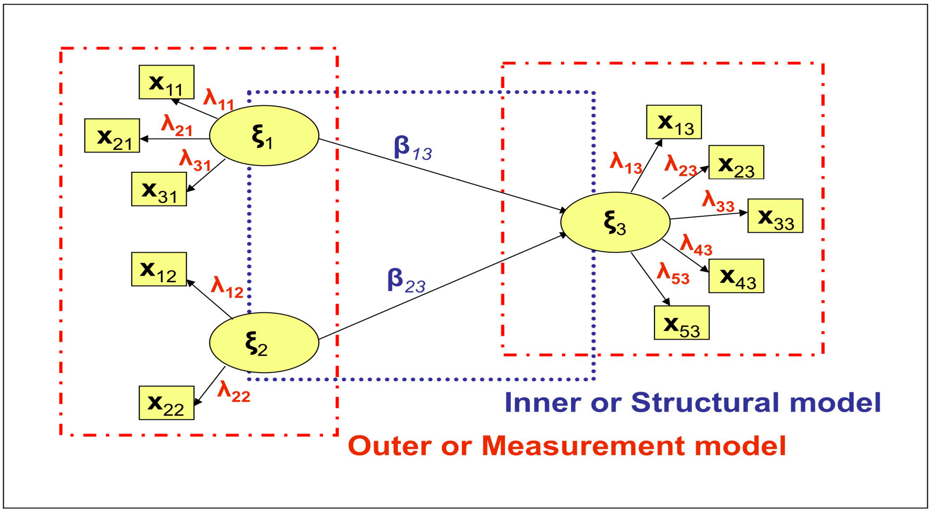
LVs in the model. It is a soft modelling exploratory technique that does not require strong assumptions on the data-generating process and has minimal demands on measurement scales, sample size and data distributions. PLS is based on simple and multiple ordinary least squares (OLS) regressions and it is particularly applicable for predictive applications and theory building (Chin, 1998; Davino et al., 2018).

Within the network of causal relations each latent variable is indirectly measured by means of a set (or block) of observable, manifest variables. A LV can be related to its MVs by either a reflective or a formative relation or by both. The set of these relationships defines the measurement or outer model.

Each LV is related to other LVs, in a systemic vision, by linear regression equations specifying the so-called "structural model" (or inner model).

PLS is an iterative algorithm that solves – separately – the blocks of the measurement model and then, in a second step, estimates the path coefficients in the structural model.

Figure 3.1
Structural and measurement models (Trinchera and Russolillo, 2010)



The estimation of the parameters of the model is performed using a procedure that computes LV scores using a PLS algorithm, and then performs ordinary least squares (OLS) regressions on them to estimate the structural equations.

PLS is prediction oriented: it is aimed at explaining at best the residual variance of the LVs and, potentially, of the MVs in any regression run in the model.

The PLS-PM algorithm provides LVs that are as much as possible correlated to each other and that at the same time explain the greatest possible amount of variance of its set of manifest variables. It assigns weights to the original variables taking into account the network of relationships between the LVs and the variance and covariance structure within and between the blocks of MVs.

It is a very flexible approach that allows the definition of both formative and reflective measurement models. In fact, each LV is approximated by a linear combination of related observed MVs.

It allows three different specifications of the measurement model, three ways to relate the MVs to their LVs:

- Reflective specification (outwards directed): each MV reflects the corresponding LV. In this case, the MVs should be highly correlated, due to the fact that they are correlated with the LV of which they are an expression. In other words, the block has to be homogeneous.
- Formative specification (inwards directed): the LV is supposed to be generated by its own MVs. Each MV or every set of MVs represents a different level of the underlying latent concept.
- The MIMIC specification: a combination of reflective and formative specifications within the same block of manifest variables.

One of the key features of PLS-PM is that it supplies LV scores. The LV scores – which represent the value of the LV for each statistical unit – are calculated as normalized weighted aggregates of the MVs. Moreover, PLS-PM provides information on the relative importance of LVs in explaining other LVs in the structural model by means of the path coefficients. The path coefficients are estimated through an OLS multiple regression among the latent variable scores (according to the path diagram structure).

It is not uncommon that the MVs used in the construction of a composite indicator express or represent different facets or nuances of a complex phenomenon; in this case they can be conceptually split into blocks of indicators. Each block can be summarized by a single composite indicator, which is considered causative with respect to a second-order composite indicator.

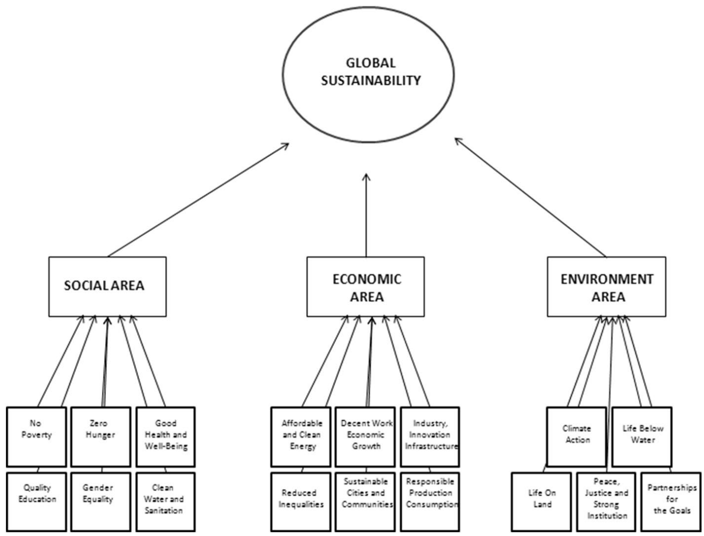
A key characteristic of the PLS-PM method is the extraction of CI scores: PLS-PM provides information on the relative importance of LVs in explaining other LVs in the structural model. Moreover, for each CI of first or of higher order, PLS-PM can provide a ranking of the units (Cataldo et al., 2020).

Whenever the latent multidimensional phenomenon (construct) is a synthesis/aggregation of latent variables (CI), or rather, whenever the MVs used in the construction of a composite indicator represent or refer to different aspects of a complex phenomenon – so they can be conceptually split into several blocks of MVs – PLS-PMs provide a very useful framework for composite indicator building, in particular for second or higher-order constructs, called "hierarchical models".

As their name states, hierarchical models enable the definition of LVs of a higher order, linked to each other by reflective or formative relationships. To avoid confusion in the terminology we will call the composite indicator the latent variable defined/estimated within the measurement model, and the complex indicator the latent variable identified as a result of the interaction among the latent variables of the structural model.

For example, Cataldo et al. (2020) conceive global sustainability as a third-order construct. The following figure represents the structural model.

Figure 3.2
The SDGs theoretical framework (Cataldo et al., 2020)



The most common approach used to estimate a hierarchical model is the repeated indicator approach, which consists in taking all the blocks of MVs of the lower-order CI and using them as a unique block of MVs of the higher-order LV (complex indicator).

As an output of the estimation of such a model we obtain two kinds of weight: one set measuring the impact of each MV on the corresponding composite indicator, the other set measuring the first and higher-order path coefficients, i.e., the relations between the composite and the complex indicators in the system.

These two levels of weights help us to understand the different aspects that affect the complex phenomenon. The model also provides composite and complex indicator scores.

3.4 Are data-driven models suitable for composite indicators?

As we have seen in this brief overview, data-driven models, and in particular those designed for the identification and measurement of latent variables, offer an extremely flexible and versatile environment, suitable for shedding light on the components of a multidimensional phenomenon and on their interactions, including when it comes to interactions of subsequent, hierarchical levels.

However, when wanting to use these models to define composite or complex indicators there is a significant drawback to account for, a limitation of all data-driven models: changing the data changes the weights, consequently the indices obtained are not comparable in time or space. It is a heavy limitation if – as in the context to which we refer – composite indicators are recognized as tools in policy analysis and public communication for comparing country performance. It is a heavy limitation of PCA, FA and PLS-PM.

References

Bandura, R. (2008). A survey of composite indices measuring country performance: 2008 Update. UNDP/ODS Working Paper, New York.

Benzecrí, J.P. (1980). Pratique de l'analyse des données, analyse des correspondances. *Expose´ Elementaire*. Paris: Dunod, Bordas.

Cataldo R., Crocetta, C., Grassia, M.G., Lauro, N.C., Marino, M., Voytsekhovska, V. (2020). Methodological PLS-PM framework for SDGs system. *Social Indicators Research*. 10.1007/s11205-020-02271-5.

Cherchye, L., Moesen, W., Rogge, N. et al. (2007). An introduction to 'Benefit of the Doubt' composite indicators. *Social Indicators Research*,*82*, 111-145.

Chin, W.W. (1998). The partial least squares approach to structural equation modeling. In G.A. Marcoulides (Ed.), *Modern business research methods* (pp. 295-336). Mahwah, N.J.: Lawrence Erlbaum Associates.

Davino, C., Dolce, P., Taralli, S., & Vinzi Esposito, V. (2018). A quantile composite-indicator approach for the measurement of equitable and sustainable well-being: A case study of the italian provinces. *Social Indicators Research*, *136* (3), 999-1029.

Kaufmann D., Kraay A., & Zoido-Lobatón, P. (1999), Aggregating governance indicators. Policy Research Working Papers, World Bank, <http://www.worldbank.org/wbi/governance/working_papers.html>.

Lauro N.C., Grassia M.G., & Cataldo, R. (2018). Model based composite indicators: New developments in partial least squares-path modeling for the building of different types of composite indicators. *Social Indicators Research*, *135*(2), 421-455.

Mazziotta, M., & Pareto, A. (2016) On the construction of composite indices by principal components analysis. *Rivista Italiana di Economia, Demografia e Statistica*, *LXX* (1), Gennaio-Marzo.

Mazziotta, M., & Pareto, A. (2019). Use and misuse of PCA for measuring well-being. *Social Indicators Research*, *142*(2), 451-476.

Nardo, M., & Saisana, M. (2008). *OECD/JRC handbook on constructing composite indicators. Putting theory into practice*. Brussels: European Commission-Joint Research Centre; Institute for the Protection and

Security of the Citizen Unit of Econometrics and Applied Statistics.

Saisana, M., & Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for composite indicator development. European Commission-JRC, EUR 20408 EN, Ispra.

Trinchera, L., Russolillo, G., & Lauro, C.N. (2008). Using categorical variables in PLS path modeling to build system of composite indicators. *Statistica Applicata*, *20*(2), 309-330.

Trinchera, L., & Russolillo, G. (2010) On the use of structural equation models and PLS path modeling to build composite indicators. Working paper n. 30 September 2010, Dipartimento di Studi sullo Sviluppo Economico, University of Macerata.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K.G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Volume 2, pp. 1-54). Amsterdam: North Holland.

PART 2

OPEN ISSUES

Chapter 4

Subjective and objective indicators for measuring well-being[*]

4.1 Why we need to measure well-being

One of the main issues when analysing multidimensional phenomena such as well-being is how to define a composite indicator (Terzi and Moroni, 2020). The first – and still ongoing – debate is about the variables and indicators that have to be included in a well-being index, and their nature. This chapter aims to shed some light on the issue, presenting some of the most important well-being indicators and their characteristics.

If we consider the economic history in the last 300 years, the First and the Second Industrial Revolutions marked a departure from the economic situation of the Middle Ages, which can be described as a stagnating trend. Several major inventions led to the possibility of automating many traditional industries. Textiles, pottery and glassware, followed by chemicals, steel and heavy machinery, to name some of the most important technology clusters of industrialization and economic growth (Grubler, 1995), expanded production and started attracting workers from the primary sector.

The first major effect of the industrial revolutions was an accelerated growth in industrial output, which made up a large share of GDP. For the period 1880-1980, estimates indicate an annual growth of 3.5% (Grubler, 1995), which followed an exponential pattern in the second part of the nineteenth century (Grubler, 1995).

This progress was tracked from the mid-nineteenth century by the most advanced economies at that time. Measures like GDP and industrial output were collected and produced to document material progress. Other statistics that were widely used measured workers' incomes, expressed as gross or net incomes. Based on these, there is

[*] This chapter is mostly written by Elena Grimaccia and Adrian Otoiu.

evidence that real wages started growing in the industrialized world. In Britain, which is considered the first industrial nation, income started to post strong increases from the second quarter of the nineteenth century (Harley, 2014), which shows an exponential trend compared to the periods before, when wages rose only when population declines occurred.

These developments were followed by other industrialized countries (Germany, Italy, Spain) where urban wages exhibited similar growth patterns.

However, the unprecedented progress in material wealth was not fully accompanied by an improvement in quality of life. Accounts from the early stage of the Industrial Revolution point to unsanitary living conditions: "sun-up to sun-down workdays" (Laslett, 2014). Other accounts show that often work was performed by all family members (Rimlinger, 1989). Only later in the nineteenth century did concerns about wage levels, working conditions and rest periods emerge, and later triggered the enactment of labour standards and rest periods.

It is fair to say that well-being and human progress were traced, in an initial phase, primarily through the use of indicators that were objective measures of it. However, there were growing perceptions that these measures were not enough. In 1934, Kuznets, in contributing to the standardization of gross national product measures for the US Department of Commerce, observed that "the welfare of a nation can scarcely be inferred from a measurement of national income" (Otoiu et al., 2014). Similar concerns expressed by other scholars (e.g., Nordhaus and Tobin, 1973; Myrdal, 1968) have led to the acknowledgement that well-being is a multidimensional concept, and that GDP, economic growth measures or similar proxies are mostly unidimensional, providing only a narrow measure of it (Otoiu et al., 2014).

The last few decades of the twentieth century established other measures of human progress that incorporated other measures that define the multidimensional concept of well-being. A synthesis of these indicators was carried out by the Stiglitz-Sen-Fitoussi Report, which defines eight main dimensions of well-being: 1) Material living standards (income, consumption and wealth); 2) Health; 3) Education; 4) Personal activities including work; 5) Political voice and governance; 6) Social connections and relationships; 7) Environment (present and future conditions); and 8) Insecurity, of an economic as well as a physical nature (Stiglitz et al., 2009). Among them, the use of income- and consumption-based measures places an emphasis on household wealth, gives greater consideration to the distribution of income and wealth, takes into account both measures of objective and subjective well-being and employs a practical way to measure sustainability, particularly the environmental aspects of sustainability (Otoiu et al., 2014).

A synthesis of the concomitant use of both objective and subjective measures in building a composite index was carried out by Michalos et al. (2011) and is, to a good extent, incorporated in the Canadian Index of Well-being. According to them, indicators of well-being and life satisfaction are only marginally affected by evolutions of objective measures such as the state of the environment, poverty and sustainability (Otoiu et al., 2014). Factors such as personal circumstances, life experiences and, most importantly, discrepancies between real conditions and some desired or achievable conditions play a major role in defining subjective well-being (Otoiu et al., 2014).

While subjective indicators are valuable, if not indispensable, components of human being and societal progress, it is the perception that both types of indicators are needed for the construction of an index of well-being.

4.2 What is "subjective" and "objective"?

Before going further in the analysis of the use of "objective and subjective" variables in composite indicators, it is worth clarifying what we intend with the use of the word "subjective" in social science.

First of all, the term "subjective" can be used with reference to the definition of phenomena. In this framework, the definition of phenol-mena is always subjective since we consider "objective" to be what is actually shared: a phenomenon may be considered to be "objective" if it is observable and if there is a high degree of intersubjective agreement on what is observed (Brulè and Maggino, 2017). The degree of consen-sus on what is observed constitutes a phenomenon's measurability.

