# Monitoring the main aspects of social and economic life using composite indicators: A literature review

Leonardo Salvatore Alaimo and Emiliano Seri

Working paper No. 21-7

Updated: 2021-12-20

REAL

Research group “Economics, policy analysis, and language”

# Monitoring the main aspects of social and economic life using composite indicators: A literature review*

Leonardo Salvatore Alaimo[†] and Emiliano Seri[‡]

Updated: 2021-12-20

**Abstract**

The aim of this project is to research and analyze the main indicators used at the international level to study and monitor equity and inclusion, relevant aspects of social and economic life. In detail, the focus of the work concerns the main measures, both present in literature and produced by international institutions, which allow an analysis of these phenomena from an international comparative perspective. The study presents the sources of data of such different measures and the conceptual and methodological steps to produce them. Moreover, a significant section of the research assesses the statistical methodologies to deal with the complexity of multidimensional systems of indicators and to synthesize them. The project includes an overview of indicators used in the language policy area, including language education and multilingual public administration, and of databases used to populate these indicators.

*The financial support of Esperantic Studies Foundation is gratefully acknowledged.

[†]Department of Social Sciences and Economics, "Sapienza" University of Rome, Rome, Italy
Email: *leonardo.alaimo@uniroma1.it.*

[‡]Department of Statistical Sciences, "Sapienza" University of Rome, Rome, Italy
Email: *emiliano.seri@uniroma1.it*

# 1 Introduction

An increasingly recognized belief is that the evaluation of the progress of a society must be made according to not only economic factors, but also social and environmental. This means that such an evaluation must include measures of inequality and sustainability, key factors in the societal development. In the last decades, there has been a growing interest among researchers in the measurement of these phenomena and the methodological aspects that this entails. It is a common understanding that all socio-economic phenomena (such as development, progress, poverty, social inequality, well-being, quality of life, infrastructure endowment, etc.) must be adequately measured and represented in a multidimensional way, i.e. with the 'combination' of different dimensions to be measured. The main problem is that the complex and multidimensional nature of the phenomena considered makes it difficult to understandably represent them. This gives rise to the need to search for a latent variable to which all the dimensions of the phenomenon considered can be ascribed and to synthesise them into such a latent variable by means of appropriate statistical methodologies. This means dealing with synthetic indicators.

This research project aims to present an overview of the topic of synthesis of statistical indicators systems, highlighting the main conceptual and methodological aspects. Secondly, we focus on the main indicators proposed by international organisations to study relevant aspects of social and economic life, in particular equity and inclusion, in order to give an international comparative perspective. For each indicator, we report the sources of data and the conceptual and methodological steps to produce it.

# 2 Synthesis of multi-indicators systems: Theory and methodology

The complexity and multidimensional nature of phenomena that defines reality (wellbeing, poverty, quality of life, development, and so on) require the adoption of different measures to analyse and understand them. The measurement process in social sciences is associated with the construction of systems of indicators, developed by means of the so-called *hierarchical design* (Maggino, 2017), graphically summarised in Figure 1. Indicators within a system are interconnected and new properties typical of the system and not of its constituent elements emerge from these interconnections. As can be easily understood, these systems are also complex adaptive systems. Therefore, a system of indicators allows the measurement of a complex concept that would not otherwise be measurable by taking into account the indicators individually. They play a key role in describing, understanding and controlling complex socio-economic phenomena. The complex nature of systems of indicators requires approaches allowing more concise views in order to analyse and understand them.

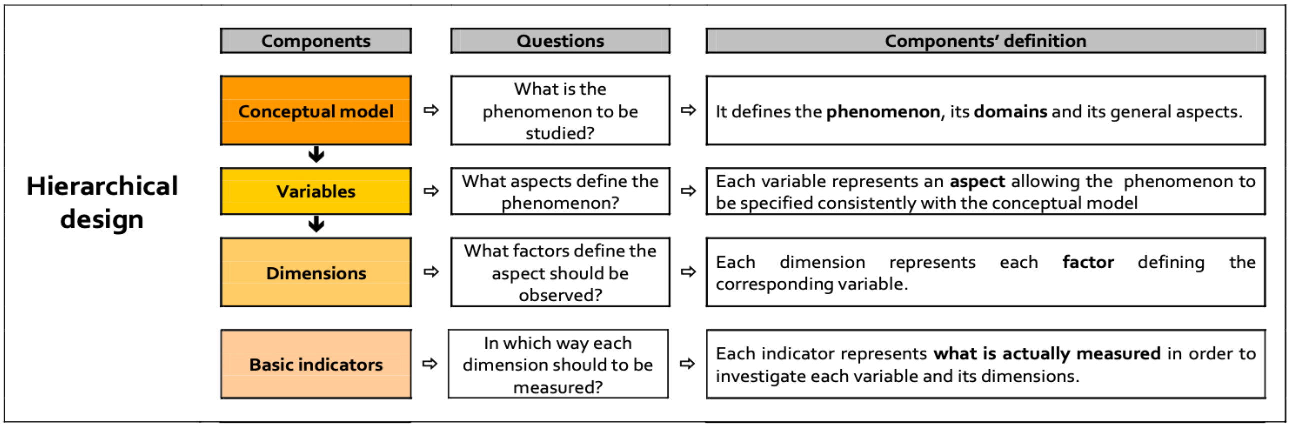


Figure 1: Hierarchical design (source: Maggino (2017).