Then secondly, the definition of "subjective" can also be attributed to the way we measure phenomena. In this second framework, the measurement methods of any phenomenon must allow us to produce comparable measurements in different places and for different people (reliability and validity of the measurements). Therefore, "subjective" measures, in the sense that they measure a phenomenon in a variable way over time, in space or in different areas, are absolutely to be avoided.

Finally, when we analyse reality, we realize that there are some aspects that are directly observable and others that are not. Some phenomena are irreducibly subjective because they are not directly observable while subjective answers of the respondents are possible: evaluations, satisfactions, perceptions. These subjective variables can be measured by employing appropriate tools, such as binary or rating scales based on subjective assessments. As with objective measures, intersubjective agreement, particularly within a culture, may allow for significant comparability.

In building a composite indicator of well-being we refer to the latter definition.

4.3 "Objective" and "subjective" indicators in composite indicators of well-being

The efforts to operationalize the well-being concept have been made in two different directions: one is the level of living approach and the other is the quality of life approach. While the former focuses exclusively on resources and objective living conditions, the latter emphasizes the subjective well-being of individuals as a final outcome of conditions and processes (Noll, 2004). While objective social indicators are statistics that represent social facts (independent of personal evaluations), subjective social indicators are measures of individual perceptions and evaluations of social conditions. This distinction is also reflected in the recent discussion on quality of life by the Stiglitz Committee (Stiglitz et al., 2009).

The most widely used and best-known composite indicator of well-being is the Human Development Index (HDI), as discussed in Terzi and Moroni (2018). The HDI was created under the human development approach, of the economist Mahbub Ul Haq (1996). The theoretical perspective of the Index is anchored in the Nobel laureate

Amartya Sen's work on human capabilities (1993). The HDI is a measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and having a decent standard of living. The health dimension is assessed by life expectancy at birth, the education dimension is measured by the mean number of years of schooling for adults aged 25 and more and expected years of schooling for children of school entering age. The standard of living dimension is measured by gross national income per capita. Life expectancy at birth corresponds to the number of years a new-born could expect to live if prevailing patterns of age-specific mortality rates at the time of birth stayed the same throughout the infant's life. Mean years of schooling is the average number of years of education received by people aged25 and older, converted from education attainment levels using official durations of each level. Gross national income (GNI) per capita is calculated as the aggregate income of an economy generated by its production and its ownership of factors of production, less the incomes paid for the use of factors of production owned by the rest of the world, converted to international dollars using PPP rates, divided by midyear population (Grimaccia and Naccarato, 2019; Ul Haq, 1996). The HDI is the geometric mean of the normalized indicators.

This index does not account for inequalities within each dimension across the population (Terzi and Moroni, 2018), thus an Inequality Adjusted HDI (IHDI) has been built by the United Nations Development Programme (see Chapter 6).

The idea of "people being the real wealth of a nation" (Ul Haq, 1996) was of course not only revolutionary but also very useful for measuring human development and for designing corresponding policies. But the HDI suffers from several drawbacks, and criticisms started from the very beginning: most critics have focused on the selection of indicators, high correlation between components, computational form and component weighting (Chowdhury, 1991; Otoiu et al., 2014).

A key development in measuring well-being and social progress is the inclusion of both subjective and objective indicators. In Diener and Suh's view (Diener and Suh, 1997), social indicators and subjective well-being measures are based on different definitions of quality of life, but despite the conceptual and methodological differences between social indicators and subjective well-being (SWB), scientific approaches to well-being need to take a comprehensive view of the phenomenon by incorporating the strengths of each perspective.

Research has acknowledged that while subjective indicators are important, as they aggregate the evolution of several factors considered important in well-being and social progress at the individual level (Michalos et al., 2011), they cannot account entirely for the measurement of well-being and social progress (Otoiu et al., 2014).

Subjective well-being has risen in prominence as a way of assessing human progress.

There are a number of composite indicators that take into account both subjective and objective variables. Among the indicators that try to gauge subjective well-being, perhaps the most well-known is the Ladder of Life. Also known by its formal name, 'Cantril's Ladder of Life Scale', it consists of answers to the following question (OECD, 2013, citing Bjørnskov, 2010): 'Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. If the top step is 10 and the bottom step is 0, on which step of the ladder do you feel you personally stand at the present time?'

Ladder of Life estimates are collected via the Gallup World Poll (2009) for more than 150 countries, representing 98% of the world population.

Other well-known measures of subjective well-being are produced by the World Values Survey (OECD, 2013, citing Bjørnskov, 2010), as an answer to the following question: 'All things considered, how satisfied are you with your life as a whole these days? Using this card on which 1 means you are "completely dissatisfied" and 10 means you are "completely satisfied", where would you put your satisfaction with life as a whole?'

The Happy Planet Index measures sustainable well-being, defining the latter as: "How satisfied the residents of each country say they feel with life overall, on a scale from zero to ten, based on data collected as part of the Gallup World Poll". It has been presented as an index of efficiency, since it compares how efficiently residents of different countries use natural resources to achieve long, high-well-being lives.

Three dimensions are considered: Experienced well-being, Life expectancy and Ecological footprint. The Experienced well-being is based on the above-cited Cantril Ladder surveyed in the Gallup World Poll. Life expectancy is calculated as the average number of years a person is expected to live for in each country based on data collected by the United Nations. The Ecological footprint is the average impact that each resident of a country has on the environment, based on data prepared by the Global Footprint Network. Ecological footprint is expressed using a standardized unit: global hectares (gha) per person. The HPI scores are calculated by multiplying the mean life expectancy of residents of a given country by the mean experienced well-being of residents in the same country. Dividing the results by the ecological footprint provides the index "weighted" by the impact of well-being on the environment. Since 2016, the index has been adjusted according to inequality (see Chapter 6).

Well-being dimensions in the SSF report are incorporated in several composite indicators, including the Legatum Prosperity Index (LPI) (Otoiu et al., 2014). The latest edition of the LPI is an aggregation of 12 subindexes, namely Safety and Security, Personal Freedom, Governance, Social Capital, Investment Environment, Enterprise Conditions, Market Access and Infrastructure, Economic Quality, Living Conditions, Health, Education, and Natural Environment (The Legatum Institute, 2019). Several subindexes are built using subjective measures (indicators) collected via polls (e.g., perceived tolerance of immigrants, confidence of national governments, ability to live on a household income, satisfaction with freedom, emotional well-being) alongside objective indicators, such as occupational mortality, unemployment, depressive disorders, etc.

International official institutions carried out a consistent investment in studies on well-being and quality of life in order to answer the emerging knowledge demand. They paid great attention to measuring subjective well-being, applying indicators that allowed them to compare human conditions in different social systems (Conigliaro, 2017). In 2011, the OECD produced the Better Life Index, and later published the methodological volume *Guidelines on Measuring Subjective Well-being* (2013). According to Durant (2015), "current" well-being is measured in terms of outcomes achieved in two broad domains: material living conditions (i.e., income and wealth, jobs and earnings, housing conditions) and quality of life (health status, work-life balance, education and skills, social connections, civic engagement and governance, environmental quality, personal security and life satisfaction). The Better Life Index has 11 domains and 24 indicators and is synthesized by a weighted arithmetic mean (with subjective weights).

The Sustainable Development Goal Index is based on the 17 Sustainable Development Goals identified in the 2030 Agenda (UN, 2015) and is calculated on 114 of the indicators identified for monitoring the SDGs.

Composite indices calculated at the nation level with the aim of being utilized in policy realms (Yi, 2009) are the above-mentioned Canadian Index of Well-being in Canada, the People's Life Indicators (PLI) in Japan and Gross National Happiness (GNH) in Bhutan.

The GNH Index explores each person's life in nine domains: (1) psychological well-being;(2) health; (3) education; (4) time use; (5) cultural diversity and resilience; (6) good governance; (7) community vitality; (8) ecological diversity and resilience; and lastly (9) living standards.

Although there are so many important examples of composite indicators of well-being that include subjective variables, also this approach requires making assumptions about the weights to apply to each life domain as well as the universality with which those weights apply across the population.

4.4. Critical issues on the use of both objective and subjective indicators in a composite indicator of well-being for the comparability of individuals and countries

Nowadays, many scholars and policy makers use subjective indicators (Glatz and Eder, 2020), since the measure of subjective well-being (SWB) can help decision makers design policies that could improve quality of life of people beyond economic growth (Diener et al., 2015). This approach focuses upon individuals' appraisals of their own life, and several surveys collect these data (Brulè and Maggino, 2017).

On the individual level, social and economic determinants have been identified as related to SWB, such as state of health, family status, social relationships, income, etc. (Dolan et al., 2008; Puntscher et al., 2015). The evidence suggests that good health, rich social capital, decent standard of living can increase SWB.

However, on the country level, the most important debate is related to the feasibility of creating an internationally valid measure for the many diverse cultural and socioeconomic contexts in the world. Moreover, according to the well-known Easterling paradox (Easterlin, 1974), the relationships between income and happiness at the country's level is not straightforward in the long term, even if, in the short-term, happiness and income present the same pattern (Easterlin et al., 2010).

"Well-being is difficult to define but it is even harder to measure. In general, well-being measures can be classified into two broad categories: objective and subjective measures. The first category measures well-being through certain observable facts such as economic, social and environmental statistics" (Conceição and Bandura, 2008). This review paper has made an important distinction between subjective and objective factors of defining well-being and social progress, and the interplay between them. While intuition may indicate that there could be similarities between them, as they refer to describing similar things, research reveals that there could be a disconnect between the two.

This disconnect has been synthesized by Michalos et al. (2011) when designing the Canadian Index of Well-being: "Viewed from one perspective, a person may be well off, but from another not at all well off. A poor person might have good family relations and spiritual fulfilment while living in a rough neighbourhood with substandard housing. Someone with a good job may suffer from a long and lonely commute every day."

The importance of the subjective aspects of well-being and their relationship with objective well-being has been empirically explored in several papers (Otoiu et al., 2014). Among them, a salient example is the one by Stevenson and Wolfers (2009), which explored the levels of happiness among women in the 1970-2006 period. Their results show that women in the US have exhibited decreased levels of happiness in recent decades, while their condition has improved in terms of health, social status and empowerment. Research has shown that, while women have experienced major gains in terms of empowerment and income, an increasing ability to have and hold jobs over a longer period of time with career advancement prospects, and the ability to accommodate family and work duties, their reported levels of happiness have not followed suit. In fact, they have decreased below the levels of happiness experienced by men starting from 1985 onwards, with a growing gap that is becoming increasingly worse for women. It appears that an increasing complexity of life, which has diverged from the classical pattern of being employed in secretarial/assistant jobs until marriage, followed by a "stay-at-home mom" status for the rest of her life (Atwood, 1968), is the main factor behind this observed decrease in happiness.

Chile is a recent example of how policies foster economic growth and private enterprise. While evidence has shown that median wages, expressed in real terms, have shown a steady increase, and poverty levels dropped from 28% in 2006 to about 8% in 2017 (BBC, 2019), the inequality has taken a heavy toll on the Chilean society, which is the most unequal among OECD countries (BBC, 2019). The tensions caused by the growing social divide erupted in the autumn of 2019 and have shown the limits of a socio-economic model based solely on free market principles, as enshrined by the Pinochet government through a fraudulent plebiscite in 1980 (Bonnefoy, 2020), which guarantees free market provision for services that are traditionally in the public domain, such as health care, education and social security (Bonnefoy, 2020). Following the public unrest, a referendum on the change of the current Constitution was approved, with 78% of voters favouring the draft of a new Constitution, where one of the main purposes, apart from reversing the dictatorial legacy of General Pinochet, was to guarantee social rights more than market conditions. This push towards a more equal society is meant to improve the fortunes of many Chileans, as about three-quarters of household income is used to pay debt, while small pensions force many people to work after retirement, and make education and health care prohibitively expensive, if not unaffordable, for many ordinary citizens.

References

Atwood, M. (1968). *The edible woman anchor*. London: Hachette Digital.

BBC (2019). Chile protests: Is inequality becoming worse? BBC News, 21 October 2019.

Conceição, P., & Bandura, R. (2008). Measuring subjective well-being: A summary review of the literature. UNDP Working Papers.

Bonnefoy, P. (2020). An End to the Chapter of Dictatorship: Chileans Vote to Draft a New Constitution, *New York Times*, Oct. 25, 2020.

Brulè, G. & Maggino, F. (2017). *Metrics of subjective well-being: Limits and improvements*. Happiness Studies Book Series. New York: Springer.

Chowdhury, O.H. (1991). Human Development Index: A Critique. *The Bangladesh Development Studies*, *19*(3), 125-127.

Conigliaro P. (2017). Life Satisfaction and Subjective Well-being ina Sustainability Perspective. *Working papers Dipartimento di Scienze Sociali ed Economiche*, n. 7/2017.

Diener, E., Oishi, S., & Lucas, R.E. (2015). National accounts of subjective well-being. American Psychologist, 70(3), 234-242.

Diener, E.& Suh, E. (1997). Measuring quality of life: Economic, social, and subjective indicators. *Social Indicators Research*, *40*,189-216.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of Economic Psychology, 29(1), 94-122.

Durand M. (2015). The OECD better life initiative: how's life? And the measurement of well-being. *Review of Income and Wealth, 61*(1), 4-17.

Easterlin, R.A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P.A. David & M.W. Reder (Eds.), Nations and households in economic growth: Essays in honor of Moses Abramovitz (pp. 89-125). New York, NY: Academic Press.

Easterlin, R.A., Angelescu McVey, L., Switek, M., Sawangfa, O., & Smith Zweig, J. (2010). The happiness-income paradox revisited. *PNAS*, 107 (52) 22463-22468.

Gallup (2009). Understanding How Gallup Uses the Cantril Scale, accessed on November 11.

Glatz, C. & Eder A. (2020) Patterns of Trust and Subjective Well-Being Across Europe: New Insights from Repeated Cross-Sectional

Analyses Based on the European Social Survey 2002-2016. *Social Indicator Research, 148*, 417-439.

Grimaccia, E., & Naccarato, A. (2019). Food insecurity individual experience: A comparison of economic and social characteristics of the most vulnerable groups in the world. *Social Indicator Research*, *143*, 391-410.

Grubler, A. (1995). Chapter 4. Industrialization as a historical phenomenon. In R. Socolow, C. Andrews, F. Berkhout, & V. Thomas (Eds.), *Industrial ecology and global change*. Cambridge: Cambridge University Press.

Harley, C.K. (2014). Chapter 16. British and European industrialization. In L. Neal, & J.G. Williamson (Eds.), The Cambridge history of capitalism (Volume I). Cambridge: Cambridge University Press.

Haq, [Ul] M. (1996). *Reflections on human development* (1st ed.) Oxford University Press.

Laslett, J.H.M. (1989). Chapter VI. State policy toward labour and labour organizations, 1830-1939: Anglo-American union movements. In P. Mathias, & S. Pollard (Eds.), The Cambridge economic history of Europe (Volume VIII). Cambridge: Cambridge University Press.

Michalos, A.C., Smale, B., Labonté, R., & Muharjarine, N. (2011). The Canadian index of well-being. In: Technical Report 1.0. Vasa, Retrieved from <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf> on February 6, 2013.

Myrdal, G. (1968). *Asian Drama: An Inquiry into the Poverty of Nations*. New York: Twentieth Century Fund.

New Economy Foundation (2016). Happy Planet Index Methods.NEF, London.

Noll, H. (2004), Social Indicators and Quality of Life Research: Background, Achievements and Current Trends. In N. Genov (Ed.), Advances in Sociological Knowledge Over Half a Century. Wiesbaden: VS Verlag fur Sozialwissenschaften. OECD (2013).

Nordhaus, W., & Tobin, J. (1973). Is Growth Obsolete? In Moss, M. (Ed.), The Measurement of Economic and Social Performance. Studies in Income and Wealth. Vol. 38. Cambridge, Mass: NBER.

OECD (2013). *OECD Guidelines on measuring subjective well-being*. OECD Publishing. <http://dx.doi.org/10.1787/9789264191655-en>.

Otoiu, A., Titan, E., & Dumitrescu, R. (2014). Are the variables used in building composite indicators of well-being relevant? Validating composite indexes of well-being. *Ecological Indicators*, *46*, 575-585.

Puntscher, S., Hauser, C., Walde, J., & Tappeiner, G. (2015). The impact of social capital on subjective well-being: A regional perspective. Journal of Happiness Studies, 16(5), 1231-1246.

Rimlinger, G.V. (1989). Chapter VII, labour and the state on the continent, 1800-1939. In P. Mathias, & S. Pollard (Eds.), The Cambridge economic history of Europe. Cambridge: Cambridge University Press.

Sen, A. (1993). Capability and well-being. In S. Nussbaum (Ed.), The quality of life. Oxford: Clarendon Press.

Stevenson, B., & Wolfers, J. (2009). The paradox of declining female happiness. *American Economic Journal: Economic Policy*, *1*(2), 190-225.

Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). Report by the commission on the measurement of economic performance and social progress. In Commission on the Measurement of Economic Performance and Social Progress, Paris, France.

Terzi, S., & Moroni, L. (2018) How can we compare rankings that are expected to be similar? An example based on composite well-being indicators. In A. Abbruzzo, E. Brentani, M. Chiodi, & D. Piacentino (Eds.), *Book of short papers*. SIS 2018.

Terzi, S., & Moroni, L. (2020). Local concordance and some applications. *Social Indicator Research*, <https://doi.org/10.1007/s11205-020-02312-z>.

The Legatum Institute (2019). The Legatum Prosperity Index™ 2019 available from <https://docs.prosperity.com/8115/8635/0367/The_Legatum_Prosperity_Index_2019.pdf>, [accessed on November 2, 2020].

UN (2015). Resolution adopted by the General Assembly on 25 September 2015 n. 70/1. Transforming our world: the 2030 Agenda for Sustainable Development, New York.

Yi, I. (2009). Quality of life index in Korea: Why we need it & how to apply. OECD World Forum on "Statistics, Knowledge and Policy", Busan, Korea.

Chapter 5

Composite indicators for dichotomous variables: a counting approach*

Counting approaches stem naturally within the construction of poverty indicators. In the one-dimensional case a poverty line is set, and the headcount ratio counts the percentage of households that fall below the poverty line.The first European use of an (implicit) poverty line was by the London School Board during the 1880s in order to exempt indigent families from paying school fees (Gillie, 1996). The poor were those whose household income was below the poverty line. Subsequently, moving to multidimensional approaches such as the basic needs and social exclusion approaches, lists of needs considered essential alongside minimum levels of achievements (cut-offs) would be specified. It is in such a context that counting the number of deprivations naturally emerged as a method of identifying the poor and of monitoring progress towards meeting basic needs. For example, in his pioneering study on poverty in the United Kingdom, Townsend (1979) constructed a deprivation index based on the number of dimensions in which a person was deprived.

Poverty measurement can be broken down into two distinct steps: identification and aggregation. In a unidimensional approach the identification step consists in defining an income threshold called the "poverty line" as a benchmark (either an objective or a relative benchmark), while aggregation means selecting a poverty index, for example a headcount ratio or – according to a different approach – a poverty gap.

The simplest and most widely used poverty measure is the headcount ratio, which is the percentage of poor people in a given population. A second index, the (per capita) poverty gap, identifies the

* This chapter is mostly written by Silvia Terzi.

aggregate by which the poor fall short of the poverty line income, measured in poverty line units and averaged across the population. Both indices can be seen as a population average, with the non-poor being assigned a value of '0'. The headcount ratio assigns a '1' to all poor persons, while the poverty gap assigns the normalized shortfall (the difference between their income and the poverty line, divided by the poverty line itself) before taking the population average (Alkire and Foster, 2007).

In a multidimensional setting a simple identification method is to aggregate all achievements into a single cardinal variable of 'well-being' or 'income' and use an aggregate threshold to determine who is poor. In this case, a person is poor if the monetary value of her/his achievements is below the aggregate threshold z. Indeed, this approach (i.e., aggregation before identification) treats the dimensional achievements as interchangeable/substitutable, converting them into one another regardless of dimension-specific cut-offs. However, when deprivations are involved, compensation among dimensional achievements and shortfalls is most often inconceivable. Take, for example, nutrition and education: in no case can being below the threshold in one of these dimensions be compensated by satisfying achievements in the other: undernourishment and illiteracy are not counter balanceable.

Consequently, as Alkire and Foster (2007, 2011) argue, it seems more appropriate/convenient to turn attention to other approaches.

A long tradition in social sciences has been concerned with measuring material deprivation by looking at a number of dichotomous indicators of living conditions, such as malnourishment or access to safe drinking water (as in the Multidimensional Poverty Index) or ownership of durables or the possibility of carrying out certain activities like a week of annual holiday away from home (as in Eurostat's Measuring Material Deprivation in the EU). The typical way to summarize the information has been to count the number of dimensions in which people fail to achieve a minimum standard, hence the label of "counting approach".

Alternative procedures sum over individuals first, to form a "dimension index" that represents total or average achievements in a single dimension, and then combines the "dimension indices" of the different dimensions. Examples are provided by Anand and Sen's (1997) poverty index, but also by the Human Development Index. Anand and

Sen recommend a human poverty index, *P*, based on three subindices related to "proportion expected to die before the age of 40", "illiteracy" and "economic deprivation". The Human Development Index is the aggregation (geometric mean) of three indices that represent the following dimensions: "a long and healthy life", "knowledge" and "decent standard of living".

The counting approach combines different elements of deprivation at the individual level, which are then summed over individuals to form an aggregate index for the country. It represents the simplest way to embed the association between deprivations at the individual level into an overall index of deprivation (Aaberge and Brandolini, 2014).

Different procedures invert the order of aggregation by first computing the proportions of people suffering in each dimension, and then aggregating these proportions into a composite index of deprivation. This different order of aggregation has the advantage that the proportions of deprived people can be derived from various sources and the composite index is easy to understand.

Moreover, if the dimensions of well-being are "independent" of each other, the order of aggregation does not matter, and the two approaches are equivalent. However, if they are dependent, and suffering from multiple deprivations has a more than proportionate effect on people's well-being – for example, because the loss of quality of life due to being both poor and sick far exceeds the sum of the two separate effects – the cumulative effects of multiple disadvantages may imply that an important aspect of hardship is missing.

Within the counting approach, a choice has to be made on who is to be considered (multidimensionally) poor. There are two different approaches: the union and the intersection approach. The union approach was theoretically formulated by Bourguignon and Chakravarty (2003), among others. Following this approach, a person is defined as poor if she/he is deprived in at least one dimension. According to the second, more extreme approach, the intersection, developed by Atkinson (2003), a person is identified as poor if she/he is deprived in all the d dimensions. Rather than selecting the union or the intersection approach, Alkire and Foster introduce a so-called "intermediate approach", defining a person as poor if she/he is deprived in at least k dimensions, where k is a number between 1 and d. They call this procedure the "dual cut-offs approach".

Alkire and Foster's dual cut-off method was introduced as a counting approach for ordinal or even dichotomous/binary variables and has led to the development of the Multidimensional Poverty Index (MPI).

It is based on two steps: identification (i.e., setting a threshold or reference value) and aggregation (i.e., counting how many thresholds are reached by each unit). It can conveniently be used to define a multivariate performance (composite) indicator such as quality of life or sustainable development; all that is needed is a goal to achieve for each variable, and this is independent of the scale of measurement of the variables (or their ordinal or cardinal nature).

As Alkire and Foster (AF) point out (2007), moving from the unidimensional to a multidimensional poverty framework raises issues, including which dimensions of interest, and how to set cut-offs and weights for each dimension, but also challenges such as: at what point in the analysis should interactions between the dimensions be reflected? One of the most interesting features of their dual cut-offs approach is the incorporation of information related to the association between the different variables/dimensions/components.

Suppose we have different areas of well-being or performance (or poverty as in the original context) and no natural definition of an aggregate variable. The different areas or dimensions could be – and indeed often will be – measured on an ordinal scale, such as: subjective perception of well-being, customer satisfaction/appreciation in the assessment of service quality, standards of living or years of school, etc.

For each dimension, one could define a specific threshold/ reference value or cut-off (as AF call it) and identify who is above and who is below each of these one-dimensional thresholds.

The second step consists in establishing a second reference value (or second-level cut-off), usually indicated with k, to define as multidimensionally effective (poor in the original context) the unit that exceeds the first-order threshold in at least k dimensions or key indicators. In other words, the second cut-off value defines how many successes/deprivations a unit must record in order to be defined as effective (or poor) *tout court*. If we set $k=d$ this would lead us to the intersection-based approach, i.e., to consider multidimensionally effective the units that reach or exceed the reference thresholds in all key indicators.

On the other hand, if we set the second cut-off value equal to 1 (i.e., $k=1$) this would lead us to the union-based setting: an effective unit is successful in any of the key indicators. For $1<k<d$ we have intermediate solutions, and this is one of the advantages of the method.

Let w_j ($j=1,...,d$) be the weight applied to the j-th dimension, and let $w_j = d$, so that the weights wj of the different dimensions add to the total number of areas d. Let c_i ($i=1,...,n$) be the weighted number of achievements reached by the i-th unit; choose a performance cut-off k such that $0 < k \leq d$, and define as multidimensionally effective the unit whose achievement count c_i is $\geq k$. Let q be the number of effective units and let $c_i(k)$ be the count of the (weighted) achievements only for the effective units. P_0, the well-being indicator, can be defined as: $P_0 = \Sigma c_i(k)/nd$, i.e., the weighted average of the number of achievements in the population.

P_0 can also be expressed as a product between two measures: the incidence of effective units (H) and the intensity of achievements (A); more precisely: $P_0 = H x A$, where $H=q/n$ and $A=\Sigma c_i(k)/dq$.

It is logical to expect that, as the cut-off k varies, both the degree of incidence and the intensity will change. More specifically, as the second cut-off k increases, H is reduced because fewer and fewer units will be able to obtain a sufficient number of achievements; but at the same time, the positive variation of k increases A, producing an opposite effect on the final indicator P_0.

In contrast, by choosing a lower value for k, the increase in incidence contrasts with the reduction in intensity, with an uncertain effect on P_0, an effect that depends on the individual univariate distributions.

It is important to underline two other important properties of the methodology proposed by AF: multidimensional monotonicity and decomposability by subgroups.

The former property implies that if an additional achievement is recorded for a statistical unit, the overall index increases.

The decomposability, on the other hand, is based on the fact that if there were two distinct populations x and y, of n_x and n_y units, the index P_0 referring to the union of the two populations will be the average of

$P_0(x)$ and $P_0(y)$ weighted with their respective size.

As the deprivation score counts the number of dimensions in which an individual fails to achieve the minimum standards, it is by definition a discrete variable ranging from *0* to the number of dimensions considered. The distribution of deprivation scores contains all the relevant information in the counting approach, which by construction implies neglecting levels of achievement in the original variables. Consequently, this approach is also suitable whenever we want to combine dichotomous and continuous variables in the same index; all we need is to define a benchmark, for example a specific goal (as in sustainable development), or a relevant quantile for each cardinal variable and transform it into an ordinal or even a dichotomous variable (effective/non-effective; deprived/non-deprived). Of course, this dichotomous transformation could lead to some loss of information. However, this loss could be compensated by a particularly advantageous feature of this methodology: the composite indicator embeds information concerning the association between the different dimensions.

References

Aaberge, R., & Brandolini, A. (2014), Multidimensional poverty and inequality, No 976, Temi di discussione (Economic working papers), Bank of Italy, Economic Research and International Relations Area.

Alkire, S., & Foster, J. (2007). Counting and multidimensional poverty measurement. Oxford Poverty & Human Development Initiative OPHI Working Paper 7.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement *Journal of Public Economics*, *95*, 476-487.

Anand, S., & Sen, A.K. (1997), Concepts of human development and poverty: A multidimensional perspective. Human Development Papers, United Nations Development Programme, New York, 1997.

Atkinson, A.B. (2003). Multidimensional deprivation. Contrasting social welfare and counting approaches. *Journal of Economic Inequality*,*1*, 51-65.

Bourgignon, F., & Chakravarty, S.R. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, *1*(1), 25-49.

EUROSTAT 2012. Measuring material deprivation in the EU. Indicators for the whole population and child-specific indicators. Eurostat Methodologies and Working Papers.

Gillie, A. (1996). The origin of the poverty line. *The Economic History Review,49*(4), new series, 715-730. doi:10.2307/2597970

Townsend, P. (1979). *Poverty in the United Kingdom*. London, Allen Lane and Penguin Books.

Chapter 6

An overview of weighting systems*

6.1 Introduction to aggregation and weights. What are they and why do they matter?

Weighting is one of the most important steps in building composite indicators. In fact, it is the point at which the actual structure of the index begins to take shape. At this stage, the input variables that have previously been selected and transformed using normalization, or a similar procedure, are fed into the structure of an indicator.

Its importance stems from the fact that, once the weighting is done, it is likely to affect the scores of the composite indicators and, consequently, the rankings of the countries (OECD, 2008), or other entities that the data refers to. Therefore, weighting is a major issue, and often a major point of contention when it comes to building a composite indicator, analysing and checking its results. This is due to the fact that weights are essentially value judgements (OECD, 2008) about the importance of a specific factor/variable and its potential impact on the multidimensional concept that is being described by it.

Because of this, weighting is closely linked with another step of building composite indicators, namely aggregation. Aggregation follows immediately after the weighting step and is the instance where the index values are obtained, prior to being ranked. Very often the two steps are closely linked. First, aggregation affects weighting, as the choice of aggregation methods can alter the influence of the input variables, and therefore can raise the question of whether the weights assigned to them in a previous stage are appropriate. This relationship between the choice of weighting methods and aggregation methods appears more clearly in the next step of composite indicator construction, uncertainty and sensitivity analysis. At this stage, in which a review of the work done in

* This chapter is mostly written by Adrian Otoiu.

the previous stages is performed (OECD, 2008), critical questions are asked about the choice of weights and aggregation methods, and whether they are able to help yield plausible measures of the multidimensional concept that is being described.

Although closely linked, and, in some cases, being derived at essentially the same stage (OECD, 2008; Gan et al., 2017), as in the case of the Benefit of the Doubt or Unobserved Components model, weighting and aggregation are distinct methodological issues that employ different specific methods and elicit thorough specific knowledge of each of them. Only through achieving a thorough understanding of both will one be able to effectively use and analyse them, see their strengths and weaknesses, and be able to assess their adequacy or potential areas of improvement.