This means that it is necessary to use a variety of elementary indicators and a criterion for summarising the information they contain. In statistics, an elementary indicator refers to indirect measures of phenomena that cannot be measured directly. In this perspective, an indicator is not simply raw statistical information, but represents a measure organically linked to a conceptual model aimed at describing different aspects of reality. In general terms, an indicator is a quantitative or qualitative measure derived from a set of observed facts that may reveal relative positions (e.g. of a country or a region) in a given area or absolute positions at a given point or range in time. Elementary indicators are appropriately constructed variables that relate to specific aspects and can be considered the first step in the construction of more complex measures. Synthetic indicators are a measure of the level of a complex phenomenon, not directly measurable, obtained by appropriately synthesizing elementary indicators according to established criteria and rules.

Synthetic indicators have been widely used in literature for assessing the progress and making comparisons between countries in different fields, such as well-being (Ciommi et al., 2017; Alaimo et al., 2020), sustainable development (Krajnc and Glavič, 2005; Kondyli, 2010; Alaimo, 2018; Alaimo and Maggino, 2020; Alaimo et al., 2021, 2020), environmental situation and conditions (Kondyli, 2010; Stebbings et al., 2021), labour conditions (Boccuzzo and Gianecchini, 2015; Giannakis and Bruggeman, 2018; Bianchi and Biffignandi, 2020), gender inequalities (Bericat, 2012; Castellano and Rocca, 2014), tourism evaluation (Perez et al., 2013; Tica and Kožić, 2015; Martìn et al., 2018; Suhartanto et al., 2020), and so on. These are only a few examples that do not do justice to the enormous multidisciplinary academic production on synthetic indicator topic. At the same time, synthetic indicators have been widely used by various international organisations and actors to measure the most diverse phenomena, some of which will be presented in the following pages. The main purpose of their success is informative. It is easier for the public to understand a synthetic indicator (one single measure) than many elementary indicators.

At this point, we must start from a question: what does synthesis of multiindicators systems mean? But before answering this question, we need to clarify what is meant by a multi-indicator system. As mentioned above, an indicator system consists of a set of measures (the elementary indicators) of various kinds, observed on a given set of statistical units and selected according to a theoretical framework consistent with the measurement of a given socio-economic phenomenon. In its simplest form, a system of indicators is a matrix of data **X**, typical of multivariate statistics:

$$\mathbf{X} \equiv \left\{x_{ij} : i = 1...N; j = 1...M\right\} = \begin{Bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{Bmatrix} \tag{1}$$

where the columns represent the M indicators, the rows the N statistical units and the generic $x_{ij}$ unit represents the determination of the j-th indicator in the i-th unit. However, in most cases the multi-indicator systems are in the form of a particular type of three-way data array: the *three-way data time array*. These data structures are characterized by a greater complexity of information, consisting in the fact that multivariate data are observed at different *occasions* (for instance, times, places, and so on). In particular, in the three-way data time array occasions are different times in which the multivariate information is collected (D'Urso, 2000; Alaimo, 2020). The three-way data time arrays can be formally represented as follows:

$$\mathbf{X} \equiv \{x_{ijt} : i = 1,...,N; j = 1,...,M; t = 1,...,T\} \tag{2}$$

where $i$ indicates the generic unit, $j$ the generic indicator and $t$ the generic temporal occasion; thus, $x_{ijt}$ represents the determination of the j-th indicator in the i-th unit at the t-th temporal occasion. The study, analysis and synthesis of three-way data time arrays can be complex and require the use of specific statistical tools, which also take into account the temporal perspective. In this work, we will focus on the matrix given in the equation, except where otherwise indicated.

We need to make a clarification. As mentioned above, the topic of synthesis of indicators has a rich and varied scientific literature. There are many approaches that have been developed, as well as many statistical methods and procedures for the synthesis. In this work, we focus on methods suitable for synthesizing systems where all indicators are cardinal. This is because all international indicators (object of this analysis) are constructed using cardinal indicators. There is a large literature on the treatment and synthesis of multidimensional systems of ordinal data using specific methods suitable for such a data. The methods used for non-cardinal indicators are part of the so-called non-aggregative approach, which will be presented briefly in Section 2.2.

From the methodological point of view, synthesis concerns different aspects of the system (Maggino, 2017; Alaimo, 2020):

- *Synthesis of units*

  The aim is to aggregate the units of observation in order to create macrounits to be compared, with reference to the indicators of interest. The statistical methods that allow this to be done are part of the cluster analysis.