6.2 An overview of weighting methods

Weighting is done using a wide variety of methods. However, most of them fall into two main categories:

1) Weights derived using quantitative techniques. Widely used and acknowledged in the literature, these methods rely on statistical techniques to infer the importance of different variables in describing an overarching concept (e.g., well-being, sustainability), and take into account the relationships between variables when deriving them.

2) Weights derived using participatory methods. Given the fact that, in the process of developing composite indicators, weights are essentially value judgements (OECD, 2008), another main methodological approach for deriving weights is to use opinions. These approaches are also justified by the fact that, often, quantitative techniques have their limitations and may lead to biases (Greco et al., 2019), and that subjective judgements can be more transparent and subject to scrutiny than quantitative methods.

However, before starting a discussion about weights we should consider the most basic case, and perhaps the oldest method of weighting, which is equal weighting.

Equal weighting was one of the first ways of aggregating different input variables to produce an index, and most certainly one of the easiest ones. Its rationale stems from the fact that all input variables are deemed to have the same importance in describing a multidimensional concept. Perhaps the most salient example is the Human Development Index, whose structure assumes equal weights for the three major dimensions of wealth, health and education, as the main drivers of human development, and has remained roughly the same since its inception, despite its critics.

Notwithstanding the existence of more advanced weighting techniques, it seems that equal weighting is still the dominant one (OECD, 2008; Gan et al., 2017). Its popularity may lie not only in its simplicity and ease of replication, but also in the lack of a sound empirical basis, or the lack of a suitable alternative (Nardo et al., 2005; McGillivray and Noorbakhsh, 2004). Simplicity is a strong argument in favour of a composite index, which goes beyond a lack of a strong foundation, and favours the ease of understanding and interpretation brought by using it (Gan et al., 2017). However, its major drawbacks are the lack of insights into the relationships among the different components of an index (Gan et al., 2017), the risk of a double-weighting characteristic (Gan et al., 2017) when variable selection points to the existence of highly collinear variables being included and the fact that some dimensions get an unequal weighting due to the fact that they are described using more variables than others (OECD, 2008).

The existence of equal weighting rests on a major question, which is how well the other methods of weighting can be used for describing a multidimensional concept. Here, it seems that there is a grey area that could potentially be motivated by a lack of knowledge, or the impossibility of using an alternative method that can generate better results.

Quantitative weighting methods are widely used in building composite indicators. Using statistical techniques, information about the links between variables is extracted, and weights are derived based on that.

The most popular methods are: factor analysis/principal component analysis (FA/PCA), Benefit of the Doubt (as an application of data envelopment analysis), multiple linear regression analysis and the Unobserved Components model (OECD, 2008; Gan et al., 2017; Greco et al., 2019).

FA/PCA is the second-most popular method used in constructing composite indicators of sustainability (Gan et al., 2017). While it comprises a family of techniques, its main goal is to retain as much information as possible from the original dataset and extract a small number of factors (which can be broadly described as uncorrelated latent variables) that capture most of the total variability of the data (OECD, 2008). The method is useful in that, using correlations between input variables, it helps associate them with factors, the latter being natural candidates for describing distinct dimensions of a multidimensional concept that is formalized in a composite indicator. Furthermore, based on the factor loadings (or component weights in PCA methods), which describe the correlation between input variables selected at a previous stage (Larose and Larose, 2015), weights can be derived based on the highest factor loadings (OECD, 2018).

However, the use of FA/PCA methods is not always straightforward. While their use is extremely helpful at the multivariate analysis stage, when the relationships between input variables are explored, there are several drawbacks that make them less suitable for weighting. Among them, the most important are their limited use when only a few input variables are available. Another one involves the fact that factors are often not clearly distinct (Ferligoj, cited by Otoiu et al., 2014), which can lead to the inability to extract factors that can be used to describe relevant dimensions of well-being.

Multiple linear regression analysis is frequently used for constructing indicators. Among sustainability indicators, it appears to be the second-most popular analytical method used (Gan et al., 2017). It involves modelling the evolution of a variable, called a "dependent variable", as a function of several explanatory variables. The method can provide useful insights into the relationships between input variables and a variable that can constitute a subindex or a raw composite index value and can be particularly useful when there are many candidate input variables. The issues associated with multiple linear regression analysis, and their standard treatment, can limit their use and point to the use of other methods. Probably the main issue is multicollinearity, an issue that, if not properly addressed, can lead to wide variations in parameter estimates, high standard errors and coefficients with implausible values (Greene, 2012). Although multicollinearity may be addressed at the stage of selecting the candidate variables through correlation analysis, by selecting variables that do not exhibit high cross-correlations (mainly in the case of formative models of composite indicators), multicollinearity can be better assessed using a non-linear model that considers all variables and relationships among them.

An issue that is specific to composite indicator building is the choice of a dependent, or target, variable. This poses a challenge in the case of formative models, where the composite indicator should be "a sum of its parts" (OECD, 2008), and the composite index is largely unknown. However, this shortcoming can be overcome by using variables that are considered essential in the definition of a composite indicator. One example is the Legatum Prosperity Index[8], which uses regression analysis to derive weights for the input variables, using GDP/capita and overall well-being as dependent variables (Legatum Institute, 2013).

The Benefit of the Doubt (BOD) method is also known as Data Envelopment Analysis (DEA). While the two terms are used in the composite indicator literature, it is worth noting that it is appropriate to use BOD as the method applied for deriving weights of composite indicators, as it is an application of DEA (OECD, 2008). DEA is used to derive an efficiency frontier, which represents a benchmark for the units (countries) assessed (OECD, 2008). Then, their performance is assessed by measuring the (relative) distance of one unit from that frontier, and from a point of origin. In order to do this, the BOD approach is used to assess the relative performance by the ratio of its relative performance to the benchmark performance. Its distinctive feature is that the weights assigned to input variables are derived so that the highest (most favourable) value is achieved for a given country, by the choice of the most favourable set of endogenously derived weights (Gan et al., 2017). While the method has its advantages, its drawbacks mainly consist of multiple potential results and the incomparability of results (Nardo et al., 2005).

Among other methods used to derive weights using statistically based methods are the Unobserved Components method, which assumes that the overarching concept is not directly observable, and that it will be obtained through a linear combination of an unobserved

[8] Before 2019, when the methodology changed, and weights are derived using expert opinion.

component and an error term, and weights are calculated as a decreasing function of the variance of a specific input indicator, and an increasing function of the other input indicators.

Participatory methods are the second type of methods whose popularity is even greater than that of quantitative methods. Their simplicity and the quite comprehensible, transparent way of deriving weights based on expert opinion made them the second-most popular methods in the development of sustainability indicators (Gan et al., 2017).

The budget allocation process (BAP), also known as "expert opinion" (Saisana and Tarantola, 2002), consists in having several experts in the field distributing a number of importance points to a set of input indicators relevant for an issue, based on their knowledge of, and experience in, the subject (OECD, 2008). This ensures that, based on a number of informed opinions, input indicators judged as being more important receive a higher score. Weights are then derived based on the relative importance of the points given for each input indicator. While the method has the advantage of being explicit and fairly transparent (Gan et al., 2017), results can be subject to certain biases, and may lean, in some cases, towards policy goals instead of making an assessment of the state of affairs of a given concept or dimension. Also, in some cases, opinions may be a reflection of local conditions, which are not applicable or transferable elsewhere (Gan et al., 2017).

Public opinion is not mentioned as often as other participatory methods, perhaps due to its simplicity. In principle, this method is similar to BAP, except for the fact that here it is the public, via opinion polls or another similar method, that attributes importance points. While this method appears to be even more participative and transparent than expert opinion, its weaknesses are even stronger than those for BAP, as the public may focus on current concerns and does not generally have its own expertise in a particular field of interest, other than that acquired through media and public debates. However, despite its drawbacks, this appears to be the most widely used participatory method for devising sustainability indicators (Gan et al., 2017).

The Analytic Hierarchy Process (AHP) can be described as a structured technique used for decision-making and is based on pairwise comparisons of selected items. According to Gan et al. (2017), the AHP

is implemented in two main steps:
1) formalizing the composite indicator structure in a hierarchical structure (Gan et al., 2017), from its overarching concept (e.g., well-being), several criteria that could be seen as subdimensions of this concept (e.g., wealth, life satisfaction) and input indicators.
2) making (ordinal) pairwise comparisons between items on the same hierarchical level (Gan et al., 2017; OECD, 2008), showing which one is more important and by how much, using a scale with nine levels, ranging from equally important (1) to much more important (e.g., 9 means that one item is nine times more important than the one it is being compared with) (OECD, 2008; Greco et al., 2019).

The results are formalized in comparison matrices, which are used to derive weights, using an eigenvector method (OECD, 2008; Greco et al., 2019). While the method is deemed to yield much more consistent results than BAP and public opinion methods, it is still prone to a redundancy error just like in the case of the other two methods. However, this error is quantifiable by using the consistency ratio of the comparison matrices. Consistency ratios need to be small, i.e., 0.1, although 0.2 is sometimes deemed acceptable (Saaty, 1980), in which case they do not affect weights in a significant way. Another useful feature is that this method can also accommodate both qualitative and quantitative variables in the process of deriving weights.

Conjoint analysis (CA) is an analytical technique that is widely used in marketing. It essentially consists in allowing potential consumers to make choices between several items (products or services) based on the preferences expressed on different sets/combinations of features (Nikou, 2017). This technique was found useful in building composite indicators as it allows individuals to select their preferences among a set of alternative scenarios (OECD, 2008; Gan et al., 2017; Greco et al., 2019). The results, expressed in terms of preferences for one alternative over another, are then decomposed by relating the choices made to the values of the input indicators particular to the alternative scenarios (OECD, 2008). Decomposition is done based on a probability of preference for the alternative scenarios, expressed as a function of its input indicators (OECD, 2008; Greco et al., 2019). Then, the importance of each input is derived as its marginal rate of substitution, and is used to calculate weights (OECD, 2008; Greco et al., 2019).

CA has some particular features that distinguish it from other methods. Although it is known as a public opinion method, it is also acknowledged to be a statistical method (OECD, 2008). Its advantage comes from its approach, decomposing overall preferences into weights for the input variables, which is the opposite of other participatory methods, which derive weights based on the results obtained at input variable level. It is often compared to the AHP method, in that its approach is conceptually similar, but their approaches are different, as the AHP starts from the input variables and works its way up to obtain an overall weighting structure for the composite indicator, whereas CA uses a top-down approach based on the decomposition of alternative preferences into weights for the input indicators (Greco et al., 2019; Gan et al., 2017). Although this approach is very appealing from the theoretical point of view, and results could provide a better composite indicator structure by deriving weights based on revealed preferences, as opposed to explicitly obtaining them, it comes with several limitations. Estimation can become difficult from the computational point of view (OECD, 2008; Greco et al., 2019) and resource-intensive in terms of setting up a sample large enough to allow estimations (Greco et al., 2019; Gan et al., 2017). Moreover, the weights obtained rest on the assumptions of substitutability of the input indicators, a property that may not be desirable (OECD, 2008).

Table 6.1 – *Methods for indicator weighting*

Method	Type	Examples	Advantages	Disadvantages	References
Equal weighting	Equal Weighting	Human Development Index (UNDP, 1990)	- Easy to use and understand, replicable - Achievable temporal and spatial comparability	- No insights into indicator relationships - Risk of double weighting	Greco et al., 2019
Principal component analysis/ Factor analysis (PCA/FA)	Statistical	Environmental State and Sustainability Index (Oțoiu and Grădinaru, 2018)	- Reduces the risk of double weighting - Reveals relationships between input variables	- Factors can be difficult to extract - Cannot accommodate a mix of discrete and continuous variables - Weights may not reveal the actual relationships between variables - Weights may be inconsistent over time	OECD, 2008
Benefit of the Doubt (BOD/DEA)	Statistical	Meta-index of Sustainable Development (Cherchye and Kuosmanen, 2004)	- Integrates weighting, aggregation and index construction steps - Can accommodate different situations when good performance in several dimensions is not penalized by other negative outcomes - Can define a benchmark of potentially achievable as policy goals	- Results can be overly optimistic; poor performance in some dimensions may not affect the overall result - Multiple achievable solutions, leading to undetermined weights - Focus heavily tilted towards policy goals, ignoring the actual state of particular dimensions - Non-unique weights may pose problems of reproducibility and temporal comparability	OECD, 2008; Greco et al., 2019

Method	Type	Examples	Advantages	Disadvantages	References
Multiple linear regression analysis	Statistical	National Innovative Capacity (Porter and Stern, 2001) Legatum Prosperity Index (Legatum Institute, 2013)	- Simple results that are reproducible and updatable - Can accommodate variables that have low cross-correlations	- Lack of dependent variable(s) relevant to the index - Linearity cannot be assumed in the relationship between variables - Low R^2 values may prevent derivation of meaningful weights	Greco et al., 2019; Gan et al., 2017
Budget allocation process (BAP)	Opinion based	The Eco-indicator 99 (Goedkoop and Spriensma, 2001) e-Business Readiness Index (Pennoni et al., 2005)	- Transparent and easy to understand results	- Potentially affected by domain-specific biases (scientific consensus on an issue) or policy priorities - Variations in relevance across regions - Circular thinking may occur when a large number of input variables/indicators are used, generating inconsistent responses	OECD, 2008
Public opinion	Opinion based	Concern about environmental problems Index (Parker, 1991)	- Transparent and easy to understand results - Better fit concerning public concerns and expectations on the multidimensional concept measured	- Circular thinking may occur when a large number of input variables/indicators are used, generating inconsistent responses - Bias towards concerns reflected in public discourse rather than objective state of affairs	OECD, 2008; Greco et al., 2019

Method	Type	Examples	Advantages	Disadvantages	References
Analytical hierarchy process (AHP)	Opinion based	Composite sustainability performance index (Singh et al., 2009	- Hierarchical structure befitting of composite indicator frameworks - Achieved better consistency among opinions, which can be objectively measured - Can employ quantitative/continuous and qualitative/discrete data together	- Biases and circular thinking may occur when a high number of alternatives are available	OECD, 2008
Conjoint analysis (CA)	Opinion and statistical based	Indicator of quality of life in the city of Istanbul (Ülengin et al., 2001)	- Hierarchical structure befitting of composite indicator frameworks - Mitigates bias and circular thinking - Based on revealed preferences rather than explicitly stated ones	- Assumes additivity of weights - Computationally complex - Relatively large samples required - Potentially inconsistent results for certain categories of respondents	OECD, 2008

Source: author's compilation, Gan et al., 2017, OECD, 2008, Greco et al., 2019.

Table 6.1 summarizes the weighting techniques used so far and offers some examples and references for readers who wish to further their knowledge of the particular weighting techniques explained.

Apart from that, it is worth wrapping up the discussion on weighting systems with some general conclusions before moving on to the next topic.

The choice of weighting methods is related to the type of input variables and whether they can accommodate a mix of quantitative and qualitative variables (see discussion in Chapter 5).

Another issue pertains to the number of input variables. In some cases, having just a few of them can make the application of some methods useless (e.g., PCA/CA, which is essentially a data reduction method). In other cases, especially for participatory methods, having a large number of input variables can greatly diminish the relevance of the weights obtained, as the choice between them can be difficult to make, especially in the case of unstructured participatory methods (BAP and public opinion).

The features of some methods can pose challenges that may not be solved in a satisfactory way. Apart from PCA/FA, which cannot generate a solution, for example, in the case when input variables are not correlated, some methods may not yield consistent results, e.g., CA, BOD, etc.

The feasibility of implementation is another issue that needs to be taken into consideration when using participatory methods. Using them requires getting input from a panel of experts, or, in the case of public opinion, from a sample of respondents. This may amount to sizeable resources needed to obtain the opinions, and time needed to collect and process the results. By comparison, statistical-based methods usually use publicly available data, which are mostly available for free, and are produced based on established methodologies and standards that are usually available and accompany data releases. This makes these methods more appealing to researchers, who can thus build composite indicators and publish/release them fairly easily.

In comparing the two types of methods, there is always the question as to which approach can lead to the best results in terms of coming up with the most relevant weighting solution. Opinions are split between the two types of methods, as a predominantly quantitative approach may lead to results that may not be relevant to the overarching concept they attempt to describe. On the other hand, participatory approaches may be biased towards public concerns or policy goals and may ignore the actual state of affairs in a particular issue (Oțoiu and Grădinaru, 2018). However, this may be the only alternative when the overarching concept is tightly linked with it (e.g., governance or public perception issues/domains).

One last mention should address the actual analytical techniques used in deriving weights. Some of the issues mentioned when presenting the statistically based methods can be addressed by using alternative statistical techniques. For example, linear regression can be replaced with non-linear models. Given the structure of the indicators, panel data models that can accommodate time and country/region effects may be the ideal candidate for regression modelling, which can address the biases particular to specific regions/subregions and improve the time and geographic consistency of the estimates. When having discrete variables, alternative methods similar to PCA/FA can be used to yield comparable results (e.g., canonical discriminant analysis as shown in OECD (2008)). Orthogonal designs can help in generating a relatively parsimonious set of alternatives, and the use of suitable conjoint techniques such as adaptive conjoint analysis (Nikou, 2017) may also help mitigate the issues associated with this method (Nikou, 2017).

6.3 Aggregation: role, importance and main approaches

Aggregation is the next important step in building a composite indicator, as, at this stage, weights are applied to input variables (which, in most cases, undergo a process of transformation), and scores for the composite indicators, or, in many cases, for their component subindexes, are derived.

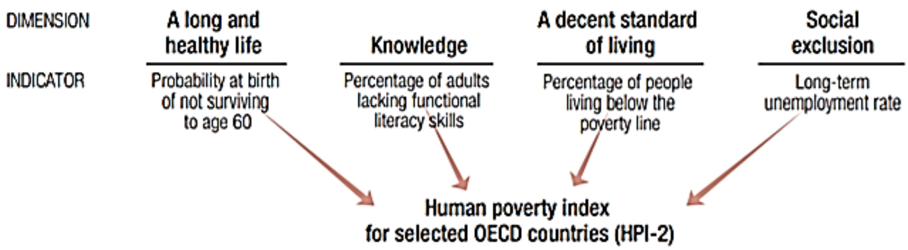
The relationship between weighting and aggregation illustrates the fact that most of the stages of building a composite indicator are linked between them. The choice of weights essentially defines the contribution of the selected input variables to the (final) composite indicator scores. The values obtained at this stage are closely related to the nature of the selected candidate variables (continuous and/or discrete) and the method of transformation that is applied to them (see Chapter 2). Also, aggregation is closely linked not only with weighting, but with variable selection, because the choice of the aggregation method needs to take into account the importance of each input variable, and the relationships between them (Greco et al., 2019; OECD, 2008). Thus, the choice of aggregation method is a natural continuation, and incorporates the insights and methods used in the previous stages.

Perhaps the most important consideration is the composite indicator structure. Some indicators use a simple structure. One example is the UNDP's Human Poverty Index for selected OECD countries (UNDP, 2007), which plugs input indicators directly into the formula:

$$\mathrm{HPI} = [1/4(\mathrm{P}_1^{\propto} + \mathrm{P}_2^{\propto} + \mathrm{P}_3^{\propto} + \mathrm{P}_4^{\propto})]^{1/\propto} \tag{1}$$

where P1-P4 are the input indicators and α is a weighting constant.

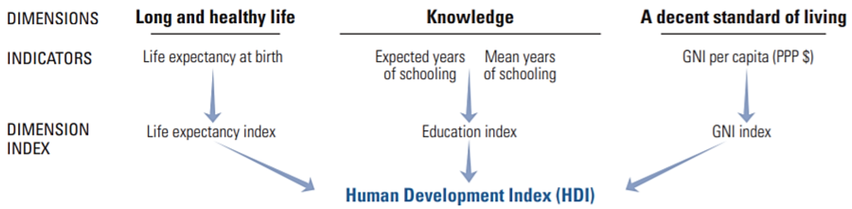
Figure 6.1
The structure of the Human Poverty Index for selected OECD countries
(UNDP, 2007)



Other indexes still keep a relatively simple structure, which diverges from the simple layout by adding in some degree of complexity. However, in many cases (e.g., Human Development Index, Multidimensional Poverty Index), simple weighting is used to weight the initial indicators that are used in the formula.

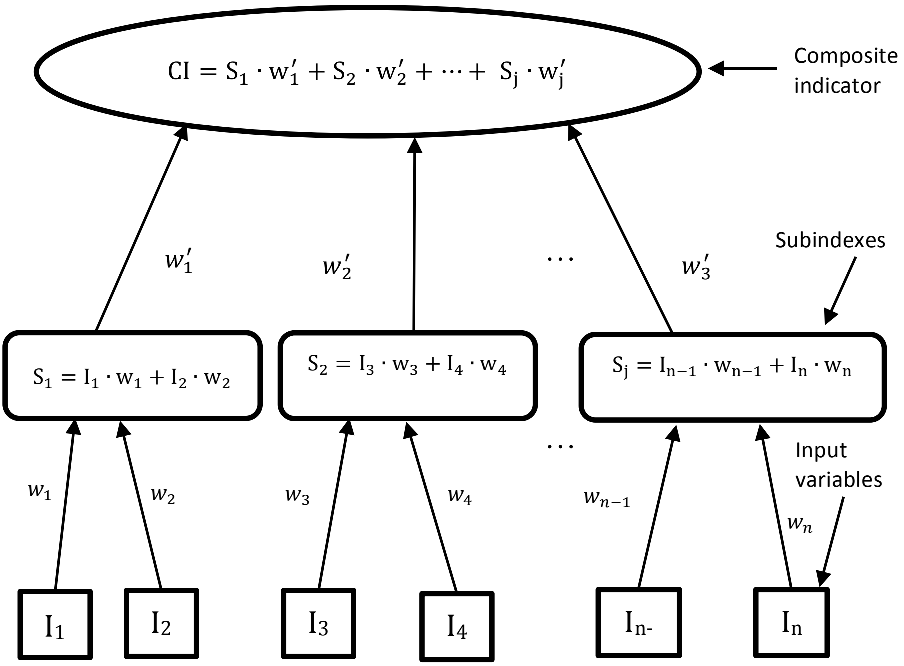
As an example, the Education Index, which is one of the three pillars used for computing the Human Development Index (HDI), is computed as a simple arithmetic average of the Expected years of schooling and Mean years of schooling input indicators, before the actual value of the index/dimension is calculated from the three input indexes.

Figure 6.2
Human Development Index structure (UNDP, 2019)



In many cases, especially when there are a large number of input indicators, the composite indicator has a hierarchical structure, very similar to the basic example presented in Figure 6.3. This corresponds to the conceptual framework of a composite indicator, where one complex, multidimensional concept has been decomposed into several subdimensions, for which subindexes are computed from the selected input variables that have undergone normalization.

Figure 6.3
The hierarchical structure of composite indicators



Note: the example above refers to the case of additive aggregation.

While, in many cases, aggregation is done through multiplication of weights by the corresponding input variables, the actual way in which aggregation is achieved can vary.

Different aggregation solutions need to keep in mind the nature of the input indicators, the characteristics of subindexes and the relationships between them.

The main question that needs to be addressed when choosing among alternative weighting methods is whether compensation is allowed among input variables – that is, whether an improved performance of one index can compensate for a lower performance of other indexes –. From the theoretical point of view, it is often the case that good performance in one area cannot adequately compensate for poor performance in another (e.g., in the case of material deprivation and environmental indexes); however, several weighting methods compute weights based on the assumption that marginal substitution is allowed among individual indicators (e.g., linear regression, CA).

The most common method is additive aggregation, when the weighted input indicators are added up, using the following general formula:

$$\mathrm{CI} = \sum_{i=1}^{m} \mathrm{I}_i \cdot \mathrm{w}_i, = \overline{1, m}\ , \mathrm{w}_i \in (0,1) \tag{2}$$

where I represent the input indicators, w their matching weights and m the number of indicators. This method assumes perfect substitution among the input indicators and is frequently used in the case of equal weighting, or when a suitable theory justifies this choice (Gan et al., 2017).

The second one is geometric aggregation, which is the geometric mean of the input indicators. A general formula is presented below:

$$\mathrm{CI} = \prod_{i=1}^{m} \mathrm{I}_i^{\mathrm{w}_i}, i = \overline{1, m}, \mathrm{w}_i \in (0,1) \tag{3}$$

where I represent the input indicators, w their matching weights and m the number of indicators. To a certain extent, this method addresses the substitution issue posed by additive aggregation. But the method cannot be considered non-compensatory, as geometric aggregation allows for some degree of compensation among variables (OECD, 2008). Its use addresses the case of diminishing marginal substitutability among items, in particular in the case where progress recorded by indicators with relatively low absolute values can be accounted for in a significant manner (OECD, 2008).

Both methods rest on the assumption that there is a certain preferential independence among input indicators. The preferential independence condition means that the trade-off ratio between two variables is independent of the values of the other variables (OECD, 2008). This also implies that weights express, to some extent, the marginal contribution of each input indicator to the composite index. However, preferential independence is found to be a very strong assumption (Greco et al., 2019; OECD, 2008), under which marginal contribution may not coincide with the importance of each input indicator (OECD, 2008; Greco et al., 2019), which should be the purpose of weights assigned to individual indicators. This results in obtaining biased composite indicator scores, for which there are no suitable adjustment procedures (OECD, 2008).

Due to the fact that, in some cases, the substitutability of input variables is deemed unacceptable (e.g., in the case of sustainability indicators, as per Gan et al. (2017)), or the resulting weights seem to diverge from the importance attributed to the input variables, non-compensatory aggregation techniques have become an appropriate choice.

These methods comprise two main approaches: the properties of the aggregation function (see Pollesch and Dale, 2015) and the multicriteria decision-making (MCDM) methods used to derive these weights (Gan et al., 2017).

The latter approaches are the most popular methods used in deriving non-compensatory weights, with the best-known one being the non-compensatory multicriteria approach (Nardo et al., 2005; OECD, 2008).

This approach, which is described in detail in the OECD *Handbook on Constructing Composite Indicators* (2008), has the following two steps (Nardo et al., 2005) : 1) creating an outranking matrix through first making pairwise comparisons of countries based on the whole set of input indicators, and then ranking countries in a complete initial order (Nardo et al., 2005); 2) processing the ranking matrix until a final solution is found. The best-known method used is the Condorcet-Kemeny-Young-Levenglick (CKYL) ranking procedure.

In the CKYL method, the final ranking of countries is the one supported by the maximum number of input indicators for each pairwise comparison, summed over all pairs of countries considered (Nardo et al., 2005).

While this method is an attempt to get non-compensatory weights, it is the least preferred among the three aggregation methods described. The CKYL method is deemed to be computationally expensive and, because it is rank based, it does not accommodate the size of the differences between input indicators (Munda and Nardo, 2005). Moreover, the method does not properly take into account the nominal importance of the input indicators (Paruolo et al., 2013, cited by Greco et al., 2019), which is the main objective of building a composite indicator that is relevant to the multidimensional concept that is being measured.

In spite of the fact that, in most cases, the composite indicator has a hierarchical structure, the weights for subindexes are not often determined using an analytical process. They usually stem from the importance given to each dimension and are assigned based on theoretical considerations rather than through the use of a weighting technique as described earlier in this chapter. The weighting techniques described above, for the case when there are several subindexes, are usually applied at subindex level.

Nevertheless, it is not plausible to assume that their determination is completely unrelated to the weights determined for the input indicators. As an example, if input variables are grouped into dimensions and those are further aggregated into composite indicators, then applying equal weighting to the variables may imply an unequal weighting of the dimension (the dimensions grouping a larger number of variables will have a higher weight). This could result in an unbalanced structure in the composite index.

References

Cherchye, L., & Kuosmanen, T. (2004). *Benchmarking sustainable development: A synthetic meta-index approach*. No. RP2004-28, WIDER Working Paper Series, World Institute for Development Economic Research (UNU-WIDER).

Gan X. et al. (2017). When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, *81*, 491-502.

Greco, S., Ishizaka, A., Tasiou, M. et al. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research, 141*, 61-94.

Greene, W. (2012). *Econometric analysis*, 7th ed. Upper Saddle River: Prentice Hall.

Goedkoop, M., & Spriensma, R. (2001). *The Eco-indicator 99: A damage-oriented method for life cycle impact assessment: Methodology report*. Product Ecology Consultants.

Larose, D.T., & Larose, C.D. (2015). *Data mining and predictive analytics,* 2nd ed. Hoboken, N.J.: John Wiley & Sons.

Legatum Institute (2013). *The Legatum Prosperity Index 2013. Methodology And Technical Appendix* [online]. Legatum Institute, United Kingdom, 2013. [cit. 02.11.2015] <http://media.prosperity.com/2013/pdf/publications/Methodology_2013_FinalWEB.pdf>.

McGillivray, M., & Noorbakhsh, F. (2004). *Composite indices of human well-being: Past, present, and future*, WIDER Research Paper, No. 2004/63, ISBN 9291906603, The United Nations University World Institute for Development Economics Research (UNU-WIDER), Helsinki.

Munda, G., & Nardo, M. (2005). *Constructing consistent composite indicators: The issue of weights*. Ispra, Italy: Institute for the Protection and Security of the Citizen, Joint Research Centre.

Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). Tools for composite indicators building. *European Commission, Ispra*, *15*(1), 19-20.

Nikou, S. (2017). Conjoint analysis. In M. Allen (Ed.), *The Sage encyclopedia of communication research methods* (Vol. 1, pp. 237-239). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781483381411.n89.

OECD (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: New York: OECD Publications.

Oţoiu, A., Grădinaru, G. (2018). Proposing a composite environmental index to account for the actual state and changes in environmental dimensions, as a critique to EPI. *Ecological Indicators*, 93, 1209-1221.

Otoiu, A., Titan, E., & Dumitrescu, R. (2014). Are the variables used in building composite indicators of well-being relevant? Validating composite indexes of well-being. *Ecological Indicators*, 46, 575-585.

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

Pennoni, F., Tarantola, S., & Latvala, A. (2005). The European e-business readiness index. Joint Research Centre (JRC) (2003-2008).

Pollesch, N. & Dale, V. (2015). Applications of aggregation theory to sustainability assessment. *Ecological Economics*. *114*, 117-127.

Porter, M.E., & Stern, S. (2001). National innovative capacity. In The global competitiveness report 2001-2002. World Economic Forum, New York: Oxford University Press.

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

Saisana, M., & Tarantola, S. (2002). State-of-the-art report on current methodologies and practices for composite indicator development (Vol. 214). Ispra: European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, Technological and Economic Risk Management Unit.