- *Synthesis of elementary indicators*

  The aim is to aggregate the values referring to several indicators for each unit of observation, obtaining a synthetic measure. From the technical point of view, statistical methods used in this case can belong to two different approaches: the aggregative-compensative and the non-aggregative.

Obviously these two aspects are not mutually exclusive; on the contrary, it is often necessary to do both for a full understanding of reality. In this research, we focus on the synthesis of elementary indicators.

## 2.1 The aggregative-compensative approach

The aggregative-compensative approach is the dominant framework in literature and it is the one used by international organisations for the construction of their synthetic indicators. Despite its success, it poses some conceptual and methodological questions, the main of which will be addressed in this work. As suggested by the term, the aggregative-compensative approach consists in the aggregation, by means of a mathematical function, of the elementary indicators. These methodologies are defined *composite* indicators (Saisana and Tarantola, 2002; OECD, 2008). Building a composite indicators is not an easy task and requires a step-by-step process (Nardo et al., 2005):

1. Definition of the phenomenon;
2. Selection of basic indicators;
3. Normalization of individual indicators; 4. Aggregation of the normalised indicators;
5. Robustness analysis and validation.

### 2.1.1 Definition of the phenomenon

The first step in any synthesis is the definition of phenomenon and the subsequent identification of the theoretical framework and the relevant variables. It is always necessary that the concept refers to and is inserted within a theoretical framework that gives it meaning. No meaning can be attributed without *subjectivity*. The role of the subject in knowledge production is clear. This is particularly evident for socio-economic phenomena. Different researchers analysing the same phenomenon, using the same definition and the same indicators may arrive at different conclusions.

Fundamental attention must be given to the analysis of the *measurement model*. Its specification is one of the main theoretical assumptions involved in the process of synthesis of indicators. Measurement model refers to the relationship between concepts and indicators. The debate on measurement models is part of the literature on the evaluation of latent variables, which has a long tradition in the social science (Duncan, 1984). Latent variables are phenomena of theoretical interest which cannot be directly observed and have to be assessed by manifest measures which are observable. Two different conceptual approaches can be identified: *reflective* and *formative* (Blalock, 1964; Bollen, 1989; Diamantopoulos and Winklhofer, 2001; Diamantopoulos and Siguaw, 2006; Diamantopoulos et al., 2008). The reflective measurement models have a long tradition in social sciences (in particular, in psychometric research) and are based on classical test theory, according to which measures are effects of an underlying latent construct (Lord and Novick, 1968; Bollen and Lennox, 1991). Therefore, causality is from the construct to the measures. Specifically, the latent variable $\eta$ represents the common cause shared by all items $x_i$ reflecting the construct, where each item corresponds to a linear function of its underlying construct plus measurement error, as shows in equation 3:

$$x_i = \lambda_i \eta + \epsilon_i \tag{3}$$

where $x_i$ is the i-th indicator of the latent variable $\eta$, $\epsilon_i$ is the measurement error for the i-th indicator and $\lambda_i$ is a coefficient capturing the effect of $\eta$ on $x_i$. Measurement errors are assumed to be independent (i.e., Cov[$\epsilon_i, \epsilon_j$] = 0, for *i* 6= *j*) and unrelated to the latent variable (i.e., Cov[$\eta, \epsilon_i$] = 0, for all *i*). A fundamental characteristic of reflective models is that a change in the latent variable, causes variation in all measures simultaneously. All indicators in a reflective model must be positively correlated. Internal consistency is fundamental: correlations between indicators are explained by the model of measurement and two uncorrelated indicators cannot measure the same construct (Bollen, 1984). Each indicator has a specific error component.

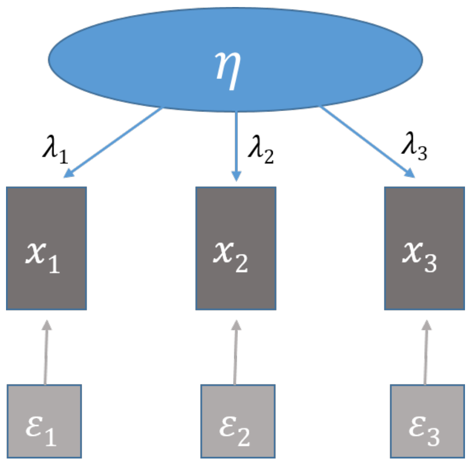


Figure 2: Reflective measurement model (source: Alaimo (2020)).

Typical examples of reflective scenarios include psychometric measures, such as those of attitudes or personality. Figure 2 shows the main components of reflective models and their relationships.