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

Ülengin, B., Ülengin, F., & Güvenç, Ü. (2001). A multidimensional approach to urban quality of life: The case of Istanbul. *European Journal of Operational Researches*, *130*(2), 361-374.

UNDP (1990). *Human Development Report 1990*. New York: Oxford University Press.

UNDP (2007). *Human Development Report 2007/2008*, New York: Palgrave Macmillan.

UNDP (2019). *Human Development Report 2019*. Technical notes.

Chapter 7

Substitutability, non-substitutability and 'balance' of indicators*

7.1 Introduction

A fundamental issue concerning composite indicator construction is the degree of compensability or substitutability of the individual indicators.

The components of a composite indicator are called 'substitutable' if a deficit in one component may be compensated by a surplus in another (e.g., in a time use index, a low value of "Proportion of people who have participated in religious or spiritual activities" can be offset by a high value of "Proportion of people who have participated in meetings of cultural or recreational associations" and vice versa). Similarly, the components of a composite indicator are called 'non-substitutable' if a deficit in one component may not be compensated by a surplus in another (e.g., in a development index, a low value of "Life expectancy at birth" cannot be offset by a high value of "GDP per capita" and vice versa).

Therefore, we can define an aggregation approach as 'compensatory' or 'non-compensatory' depending on whether it permits compensability or not (Casadio Tarabusi and Guarini, 2013). An in-between approach based on an 'imperfect substitutability' across all components of a composite indicator is called 'partially compensatory'.

Compensability is closely associated with the concept of unbalance, i.e., a disequilibrium among the individual indicators that are used to build the composite indicator. For example, in the case of two individual indicators X and Y whose normalized values range between 0 and 1, there is perfect balance when X = Y, whereas the maximum unbalance occurs when X = 0 and Y = 1 or vice versa. In any composite indicator each dimension is introduced to represent a relevant aspect of the phenomenon considered, therefore a measure of unbalance among

* This chapter is mostly written by Matteo Mazziotta and Adriano Pareto.

dimensions may help the overall understanding of the phenomenon. In a non-compensatory or partially compensatory approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is often used.

7.2 The concept of 'balance' or 'equilibrium'

As is well known, one of the most important steps for constructing a composite indicator is the normalization of individual indicators. Normalization aims to make the indicators comparable and it is required before any data aggregation as the indicators in a data set often have different measurement units. Thus, it is necessary to bring the indicators to the same standard by transforming them into pure, dimensionless numbers (Mazziotta and Pareto, 2017).

The literature offers a wide variety of normalization methods, each with its pros and cons. The most common are: Standardization (or transformation in *z*-scores), Rescaling (or Min-Max method) and Indicization[9] (or transformation into index numbers).

However, the normalization method has a strong impact on results because it creates a 'correspondence system' (also denoted as a 'correspondence grid') between different indicators (McGranahan, 1970). The 'correspondence system' defines what level of any one indicator tends to go with (corresponds to) given levels of other indicators (e.g., what level of "Life expectancy" should be found normally with a given level of "Gross national income per capita" and vice versa). These correspondences are particularly important when a non-compensatory approach (i.e., an approach based on the concept of 'unbalance' or disequilibrium among individual indicators) is followed[10].

In such a case, in fact, it is necessary to define what is meant by 'balance' and this definition depends on the normalization method adopted. For example, if indicators are converted to a common scale with a range of [0, 1], then the set of maximum values and the set of

[9] This method is also called 'Distance to a reference' (OECD, 2008).

[10] We say that an approach is non-compensatory when it is not fully compensatory.

minimum values will be considered 'balanced'[11], whereas the set of mean values could be considered 'unbalanced'. By contrast, if indicators are converted to a common scale where the mean value is set equal to 100, then the set of mean values will be considered 'balanced', whereas the set of maximum values and the set of minimum values could be considered 'unbalanced'.

Therefore, an incorrect choice of normalization method for constructing a non-compensatory composite indicator can lead to an unacceptably large degree of distortion of results.

7.3 The 'effect' of normalization

In order to illustrate the effect of normalization on the distributions of individual indicators, we consider the three indicators represented in Figure 7.1.a (Mazziotta and Pareto, 2020).

The first indicator has an exponential distribution (Exp), the second has a normal distribution (Nor) and the third has a Beta distribution (Bet). The indicators have different means and variances, as we suppose that they represent the most disparate phenomena. Figures 7.1.b, 7.1.c and 7.1.d show, respectively, the distributions of normalized indicators by standardization, rescaling and indicization.

As we can see, the distributions of indicators transformed into z-scores (Figure 1.b) are 'centred' around the origin (mean=0) and 'elongated' or 'shortened' to have the same variability (variance=1).

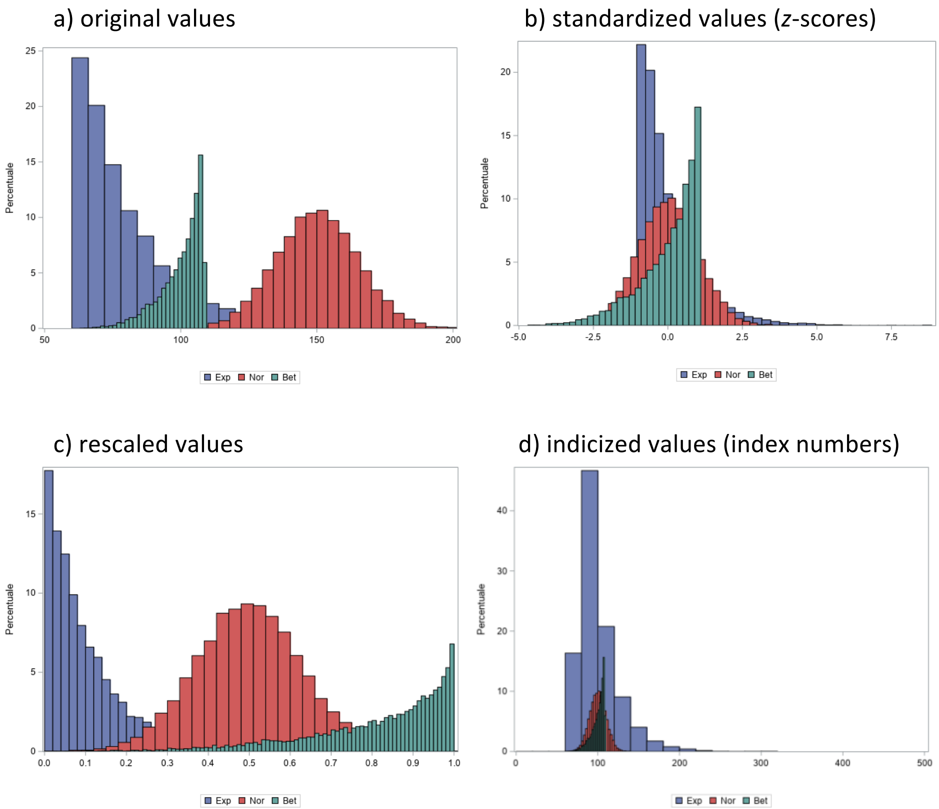
Rescaling also makes the variances more homogeneous (but not equal), bringing all the values into a common interval (Figure 7.1.c). However, the distributions of indicators are not 'centred' and this leads to the loss of a common reference value, such as the mean. It follows that equal normalized values (i.e., balanced normalized values) can correspond to very unbalanced original values. For example, the normalized value 0.2 for the Exp indicator corresponds to a high original value, whereas for the Nor and Bet indicators it corresponds to a very low original value. Therefore, the use of a simple rescaling for

[11] Note that this is a strong and less plausible assumption, because the minimum and the maximum of a distribution are often 'outliers' (i.e., 'abnormal' values).

aggregating individual indicators with an unbalance adjustment method, such as the geometric mean, can lead to biased results. Moreover, the normalized value 0.5 is the mean of the range, but not of distributions, and thus it cannot be used as a reference for reading results (for instance, if the normalized value of a given unit is 0.3, we cannot know if its original value is above or below the mean).

Indicization with the mean as a base set to 100 (Figure 7.1.d) ‘centres’ all distributions around the mean but does not ‘normalize’ their variability (e.g., the range of the Bet indicator is very short, whereas the range of the Exp indicator is very large). Indeed, indicized indicators have the same coefficients of variation as original indicators.

Figure 7.1
Original and normalized indicators with different distributions



7.4 The 'correspondence grid'

A correspondence grid emerges when for each indicator the original values that are identified with each level of the common scale are shown in a table. For example, at level 2 of the correspondence grid of standardization are given the original values (for each indicator) that have the standardized value of 2 and that correspond to each other. The result is a list of 'correspondence points', each of which represents a set of original values that will be considered 'balanced' (Mazziotta and Pareto, 2021).

The correspondence grid must be carefully constructed and evaluated by the researcher, because it can yield an 'artificial' or 'inconsistent' model of balance of original indicators. Table 7.1 shows the correspondence grids for the three normalizations of Figure 7.1.

There are a number of points of interest in the table. In particular, all normalization methods, except rescaling, consider 'balanced' the set of mean values. Standardization considers balanced a set of values when they are 'equidistant' from the mean in terms of standard deviations. For example, at level 3 of the correspondence grid all the original values are given that are equal to the mean plus 3 standard deviations.

Indicization considers balanced a set of values when they are 'equidistant' from the mean (the base) in percentage terms. For example, at level 200 of the correspondence grid all the original values are given that are double the mean. In this case, the set of null values is also considered balanced, so the transformation into index numbers should be applied only to variables that have an 'absolute zero' point (e.g., "Height" and "Weight").

Finally, rescaling considers balanced the two sets of extreme values and it creates 'artificial' correspondence points (i.e., artificially balanced sets of values) in the middle. For example, the set of mean values corresponds approximately to the set of normalized values (0.1, 0.5, 0.8), and then it will be considered very unbalanced. The greater the differences between the indicator distributions, the greater the distortion of the correspondence points.

Table 7.1 – *Correspondence grid for different normalization methods*

Standardization				Rescaling				Indicization			
Scale	**Exp**	**Nor**	**Bet**	**Scale**	**Exp**	**Nor**	**Bet**	**Scale**	**Exp**	**Nor**	**Bet**
2.5	128.5	187.6	119.4	1.0	250.0	209.5	108.4	200	160.0	300.0	200.0
2.0	118.8	180.1	115.5	0.9	231.0	197.7	103.9	180	144.0	270.0	180.0
1.5	109.1	172.6	111.6	0.8	212.1	185.8	**99.4**	160	128.0	240.0	160.0
1.0	99.4	165.0	107.8	0.7	193.1	174.0	94.9	140	112.0	210.0	140.0
0.5	89.7	157.5	103.9	0.6	174.2	162.2	90.5	120	96.0	180.0	120.0
0.0	**80.0**	**150.0**	**100.0**	0.5	155.2	**150.3**	86.0	**100**	**80.0**	**150.0**	**100.0**
-0.5	70.3	142.5	96.1	0.4	136.3	138.5	81.5	80	64.0	120.0	80.0
-1.0	60.6	135.0	92.2	0.3	117.3	126.7	77.0	60	48.0	90.0	60.0
-1.5	50.9	127.4	88.4	0.2	98.4	114.8	72.5	40	32.0	60.0	40.0
-2.0	41.2	119.9	84.5	0.1	**79.4**	103.0	68.1	20	16.0	30.0	20.0
-2.5	31.5	112.4	80.6	0.0	60.5	91.2	63.6	0	0.0	0.0	0.0

7.5 Conclusions

In a non-compensatory or partially compensatory approach, all the dimensions of the phenomenon must be balanced and an aggregation function that takes unbalance into account, in terms of penalization, is often used. This methodological approach requires what is meant by 'balance' and this definition depends on the normalization method adopted.

It is good to specify that, in general, the perfect normalization method does not exist (Freudenberg, 2003). Each method has strengths and weaknesses, and the choice depends on the aims of the research and/or on the aggregation function used for constructing the composite indicator. In any case, a realistic 'correspondence grid' (not artificial or meaningless) must be constructed in order to consider a correct 'balancing model' of the values.

For this reason, many composite indices based on the classical Min-Max method should be revised. This is the case with the Human Development Index (HDI) (UNDP, 2019), where, however, the goalposts[12] are 'reasoned' (i.e., they act as 'natural zeros' and 'aspirational targets') and have been set by experts. In other cases, as in the composite indices summarizing the SDGs published recently by Istat (Istat, 2020), the normalization is based on a simple rescaling with 'observed' goalposts. This procedure can lead to distortions of the 'balancing model' of individual indicators and, therefore, to incorrect or misleading results. On the other hand, the composite indices developed by ASVIS to describe Italy's performance with respect to the SDGs are normalized with a revised Min-Max method (AMPI methodology) based on a real 'balancing model' corresponding to the set of values in 2010 (ASVIS, 2020).

In conclusion, the construction of a composite indicator must follow a precise work paradigm and the 'balancing model' must be carefully defined, since composite indicators have a great responsibility: measuring multidimensional phenomena in order to better understand the reality.

[12] The goalposts are the minimum and maximum values used for transforming original indicators expressed in different units into indicators normalized between 0 and 1 (UNDP, 2019).

References

ASVIS (2020). L'Italia e gli Obiettivi di Sviluppo Sostenibile. Rapporto ASVIS 2020. Roma: Alleanza Italiana per lo Sviluppo Sostenibile.

Casadio Tarabusi, E., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research*, *112*, 19-45.

Chiappero-Martinetti, E., & von Jacobi, N. (2012). Light and shade of multidimensional indexes. How methodological choices impact on empirical results. In F. Maggino, & G. Nuvolati (Eds.), *Quality of life in Italy, research and reflections* (pp. 69-103). Social Indicators Research Series, 48. Cham: Springer.

Freudenberg, M. (2003). *Composite indicators of country performance: a critical assessment*. Paris: OECD, DSTI/IND, 5.

Istat (2020). *Rapporto SDGs 2020*. Roma: Istituto Nazionale di Statistica.

McGranahan, D. (1970). The interrelations between social and economic development, *Social Science Information*, *9*, 61-77.

Mazziotta, M., & Pareto, A. (2017). Synthesis of indicators: The composite indicators approach. In F. Maggino (Ed.), *Complexity in society: from indicators construction to their synthesis* (pp. 159-191). New York: Springer.

Mazziotta, M., & Pareto, A. (2020). *Gli indici sintetici*. Torino: Giappichelli.

Mazziotta, M., & Pareto, A. (2021). Everything you always wanted to know about normalization (but were afraid to ask). *Italian Review of Economics, Demography and Statistics*, LXXV, 1, 41-52.

OECD (2008). *Handbook on constructing composite indicators. Methodology and user guide*. Paris: OECD Publications.

UNDP (2019). *Human Development Report 2019*. New York: United Nations Development Programme.

Zhou, P., Ang, B.W., & Zhou, D.Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research*, *96*, 169-181.

Chapter 8

Inequality and its role in measuring well-being*

8.1. Introduction

Inequality, defined as "the unfair situation in society when some people have more opportunities, money, etc. than other people" (CUP, 2008), has been traditionally associated with income inequality, a situation in which different individuals or groups have more material resources than others.

Economic inequality is not synonymous with, and has a broader scope than, income inequality.

However, a good part of the economic inequality has been explained in terms of individual or social circumstances that either helped an individual to achieve, or prevented them from achieving, the best outcomes (Sen, 1999), most of which are ex-ante conditions that define the lack of an even, and hence equitable, starting place (Alfonso et al., 2015).

A second aspect to highlight is that the issue of inequality can be addressed as evidence to be incorporated into the composite indicator or as a matter to be evaluated/assessed in its own right. Within multidimensional well-being a further distinction is between the assessment of vertical inequality, i.e., inequality in the distribution of each elementary indicator ("individual achievement", as it is usually called in a well-being context) and concern about an association across achievements pertaining to the same individual (attributes). In the former case we obtain composite indicators that account for overall inequality (or overall inequality measures, if we are addressing inequality in itself) aggregating inequality over each of the univariate distributions (attributes). This approach forms the basis of the Inequality-adjusted Human Development Index (IHDI). If we invert the order of aggregation, we derive an overall measure of inequality that aggregates

* This chapter is mostly written by Elena Grimaccia, Adrian Otoiu and Silvia Terzi.

synthetic functions of the indicators across individuals. The latter approach embeds the association between the achievements in the various dimensions into an overall indicator of individual achievements (Aaberge and Brandolini, 2014). The first form of inequality is distribution-sensitive inequality and the second is association-sensitive inequality. Many indices of social welfare, inequality or poverty, such as the Human Development Index and Human Poverty Index, are insensitive to either of these forms of inequality (Seth, 2013).

This state of unevenness has been quantified by well-known measures of inequality, among which the best known are the Gini index and the Lorenz curve, which help depict the extent to which the distribution of a resource (e.g., income, wealth) differs from a perfectly even distribution, e.g., where the share of wealth of a certain group matches its share in the population.

However, inequality in fact has to do with differences in economic well-being between population groups and unequal distribution of resources or opportunities, as well as wealth or income, and it has taken a long time before any measures of this broader notion of inequality were introduced.

For the longest time, the measurement and assessment of inequality has focused on income and/or wealth inequality. Inequality in these measures is perceived as the original cause of most other instances of inequality (Piketty, 2013, cited by De Muro, 2016), where material deprivation leads to decreased access to health, lower educational attainment, worse living conditions, etc., and is often associated with unemployment and a lower social status (Layard et al., 2005).

Since knowledge has progressed, the perception and assessment of inequality has acknowledged the fact that inequality is in fact a multidimensional concept and contrary to the traditional view, cannot be limited to income and/or wealth inequality. Efforts have been made to distinguish income/wealth inequality from other forms of inequality, in relation to the view that inequality of outcomes in health, education, etc., cannot be fully explained by inequalities in income or consumption alone (Alfonso et al., 2015). Moreover, it has been pointed out that, in fact, material wealth is not an end purpose in itself, or in other words, not a sole objective in achieving a higher overall well-being (Otoiu et al., 2014). De Muro (2016) also points out that.

8.2 Vertical inequality: meaning and applications

While there are approaches that have expanded the concept of inequality as stemming from the initial conditions and specific circumstances faced by individuals, vertical inequality has been mostly shaped as an explanation of the inequality in outcomes, or in other words, the dissimilarity in how a certain characteristic is spread across individuals, groups or countries.

Several measures have emerged in quantifying these dissimilarities, by looking at the distribution of a certain characteristic across the entities of interest; among the best-known measures is Atkinson's (1970) family of inequality measures. Atkinson's inequality index A is defined for each distribution x by:

$$A(\varepsilon)= 1 - [M_{(1-)}/M_1]\ \forall\varepsilon>0 \qquad (1)$$

An Atkinson measure compares a 'bottom-sensitive' general mean M_q (with $q = 1\text{-}\varepsilon<1$) and the 'neutral' arithmetic mean (with $q=1$) (Foster et al., 2005). General means of the order $q <1$ are always smaller than the arithmetic mean, and so the ratio $[M_{(1-\varepsilon)}/M_1]$ is less than 1 but greater than 0. Greater inequality is reflected in a larger relative gap between $M_{(1-\varepsilon)}$ and M_1 and hence a higher value for the inequality measure. The practical intuition behind the use of the two means is that the general mean $Mq,\ q<1$, is a well-known measure used to take into account the relative distribution of the individual values, which gauges uneven concentrations of data (e.g., a higher incidence of smaller values) and also accounts to some extent for their variability, as opposed to the arithmetic mean. In the formula, $Mq/M1 = 1$ reflects the (reference) state of perfect equality, hence the final result reflects the extent of the deviation from this situation.

Greater inequality is reflected in a larger relative gap between $M_{1-\varepsilon}$ and M_1, and hence a higher value for the measure. Since $M_{(1-\varepsilon)}$ is a decreasing function of ε, the index $A(\varepsilon) = 1 - [M_{(1-\varepsilon)}/M_1]$, is an increasing function of ε, so that the parameter ε can be interpreted as an inequality awareness parameter (or *inequality aversion parameter*, as Foster defines it).

In acknowledging the need for inequality to be multidimensional, several composite indicators have incorporated inequality in their methodology/construction.

The most frequently used procedure to incorporate inequality into an elementary indicator is to use an inequality index to penalize it. Let X be a welfare variable, somehow affected by inequality. Let $\bar{X}$ be its mean value and IX an inequality index. We can define an inequality-adjusted index X as:

$$IAX = \bar{X}(1\text{-}I_X) \tag{2}$$

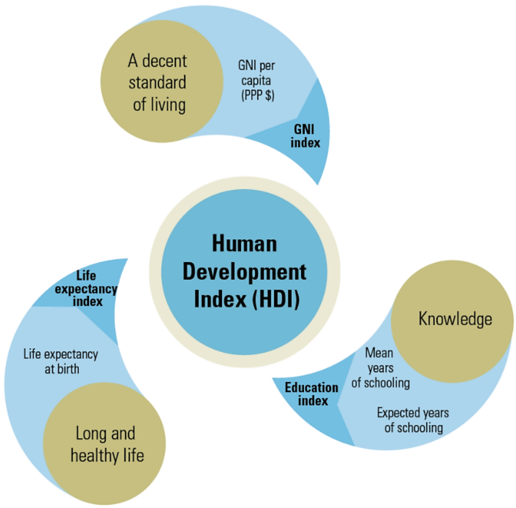
The best-known and most authoritative index, the Human Development Index, developed by the UNDP and available since 1990, has been computing the inequality-adjusted human development index since 2010, by adjusting for inequality each of the three pillars of the main HDI index, i.e., Life expectancy index, Education index and Gross National Income per capita index (UNDP, 2019). This follows several previous attempts to incorporate inequality in the measurement of well-being in the HDI.

The first attempt dates from 1991 when the Income Distribution-Adjusted HDI (IDAHDI) was computed by adjusting the original HDI values by the Income Gini Index (GI). This approach, although incomplete, had been used in order to adjust for the most uneven dimension of well-being, the income dimension (Kovacevic, 2010), and, given the correlation between the GI and the three dimensions of well-being (see Figure 8.1), it was deemed to account for most of the inequality (Kovacevic, 2010). Incorporation of the GI in the IDAHDI was done by multiplying the income pillar by (1-GI), and the GI was interpreted as the loss of welfare as a percentage of the maximum achievable welfare, caused by income inequality (Kovacevic, 2010).

Despite its value, the IDAHDI was computed for a limited number of countries, and discontinued in the 1994, among concerns about the availability of income inequality measures (Kovacevic, 2010).

However, the interest behind accounting for inequality remained strong, and several options have been explored to account for inequality in human development.

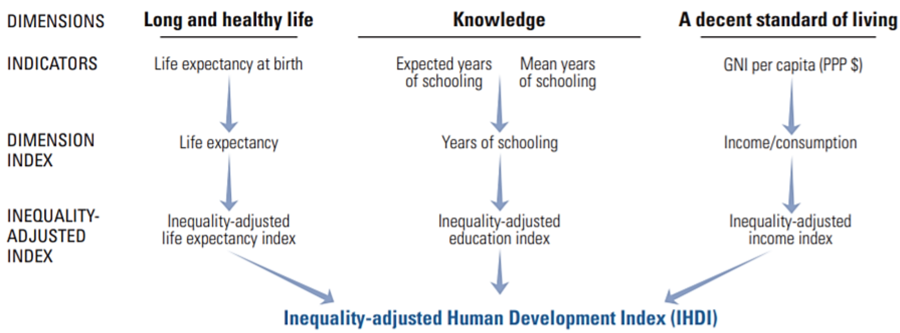
Figure 8.1
The three pillars of the Human Development Index (UNDP, 2020)



Among them, there is Hicks' Inequality-Adjusted HDI (IAHDI) (Hicks, 1997) and the more advanced version of Foster et al.'s method (2005), which uses and extends Atkinson's class of welfare functions to a multidimensional measure of inequality (Kovacevic, 2010). A (further) revision of Foster et al.'s (2005) method was proposed by Alkire and Foster (2010) and has been used for computing the Inequality-adjusted HDI (IHDI) since 2010.

Essentially, the Alkire and Foster methodology consists of adjusting for inequality each pillar of the (unadjusted) HDI using the methodological layout presented in Figure 8.2.

Figure 8.2
The structure of the IHDI (UNDP, 2019)



The adjustments take into consideration inequality within each of the three pillars, Life expectancy, Education and Income, using granular data available for each dimension, as follows:
– for life expectancy, inequality is computed using the UNDESA abridged life tables with mortality and average age at death provided in five-year intervals up to 100 (UNDP, 2019).
– mean years of schooling for harmonized household data (a complete list is included in the UNDP (2019) technical notes).
– disposable household income or consumption per capita for most countries, using similar data sources to those used for computing mean years of schooling.

Calculation of the inequality-adjusted index is done using the following methodology:
1) calculation of Atkinson indexes for each dimension using formula (1) with $\varepsilon = 1$. To account for the fact that no zeros or negative values can be used to compute a geometric mean, some adjustments are made to compensate for these cases.
2) adjustment of each subindex (pillar) with the value of the corresponding Atkinson indexes (A_x), (with $\varepsilon = 1$), using the formula:

$$I'_x = (1 - A_x) \cdot I_x \tag{3}$$

where I_x is the original HDI index for one of the three dimensions, and I'_x the corresponding inequality-adjusted index.
3) computation of the IHDI index in a similar way to that used for the HDI, as a geometric mean of the inequality-adjusted subindexes.
In addition, HDI methodology computes the overall loss of HDI due to inequality using the geometric mean formula:

$$Loss = \sqrt[3]{\prod_{x=1}^{3}(1 - A_x)} \tag{4}$$

along with the coefficient of human inequality, computed as an arithmetic average of the three Atkinson indexes A_x.

Despite the technical complexities, the IHDI index has managed to achieve the calculation of index values for 150 out of 189 countries covered by the regular HDI, which ensures that subgroup consistency is achieved. This means that changes for subgroups are captured and alter the distribution of inequality for the entire group (or country, in the case of (I)HDI).

The main disadvantage of using this method is that the IHDI is not association-sensitive in the sense that it does not capture inequalities that overlap across its dimensions (UNDP, 2019). This would be currently impossible in the absence of having a single data source for all countries (UNDP, 2019).

Despite some of its disadvantages, the IHDI measure has managed to bring additional insights in measuring human development. Thus, it makes sense to consider that, in the light of Sen's work, and in line with the HDI theoretical goals, which focus on individual well-being, the opportunity to live the life desired and to have the proper environment that allows an individual to fulfil their potential (UNDP, 2020), the IHDI quantifies the achieved well-being as opposed to potential well-being for a given country in a given year, as computed by the HDI.

Another criticism of this aggregation method is that within-group inequality is not accounted for, so a strong association-sensitive condition is not satisfied. However, given the different data sources on which the HDI is based across dimensions and, in many cases, across countries, this condition cannot be satisfied for the IHDI in a feasible way.

Other authors have suggested and applied different inequality indicators to adjust for vertical inequality, but the prominent features of these alternative inequality-adjusted indicators are closely related to those just illustrated. For example, Ciommi et al. (2017) propose a composite index computed as a weighted average of the elementary indicators with weights based on the Gini index of concentration (Gini-based weighted average (GW)), and this is tantamount to using the Gini index instead of Atkinson's index of inequality.

Of course, the Gini-based weighted average is not association sensitive either. Association sensitivity is an important concept, which, in the context of composite indicators, captures both the vertical distribution of inequality for individual dimensions and the correlation between them (Kovacevic, 2010) in order to provide a fair estimation of multidimensional inequality. This, to some extent, addresses the issue of non-substitutability as a desirable property of composite indicators, set forth in the first chapter of this book, as poor performance in one dimension should not automatically be compensated by an above-average performance in another (Mazziotta and Pareto, 2016).

Another way to address inequality, this time straightforwardly, is used in the World Happiness Index. The Index includes data for 60 countries, and it derives from the Gross National Happiness index (Bhinde, 2017). It comprises four chapters: Peace and security; Freedom, democracy, human rights; Quality of life; Research, education, information, communication, culture. The quality of life chapter also includes an inequality component, namely the Gini coefficient, summed up together with GDP per capita, life expectancy, the incidence of suicides and air quality. The component score is the sum of the ranks of the indicators:

$$QL = GDP\ rank + GINI\ rank + LIFE\ rank + SUIC\ rank + AIR\ rank \qquad (5)$$

8.3 Dispersion or inequality (across individuals): importance and issues

The notion of inequality across dimensions is closely related to the notion of unbalance or disequilibrium among indicators. So, we can say that the inequality issue is indirectly addressed within non-compensatory or partially compensatory aggregating functions (as in the preceding chapter). However, the notion of inequality across dimensions is also connected with the association among the components of a composite indicator and with the concept of horizontal dispersion.

When addressing the issue of dispersion in a univariate context (defined with respect to a location parameter or mean value), this is introduced as a measure of the reliability of the mean value. In fact, knowledge of the central tendency is accompanied by the need to describe how well the mean value represents the values assumed by the units of the observed population. In other words, the aim is to understand how close or far the modalities observed in the population are from the “centre” of the distribution.

If we extend/apply this concept to composite indicators, the lower the horizontal dispersion/variability, the better the reliability/ representativeness of the individual well-being achievement (i.e., CI’s individual score). And with a further step in this same direction, we can state that whenever the horizontal variability is moderate for all units, the CI is a good representation of the multidimensional phenomenon.

Horizontal dispersion can be a useful reference and tool in the context of multivariate association (concordance) among the single components of the CI, but even more so when addressing issues concerning inequality across dimensions.

For many multidimensional phenomena there is not only a question of how to define the best composite or indicator but also how to take into account the joint distribution of the single components. For example, among the issues in the measurement of well-being of the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz et al., 2009) we find the recommendation to give more prominence to the joint distribution of the dimensions of people's well-being.

But also, carrying on the recommendations with more and more in-depth details, we also read: *quality-of-life indicators in all the dimensions covered should assess inequalities in a comprehensive way, taking into account linkages and correlations. […] The consequences for quality of life of having multiple disadvantages far exceed the sum of their individual effects. Developing measures of cumulative effects requires information on the "joint distribution" of the most salient features of quality of life across everyone in a country.*

As already stressed, the inequality issue can be indirectly addressed within non-compensatory or partially compensatory aggregating functions. Of course, compensability/non-compensability does not imply dependence/independence and vice versa, however these two topics share some common aspects.

The common assumption of the two issues is that a fully compensatory approach cannot be taken for well-being indicators or for other composite indicators referring to social indicators. The different indicators and/or dimensions are not interchangeable, unless the theoretical framework suggests the specification of a reflective measurement model. (Remember that for the latter model the removal of one of the indicators does not change the essential nature of the underlying concept and that all elementary indicators must be correlated.)