Officially, the formative measurement model was proposed for the first time by Curtis and Jackson (1962). The authors question the need for the measures to be necessarily positively correlated and argue that in specific cases the measures show negative or no correlations, despite the fact that they adopt the same concept. Other authors (Blalock, 1964, 1968; Land, 1970) have subsequently discussed the main specifications of this model, according to which measures are causes of the construct rather than its effects. Indicators determine the latent variable giving it its meaning. The model is specified as follows (equation 4):

$$\eta = \sum_{i=1}^{n} \gamma_i x_i + \zeta \tag{4}$$

where $\gamma_i$ is a coefficient capturing the effect of indicator $x_i$ on the latent variable $\eta$, and $\zeta$ is the error term. The latter includes all remaining causes of the construct which are not represented in and not correlates to the indicators (i.e., Cov[$x_i$,$\zeta$] = 0)[1]. Indicators are not interchangeable; thus, omitting an indicator is omitting part of the construct (this changes the construct). Correlations among indicators are not explained by the measurement model and internal consistency is of minimal importance. There are no specific expectations about patterns or magnitude of correlations among the indicators; formative indicators might correlate positively or negatively or lack any correlation (Bollen, 1984). Indicators have no specific measurement error terms (Edwards and Bagozzi, 2000); in formative models, we only observe disturbance term ($\zeta$) un-correlated with $x_i$ (Edwards, 2011).

Almost all measurement processes of socio-economic phenomena adopt a formative model. Figure 3 shows the main components of formative models and their relationships.

The literature about the difference between reflective and formative models is rich. The state of the theory on formative models has been in intense discussion for some years. Several authoritative scholars (for instance, see: Howell et al., 2007; Wilcox et al., 2008; Edwards, 2011; Aguirre-Urreta et al., 2016) have questioned the validity of this method and published appeals to no longer host its applications in scientific journals. Nowadays, it is quite evident the appropriateness of formative models for measuring a large number of constructs and it is that of all the indicators reviewed in Section 3. At the same time, the result of an incorrect specification of the model is also evident. Important scholars support the validity and effectiveness of the formative models (for instance, see: Bollen, 2007; Diamantopoulos et al., 2008; Bollen and Diamantopoulos, 2017). The debate continues in literature and seems to be far from being resolved. We would like to point out that the choice between the two types of model does not depend directly on the researcher, but exclusively on the

[1] Equation 4 represents a multiple regression equation and, in contrast to equation 3, the latent variable is the dependent variable and the indicators are the explanatory variables.

nature and direction of relationships between constructs and measures (Alaimo, 2020). If the the direction of the relationship is from the construct to the measures we have a reflective or effect model. On the contrary, if the direction of the relationship is from the measures to the construct, we have a formative or causal model. Analysing the measurement model represents a fundamental stage of the process of synthesis, also because it allows the operational definition of the concept. This important issue influences the selection of indicators and the aggregation steps. We focus on phenomena in the economic and sociological field, most of which require a formative measurement model. Therefore, from now on, it is assumed that the measurement model is formative.

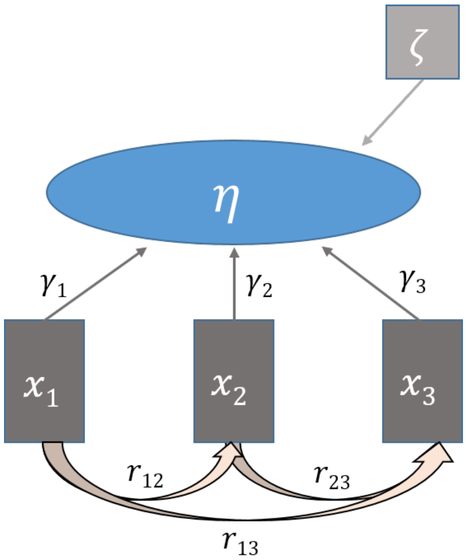


Figure 3: Formative measurement model (source: Alaimo (2020)).

The literature about the difference between reflective and formative models is rich. The state of the theory on formative models has been in intense discussion for some years. Several authoritative scholars (for instance, see: Howell et al., 2007; Wilcox et al., 2008; Edwards, 2011; Aguirre-Urreta et al., 2016) have questioned the validity of this method and published appeals to no longer host its applications in scientific journals. Nowadays, it is quite evident the appropriateness of formative models for measuring a large number of constructs and it is that of all the indicators reviewed in Section 3. At the same time, the result of an incorrect specification of the model is also evident. Important scholars support the validity and effectiveness of the formative models (for instance, see: Bollen, 2007; Diamantopoulos et al., 2008; Bollen and Diamantopoulos, 2017). The debate continues in literature and seems to be far from being resolved. We would like to point out that the choice between the two types of model does not depend directly on the researcher, but exclusively on the nature and direction of relationships between constructs and measures (Alaimo, 2020). If the the direction of the relationship is from the construct to the measures we have a reflective or effect model. On the contrary, if the direction of the relationship is from the measures to the construct, we have a formative or causal model. Analysing the measurement model represents a fundamental stage of the process of synthesis, also because it allows the operational definition of the concept. This important issue influences the selection of indicators and the aggregation steps. We focus on phenomena in the economic and sociological field, most of which require a formative measurement model. Therefore, from now on, it is assumed that the measurement model is formative.