An equivalent way of taking into account the Stiglitz Commission's recommendation is to embed in the composite indicator some information on the joint distribution of single components, as for example, for ordinal variables, in the case of the Alkire-Foster dual cut-off method or the Mazziotta-Pareto Index (MPI) (Mazziotta and Pareto, 2016), or to correct a composite indicator to account for joint distribution as in the adjusted Human Development Indicator (HDI) (Terzi, 2013). Once again, we are faced with different paths, and none of them are without limitations.

As already underlined, the counting approach combines different elements of achievement at the individual level, which are then summed over individuals to form an aggregate index. It represents the simplest way to embed the association between achievements or, vice-versa, deprivations, at the individual level into an overall index. Thus, within well-being indicators we could resort to the Alkire-Foster dual cut-off suggestion to build a multidimensional achievement index or a multidimensional deprivation index. The method is geared for ordinal variables, so we need to replace first-order cut-offs with relevant percentiles. An example of a well-known dual cut-off-based method is the Multidimensional Poverty Index <http://hdr.undp.org/en/content/multidimensional-poverty-index-mpi>.

In the case of Mazziotta and Pareto's suggestion (Mazziotta and Pareto, 2016), that has led to the Mazziotta-Pareto Index (MPI); given the availability of the necessary data, it is possible to discount for horizontal variability in a manner similar to that illustrated for vertical inequality.

First of all, individual indicators are converted to a common scale with a mean of 100 and standard deviation of 10. Now, turning attention to the rows of the normalized data matrix, let M_{zi}, S_{zi} and CV_i denote, respectively, the mean value, the standard deviation and the coefficient of variation for unit *i*.

$$MPI_i = M_{zi} - S_{zi} \cdot CV_i \tag{6}$$

Therefore, the MPI decomposes the score of each unit into two parts: a mean level (M_{zi}) and a penalty ($S_{zi} \cdot CV_i$). The penalty is a function of the indicators' variability in relation to the mean value ('horizontal variability') and it is used to penalize the units. The aim is to reward the units that have a greater balance among the indicators' values. It is then possible to correct the value of the composite indicator taking into account its vertical variability to derive an inequality adjusted MPI.

The third path leads to setting a correction to a composite well-being index, to account for the joint distribution of achievements. The original suggestion (Terzi, 2013) led to the correction of the Human Development Index (HDI) by means of the Multidimensional Poverty Index. The idea this suggestion stems from is that of integrating the two different perspectives: horizontal or micro-level aggregation (households) and vertical or macro-level (regional) aggregation. Given an association-sensitive indicator P (such as Multidimensional Poverty) and given a well-being indicator *WB*, the corrected version of the well-being indicator is:

$$WBc = WB\ (1\text{-}P) \tag{7}$$

Of course, this same correction could be applied to the inequality-adjusted HDI or to other vertical inequality-adjusted well-being indicators.

It is otherwise possible to address straightforwardly the two issues, with exploratory intent. As far as the inequality issue is concerned, we could compute horizontal standard deviation for each unit as in the MPI or other measures of dispersion.

As for the association among components, in a recent paper, Terzi and Moroni (2020) suggest a counting-based approach to plot the association between the different dimensions of a multivariate phenomenon such as social vulnerability. The type of association to detect is comonotonic/positive association, i.e., concordance. Taking moves from Kendall's concept of concordance/agreement and from his well-known concordance coefficient *W* (Kendall and Babington Smith, 1939), they introduce the notion of local concordance and suggest a local concordance coefficient and a local concordance curve in order to detect different degrees of concordance in the head, tail or centre of the multivariate distribution of the components of a well-being indicator.

Let X be the data matrix of n units and d variables: $X = (x_{ih})$, $i = 1,\dots,n$ and $h = 1,\dots,d$. For each column x_h of the matrix, they order (non-increasingly) the observations and divide them into slices of a fixed size s, where s is a function of n and d. They then count how many times a unit ranked in the *r*-th slice of any of the *d* distributions also belongs to the *r*-th slice of any of the others.

The smaller the number of units ranked in the union of the *r*-th slices, the greater the local concordance between the *d* distributions. In fact, in the case of maximum local concordance, the s units that achieve the *s* ranks pertaining to a certain slice of one distribution also achieve the same set of ranks in all the others.

This is the simple and intuitive idea that the local concordance coefficient stems from. By computing the local concordance coefficient for all the distinct slices of size *s* of the multivariate distribution, a local concordance curve can be derived.

Of course, local concordance and horizontal inequality are strictly related concepts. Maximum local concordance means no horizontal dispersion among ranked observations: more or less balanced indicators and horizontal equality.

Both these measures are local measures that can be easily aggregated or summarized to obtain an overall measure of concordance (like Kendall's W) or of horizontal inequality. In turn, this overall measure could be used to correct the composite well-being indicator as in (7).

The approaches described above can be used to address the issue of association/complementarity in the context of multidimensional well-being in several proof-of-concept applications.

One of the most important attempts to fit the dispersion of inequality into a composite index was carried out by Foster et al. (2005). In their paper, Foster et al. (2005) pointed out the limitation of the mean of the dimensional means approach in incorporating inequality in a composite indicator of well-being and made the case for an inequality adjusted HDI that accounts for the horizontal dispersion across dimensions at unit level (state level in this case). The formula used by Foster et al. (2005) is:

$$\mathrm{H}_{\varepsilon}(\mathrm{D}) = \mathrm{H}(\mathrm{D})(1 - \mathrm{A}_{\varepsilon}) \tag{8}$$

where *D* stands for a matrix of dimensions at unit level, $\mathrm{H}_{\varepsilon}(\mathrm{D})$ is the inequality-adjusted composite indicator, $\mathrm{H}(\mathrm{D})$ the original, unadjusted indicator and A_{ε} an Atkinson-type horizontal inequality measure computed at unit level, different from similar measures computed at dimension level (e.g., the IHDI).

Achievements, or in other words, the performance recorded for each unit and for each dimension, are treated as complements rather than substitutes, rising with a higher ε (remember that ε is the inequality aversion parameter).

An analysis for Mexico shows an application of this method using 2000 census data for computing the HDI-Generalized Mean (HDI-GM), an index similar to the HDI, albeit with different data, for the 32 Mexican states, using actual and imputed data at individual level in three dimensions: income, education and health. A comparison made between an index with $\varepsilon=0$, meaning that no inequality is factored in, and an index with $\varepsilon=3$, meaning extreme aversion to inequality, shows that the HDI performance of states changes markedly, favouring states with smaller discrepancies between individual performances for the three dimensions. Also, it shows that education is the highest source of inequality, followed by income, both effecting a loss of HDI of above 13%. The index has proven itself to satisfy the condition of subgroup consistency, where changes in human development for a particular group in society are reflected in its overall, composite indicator measure.

Another proof-of-concept application was carried out by Seth (2013) using Indonesian data from 1997 and 2000. This index, which is similar to the HDI in terms of construction and dimensions, seeks to factor in the distributional and associative sensitivity of multidimensional welfare indices by considering both forms of inequality: in the dispersion of inequality for each dimension across population, and the association/ correlation across different dimensions (Seth, 2013).

The analysis benefits from making comparisons between the two years as an assessment of the presumed loss of human development effected by the Asian crisis, as revealed by the index results in the year preceding it (1997) and after it (2000). In order to ensure comparisons between the two periods and allowing for the use of both methods of two-stage aggregation (row-first and column-first), normalization is carried out first, using dimension-specific threshold measures that define deprivation.

A comparison between an index similar to the HDI, computed as a mean for each dimension across individuals, followed by a mean of dimensional means, (Seth, 2013), and an index that first computes the means across dimensions, shows the importance of factoring in inequality in the construction of a composite indicator. In Seth (2013) they are concisely formalized as $W(\alpha,\beta)$, with α being an inequality aversion parameter, and β being the dimension substitutability parameter.

The first index, which is similar to the HDI, which is indifferent to inequality, assumes a $\varepsilon=0$ (expressed as $\alpha=1$) and has perfect substitutability among dimensions (expressed as $\beta=1$) shows a generalized decrease in welfare across all regions. The second one, with $\alpha=-1$ (equivalent to a relatively high inequality aversion of $\varepsilon=2$), and a $\beta=0.1$, which stands for almost no substitutability between the dimensions, shows that welfare has in fact increased from 1997 and 2000, as disparities between individuals have, to a certain extent, decreased.

Other important conclusions have emerged from this analysis. It appears that the fall in real per-capita income has led to reduced inequality (Seth, 2013), a conclusion compatible with Foster et al. (2005), and that an increase in years of education has helped reduce inequality (Seth, 2013). And, even though pairwise correlations among dimensions have increased from 1997 and 2000, leading to an increase in (overall) inequality, the reduction of distributional inequality has more than offset it, leading to a sizeable decrease in inequality and, therefore, to increased overall welfare.

References

Aaberge, R., & Brandolini, A. (2014). *Multidimensional poverty and inequality*, No 976, Temi di discussione (Economic working papers), Bank of Italy, Economic Research and International Relations Area.

Afonso, H., LaFleur, M., & Alarcón, D. (2015). *Concepts of inequality. Development Issues* No. 1, Development Policy and Analysis Division of UN/DESA, retrieved from <https://www.un.org/development/desa/dpad/publication/no-1-concepts-of-inequality/>, on November 27th, 2020.

Alkire, S., & Foster, J. (2010). Designing the Inequality-Adjusted Human Development Index (IHDI). 2010 OPHI Working Paper No. 37, Oxford Poverty & Human Development Initiative (OPHI), University of Oxford.

Atkinson, A.B. (1970). On the measurement of inequality. *Journal of Economic Theory, 2* (3), pp. 244-26.

Bhinde, H.N. (2017). An outline of 'Happiness Index. *CnR's International Journal of Social & Scientific Research, 03* (I), 31-34.

Ciommi, M., Emili, A., Gigliarano, C., Taralli, S., & Chelli, F.M. (2017). A new class of composite indicators for measuring well-being at the local level: An application to the equitable and sustainable well-being (BES) of the Italian Provinces. *Ecological Indicators*,*76*, 281–296.

CUP (2008). *Cambridge advanced learner's dictionary* (3rd edn.). Cambridge, UK: Cambridge University Press.

De Muro, P. (2016). Not just slicing the pie: The need for a broader approach to economic inequality. In S. Fadda, & P. Tridico (Eds.), *Varieties of economic inequality* (Volume 29, pp. 11-21), London and New York: Routledge Advances in Heterodox Economics.

Foster, J.E., Lopez-Calva, L., & Szekely, M. (2005). Measuring the distribution of human development: Methodology and application to Mexico. *Journal of Human Development*, *6*, 5-29.

Hicks, D. (1997). The inequality adjusted Human Development Index: A constructive proposal. *World Development*, *25*, 1283-1298.

Istat 2020. BES 2019. *Equitable and sustainable well-being in Italy*. Istituto nazionale di statistica, Rome IT.

Kendall, M.G., & Babington Smith, B. (1939). The problem of *m* rankings. *The Annals of Mathematical Statistics*, *10*, 275-287.

Kovacevic, M. (2010). *Measurement of inequality in human development: A review*. Human Development Research Paper. UNDP-HDRO, New York <http://hdr.undp.org/sites/default/files/hdrp_2010_35.pdf>

Layard, R., Nickell, S., & Jackman, R. (2005). *Unemployment: Macroeconomic performance and the labour market*. Oxford: Oxford University Press.

Mazziotta, M., & Pareto, A. (2016). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social Indicator Research 127*(3), 983-1003.

Otoiu, A., Titan, E., & Dumitrescu, R. (2014). Are the variables used in building composite indicators of well-being relevant? Validating composite indexes of well-being. *Ecological Indicators*, *46*, 575-585.

Sen, A.K. (1999). *Development as freedom*. New York: Anchor Books.

Seth, S. (2013). A class of distribution and association sensitive multidimensional welfare indices. *Journal of Economic Inequality*, *11*, 133-162.

Stiglitz, J.E., Sen, A., & Fitoussi, J.P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Commission on the Measurement of Economic Performance and Social Progress, Paris.

Terzi, S. (2013). How to integrate macro and micro perspectives: An example on human development and multidimensional poverty. *Social Indicators Research*, *114*(3), 935-945.

Terzi, S., & Moroni, L. (2020). Local concordance and some applications. *Social Indicators Research,* <https://doi.org/10.1007/s11205-020-02312-z>.

UNDP (2020). About human development. Human Development Reports. Available at <http://hdr.undp.org/en/humandev> [accessed on November 28, 2020].

UNDP (2019). *Human Development Report 2019. Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*. New York <http://hdr.undp.org/en/content/human-development-report-2019>.

Glossary

Complex indicator – A complex indicator is an aggregation of composite indicators, a third or higher level of aggregation.

C*omposite indicator* – A composite indicator is a mathematical combination of a set of normalized variables. Within multidimensional data analysis a composite indicator is thought to be a latent phenomenon or construct or variable that is affected by manifest variables or that influences them. A composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model of the multidimensional concept that is being measured.

Index – An index is sometimes a scaled composite indicator or a summary measure. In a composite indicator it is also the highest level of aggregation, opposed to elementary indicator, which is the first level, and subindex or pillar, which is an intermediate level or *dimension*.

Indicator – An indicator is something that points to, measures or otherwise provides a summary overview of a specific concept. We call an *individual* or *elementary indicator* a normalized variable. *Input variable* and *input indicator* have similar meanings to indicator and are used in discussing variable selection and weighting techniques. The statistical literature on inequality measures uses the term "attribute" as a synonym of elementary indicator, and the term "achievement" to denote the observed value of the indicator.

Pillar, subindex, subindicator – Sometimes indicators are aggregated in pillars (also called subindices or subindicators), an intermediate level of aggregation in the construction of a composite indicator; each pillar represents a *dimension* of the composite multidimensional indicator.

Polarity – The 'polarity' of an individual/elementary indicator is the sign of the relation between the indicator and the concept to be measured.

The book addresses some open questions in the construction of composite indicators. It complements well-established references such as the OECD Handbook and provides an insight into the main developments in the field of composite indicators in the near future, especially in the field of well-being and human progress. The first part of the book presents methodologies reflecting the current state of knowledge, while the second part untangles several recent and more critical issues. The book provides useful tools both for researchers with limited specific knowledge on the subject and for scholars who need an update on the latest and most advanced issues in composite indicators.

Silvia Terzi
PhD, is full professor of Statistics, Roma Tre University. Her main research interests are Partial Least Square path models, composite indicators, sustainable development and well-being.

Adrian Otoiu
PhD, is a full-time Lecturer in Statistics and Econometrics at the Bucharest University of Economic Studies. His research interests are labour market issues, composite indicators, demography, and machine learning.

Elena Grimaccia
PhD, is a researcher at the Italian Institute of Statistics (Istat). Her main research interests are composite indicators, multidimensional analysis of socio-economic issues, sustainable development and well-being, structural equation modeling.

Matteo Mazziotta
PhD, is a researcher at the Italian Institute of Statistics (Istat). His main research interests are composite indicators, methods for managing complex surveys and methods for measuring non sampling errors.

Adriano Pareto
PhD, is a researcher at the Italian National Institute of Statistics (Istat). His main research interests are composite indicators, multivariate data analysis and data mining.

ROMA TRE
UNIVERSITÀ DEGLI STUDI