### 2.1.2 Selection of basic indicators

The following step is the selection of elementary indicators, which is generally based on theory, empirical analysis, pragmatism or intuitive appeal (Booysen, 2002). The quality of indicators influence the quality of the resulting composite indicator. The selection process must be guided by the measurement model. This is one of the most delicate phases; although there is no universal rule for the selection of elementary indicators in a system, it is possible to identify some guiding elements:

a) All the dimensions of the phenomenon must be represented and measured.
b) Elementary indicators should not be *redundant*.
c) Elementary indicators must not have missing data.
d) *Polarity* of each indicator must be clear.
e) Assumptions about the nature of the indicators must be made explicit: *substitutability/non-substitutability*.

In general, redundancy can be defined as the excess of significant elements and information in a message over what is strictly necessary for the correct understanding of the message itself. The redundancy of indicators in a system is useful to increase the *reliability* of the measurement. The presence of several indicators (multi-indicator approach) effectively reduces random error. However, we are often faced with systems with too many indicators and synthesis is not possible. It is therefore necessary to reduce their number by excluding some of them from the system. There is no always valid rule for this choice. The theoretical framework and the measurement model must always be kept in mind. In the case of a reflective measurement model, if it is necessary to eliminate indicators from the system, one will certainly start with those that are not correlated with the others (because they do not measure the latent reflective variable considered). Moreover, in such a measurement model, eliminating an indicator has in no case an effect on the latent variable, which remains unchanged. In formative models, the things are different. The exclusion of an elementary indicator always has an effect on the latent variable one wants to measure (hence, on the composite indicator). Moreover, in the case where an elementary indicator has to be eliminated, it is more appropriate to act on indicators which are highly correlated with each other and, consequently, measure the same aspect (dimension) of the phenomenon, rather than to eliminate indicators which are not correlated with each other and, consequently, measure different aspects of the phenomenon.

Substitutability is one of the main assumptions about indicators. The components of a synthetic index are called substitutable if a deficit in one component can be compensated by a surplus in another. The assumption of substitutability of components implies the adoption of additive aggregation methods (e.g. arithmetic mean). The components are defined non-substitutable if no compensation is allowed between them. In the case of partial substitutability or non-substitutability of components, generally, multiplicative (e.g. geometric mean) or non-compensative methods are adopted. Thus, this conceptual assumption has an important effect on the other steps of the composites' construction, in particular the selection of the aggregation function. Aggregation methods can be *compensative* or *non-compensative*, depending on the adoption or not of compensation, and can be classified in *additive aggregation methods* and *multiplicative aggregation methods*. A possible solution identified in literature (Tarabusi and Guarini, 2013; Mazziotta and Pareto, 2016) is the adoption of a *partially compensative method*[2], i.e. allowing it "up to a certain point"; however, the question would arise as to what is the permissible and tolerable threshold of compensability. This issue raises two main problems. On the one hand, choosing one approach over another influences the results, and this is in addition to the effect of different standardisation techniques. Moreover, this choice is often made arbitrarily by the researcher, without taking into account the conceptual assumptions that may justify compensative or non-compensative approach.

The polarity of an elementary indicator is the sign of the relation between the indicator and the phenomenon to be measured (Mazziotta and Pareto, 2017). Therefore, the type of composite we want to construct defines polarity. In other words, some indicators may be positively related with the phenomenon to be measured (positive polarity), whereas others may be negatively related with it (negative polarity) (Alaimo, 2020). For example, in the case of human development, the GDP has positive polarity. However, in the case of multidimensional poverty, GDP has negative polarity. When a composite index must be constructed, all the individual indicators must have positive polarity, so it is necessary to 'invert' the sign of the indicators with negative polarity. Inversion of polarity may be performed before normalizing or jointly, the results are in most cases the same. The main inversion polarity techniques are the following:

- the *linear transformation*, in which we take the complement with respect to maximum value. Given the data matrix $\mathbf{X} \equiv \{x_{ij}\}$ in the equation 1, the linear inversion of polarity is calculated as follows:

$$x'_{ij} = Max_{x_j} - x_{ij} \tag{5}$$

[2] Among those summarised in Table 9, partially compensative methods are MPI and AMPI, and Mean-Min Function

where $Max_{xj}$ is the absolute maximum value of the j-th indicator, $x_{ij}$ is the value of the indicator j-th in the unit i-th and $x'_{ij}$ is the inverted value. This is the simplest technique, and it allows to save the same *distance* between units, with a different origin. It is particularly used with ranking, standardization, and re-scaling normalisation methods.

- The *non-linear* transformation, in which we take the reciprocal of the value to be inverted. Given the data matrix $\mathbf{X} \equiv \{x_{ij}\}$ in the formula 1, the inversion is calculated as follows:

$$x'_{ij} = \frac{1}{x_{ij}} \tag{6}$$

This technique, typically used with indicization, is criticized because it modifies the distances between units and it requires all values are greater than 0.

- A particular case is the so-called *double polarity*, in which we observe an indicator presenting positive polarity below a certain threshold and negative above it or vice versa. Examples of such an indicator is female to-male ratios, i.e. the ratio between the percentage of female and the percentage of males. These ratios are particularly used for measuring *gender gap* (WEF, 2020): they have a positive polarity up to the value of 1 (which expresses the gender equality between women and men); from 1 on, the polarity is reversed (in this case, it expresses a situation of disadvantage of the men with respect to the women). In this case, we can use the *triangular* transformation:

$$x'_{ij} = |\lambda_{x_j} - x_{ij}| \tag{7}$$

where $\lambda_{xj}$ is the value of the indicator j-th in which the polarity inverts (the threshold).

### 2.1.3 Normalization of individual indicators

In the normalization step, the researcher must select a mathematical transformation in order to make the indicators comparable. Normalization is required before any data aggregation as the indicators in a data set often have different measurement units and ranges. Moreover, this step allows all basic indicators to have the same *polarity*. Normalization is a very delicate step, because it can change the distribution and internal variability of indicators. There are different methods, each of which presents advantages and drawbacks. Choosing one rather than another has effects on the synthesis obtained. This problem can be partially overcome by performing a robustness analysis to evaluate the effects of different procedures on the results obtained (Freudenberg, 2003). However, from a conceptual point of view, normalisation does not solve the problem of putting together different measures, of *mixing apples and oranges* (Alaimo, 2020). Generally, given the matrix $\mathbf{X} \equiv \{x_{ij}\}$ reported in equation 1, with the normalization we obtain a new matrix, $\mathbf{R} \equiv \{r_{ij}\}$, in which the generic element $r_{ij}$ is the normalised value of the j-th indicator for the i-th unit. We report some of the most common normalization methods, their formulas, their pros and cons. For each method we also report an example, using the data of Italy, Germany, France and Spain of the three dimensions of the Human Development Index (HDI) (the index is described in Section 3)

- **Ranking**:

$$r_{ij} = rank(x_{ij}) \tag{8}$$

The generic normalized value $r_{ij}$ is obtained by ranking the values of each elementary indicator. Units with the same value receive a rank equal to the mean of the ranks they span. If indicator j has negative polarity, the rank order must be reversed. The main advantage is that this method is not affected by the presence of outliers, while the main drawback is that it assumes the same distance between every unit. An example of ranking normalization is in the following tables:

| | Health | Education | Income |
|---|---|---|---|
| ITA | 83,51 | 13,23 | 10,66 |
| DEU | 81,33 | 15,56 | 10,92 |
| FRA | 82,66 | 13,56 | 10,76 |
| ESP | 83,57 | 13,93 | 10,62 |

Table 1: Original values

| | Health | Education | Income |
|---|---|---|---|
| ITA | 2 | 4 | 3 |
| DEU | 4 | 1 | 1 |
| FRA | 3 | 3 | 2 |
| ESP | 1 | 2 | 4 |

Table 2: Rankings

- **Re-scaling (or Min-Max)**:

$$r_{ij} = \frac{x_{ij} - \min_i(x_j)}{\max_i(x_j) - \min_i(x_j)} \tag{9}$$

Where $\min_i(x_j)$ and $\max_i(x_j)$ are, respectively, a minimum and a maximum value (commonly the observed ones) that represent the possible range of the indicator j. The main advantage of this method is that the range for indicators with very little variation will increase and these will contribute more to the composite indicator than they would using another method; moreover, the range [0,1] gives an easy reading of the considered phenomena. The drawback is that it is based on the range and, consequently, it is sensitive to outliers. Dealing with multi-indicator system over time, this is the most commonly used normalization method. An example of Min-Max normalization is in the following tables:

|  | Health | Education | Income |
|---|---|---|---|
| ITA | 83,51 | 13,23 | 10,66 |
| DEU | 81,33 | 15,56 | 10,92 |
| FRA | 82,66 | 13,56 | 10,76 |
| ESP | 83,57 | 13,93 | 10,62 |

Table 3: Original values

|  | Health | Education | Income |
|---|---|---|---|
| ITA | 0,97 | 0,00 | 0,13 |
| DEU | 0,00 | 1,00 | 1,00 |
| FRA | 0,59 | 0,14 | 0,47 |
| ESP | 1,00 | 0,30 | 0,00 |

Table 4: Re-scaled values

- **Standardization (or Z-scores)**:

$$r_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_j} \tag{10}$$

where $x_j$ is the arithmetic mean of the indicator j-th, and $\sigma_j$ is its standard deviation. The main advantage of this method is that reports all indicators to the same distribution, i.e. a standard Gaussian $N \sim (0;1)$, and, consequently, this greatly simplifies the analysis. The main drawbacks are that the presence of negative values can be a limitation for some aggregation methods (i.e. geometric mean) and it can not be used for time series, as in that case the standard deviation may be affected by the variability in the time units prior to the one considered. An example of Z-score normalization is in the following tables:

|  | Health | Education | Income |
|---|---|---|---|
| ITA | 83,51 | 13,23 | 10,66 |
| DEU | 81,33 | 15,56 | 10,92 |
| FRA | 82,66 | 13,56 | 10,76 |
| ESP | 83,57 | 13,93 | 10,62 |

Table 5: Original values

|  | Health | Education | Income |
|---|---|---|---|
| ITA | 0,66 | -0,79 | -0,53 |
| DEU | -1,28 | 1,40 | 1,20 |
| FRA- | 0,10 | -0,48 | 0,13 |
| ESP | 0,71 | -0,13 | -0,80 |

Table 6: Z-score values

- **Distance from a reference (or Indicization)**:

$$r_{ij} = \frac{x_{ij}}{x^*_{\cdot j}} * 100 \tag{11}$$

where $x^*_{\cdot j}$ is a reference value belonging to or calculated on the same distribution of $x_j$. The main advantage is that this method retains the relative distance between the different units, while the main drawback is that it is highly sensitive to outliers. An example of indicization normalization is in the following tables, with the arithmetic mean of each indicator chosen value of $x^*_{\cdot j}$:

For more details on the different normalization methods, please see: Nardo et al. (2005); OECD (2008); Mazziotta and Pareto (2017).

| | Health | Education | Income |
|---|---|---|---|
| ITA | 83,51 | 13,23 | 10,66 |
| DEU | 81,33 | 15,56 | 10,92 |
| FRA | 82,66 | 13,56 | 10,76 |
| ESP | 83,57 | 13,93 | 10,62 |

Table 7: Original values

| | Health | Education | Income |
|---|---|---|---|
| ITA | 100,90 | 94,03 | 99,26 |
| DEU | 98,26 | 110,59 | 101,68 |
| FRA | 99,87 | 96,38 | 100,19 |
| ESP | 100,97 | 99,00 | 98,88 |

Table 8: Indicization values

### 2.1.4 Aggregation of the normalised indicators

The following step is the *aggregation* of normalized indicators. In literature many methods have been proposed for constructing composites (Saisana and Tarantola, 2002). Obviously, each method has its pros and cons; there is no such thing as the *best method*. The method used has an impact on the results obtained; in particular, the *weighting* and the *aggregation* are critically important steps.

The choice of weighting has a large impact on values and, consequently, on the meaning of the composites. Thus, it is essential to understand the effects of one choice over another. In literature, there are different approaches to the weighting issue (Gan et al., 2017). No agreed methodology exists to weight basic indicators (for a more detailed analysis, please see: Alaimo and Maggino (2020). Equal weighting is the simplest and the most commonly used, but it is not without criticism. In fact, giving the same weight does not mean not weighting, but giving all indicators the same weight and, consequently, the same importance. Probably, a better choice than others could be the public/expert opinion-based weighting. When the latter cannot be used, a good strategy could be the selection of a limited number of robust indicators, giving them the same weight.

One of the main issues of aggregative methods is related to the way in which they are calculated, i.e. as a combination of basic indicators. Different methods of aggregation exist and different classifications for those methods have been proposed. We focus on the classification based on the degree of compensation (substitutability) tolerated, according to which the widely used aggregation methods include:

- *Additive aggregation methods*

  These methods employ functions that sum up the normalized values of basic indicators to form a composite index. The most widespread additive method is the *weighted arithmetic mean*. Starting from the matrix of the normalized indicators $\mathbf{R} \equiv \{r_{ij}\}$ obtained from the matrix $\mathbf{X} \equiv \{x_{ij}\}$ by means of a normalisation method, the value of the composite indicator for a statistical unit i-th ($ci_i$) is obtained as follows:

$$ci_i = \sum_{j=1}^{M} r_{ij} w_j \tag{12}$$

where $w_j$ is the weight assigned to the j-th indicator. In the case of equal weighting, i.e. if $w_j = \frac{1}{M}$, we have the simple arithmetic mean. Similarly, given the three-way data array $\mathbf{R} \equiv \{r_{ijt}\}$ of the normalized data obtained from the original three-way data array $\mathbf{X} \equiv \{x_{ijt}\}$, the composite is given by:

$$ci_{it} = \sum_{j=1}^{M} r_{ijt} w_j \tag{13}$$

This technique implies full compensability, such that poor performance in some indicators can be compensated for by sufficiently high values in other indicators. The main advantage of this method is that it is simple, largely known and gives easy-to-understand results. The main drawback is that it is a full *compensative* method. This assumption, as is evident, is very strong and has a great impact on the results obtained, leading in many cases to an extreme flattening of the differences between the units. The additive methods, and in particular the arithmetic mean, have been widely used in the literature. For example, it has been used initially in the calculation of HDI, and it is still used for the calculation of the dimensional indicator of Education in the HDI.

- *Multiplicative aggregation methods*

Geometric aggregation methods use multiplicative instead of additive functions. The most widespread geometric aggregation function is the *weighted geometric mean*. Starting from the matrix of the normalized indicators $\mathbf{R} \equiv \{r_{ij}\}$, the value of the composite indicator for a statistical unit i-th ($ci_i$) is obtained as follows:

$$ci_i = \sqrt[M]{\prod_{j=1}^{M} r_{ij}^{w_j}} \tag{14}$$

where $w_j$ is the weight of the indicator *j*. In the case of equal weighting, we have the simple geometric mean. Similarly, given the three-way data array $\mathbf{R} \equiv \{r_{ijt}\}$ of the normalized data, the synthetic measure is given by:

$$ci_{it} = \sqrt[M]{\prod_{j=1}^{M} r_{ijt}^{w_j}} \tag{15}$$

Multiplicative methods only allow compensability between indicators within certain limitations (partially compensative). This requirement exists because of the "geometric-arithmetic means inequality" (Gan et al., 2017), which limits the ability of indicators with very low scores to be fully compensated by indicators with high scores. Simultaneously, significant marginal effects may be measured using geometric methods when increasing the values of indicators with relatively low absolute values (OECD, 2008).

- *Non-compensatory aggregation methods*

Additive and multiplicative aggregations imply the (respectively, total and partial) compensation among basic indicators. When substitution between indicators is deemed unacceptable, *non-compensatory* aggregation methods become important. A non-compensatory approach generally requires the use of non-linear functions (Munda and Nardo, 2005). We present two of the most used methods belonging to this approach.

– Mazziotta-Pareto Index (*MPI*) and Adjusted Mazziotta-Pareto Index (*AMPI*):
  The *MPI* and the *AMPI* has been developed in Mazziotta and Pareto (2016). In the *MPI*, the indicators must be normalized using a variant of the Z-scores method, as follows:

$$r_{ij} = \frac{x_{ij} - \overline{x}_j}{\sigma_j} * 10 + 100 \tag{16}$$

  In this way, we do not have negative values and the synthetic values will be in the range [70,130]. The composite index is given by:

$$MPI_i = \mu_{r_i}(1 \pm cv_i^2) \tag{17}$$

  Where $\mu_{r_i} = \frac{\sum_{j=1}^{J} r_{ij}}{J}$ with $j = 1,\dots,J$ elementary indicators, $cv_i = \frac{\sigma_{r_i}}{\mu_{r_i}}$ is the coefficient of variation of the normalized values of the ith unit with $\sigma_{ri}$ vector of standard deviations and the sign + or − depends on the polarity of the considered phenomena (+ for negative polarity and − for positive polarity). The idea is to penalize indicators with high variability in order to reduce the *compensation* problem. This method allows relative comparisons of a generic unit i-th to the other units. Since it is based on a Z-score normalisation, it cannot be used for the three-way data time arrays.
  To overcome this limitation, the *AMPI* has been proposed. It is a variant of the *MPI*, based on a Min-Max normalization and a rescaling of the basic indicators in a range (70;130), according to two goalposts, representing a minimum and a maximum value of each variable for all units and time periods. Given the three-way data array $\mathbf{X} \equiv \{x_{ijt}\}$, first data is normalised by using a variant of the Min-Max method as follows:

$$r_{ijt} = \frac{(x_{ijt} - MIN_{x_j})}{(MAX_{x_j} - MIN_{x_j})} * 60 + 70 \tag{18}$$

  where $x_{ijt}$ is the value of the j-th indicator in the i-th unit at the t-th time; $MIN_{xj}$ and $MAX_{xj}$ are the two goalposts of the indicator $j$ and $r_{ijt}$ is the normalized value. If the basic indicator has positive polarity, the formula 18 is used; otherwise, the formula 19 is calculated:

$$r_{ijt} = \frac{(MAX_{x_j} - x_{ijt})}{(MAX_{x_j} - MIN_{x_j})} * 60 + 70 \tag{19}$$

The two goalposts are defined as follows:

$$Ref_{x_j} \pm \Delta \quad \text{with} \quad \Delta = \frac{(\underset{it}{Max}_{x_{ijt}} - \underset{it}{Min}_{x_{ijt}})}{2} \tag{20}$$

where $Ref_{xj}$ is the reference value, i.e. the value of the j-th indicator in a specific unit t at a specific time t and $Max_{Xijt}$ and $Min_{Xijt}$ are, respectively, the maximum and the minimum value of the jth indicator in all units and all time periods. Thus, each indicator assumes the value 100 for the reference unit considered in the time occasion considered in all basic indicators; all the other values of each unit for all the time occasions will be expressed in reference to this value, allowing a comparison in time and space. Finally, *AMPI* is computed as follows:

$$AMPI^{\pm} = \underset{it}{\mu}_{r_{ijt}} \pm \underset{it}{\sigma}_{r_{ijt}} * \underset{it}{cv}_{r_{ijt}} \tag{21}$$

Where $\underset{it}{\mu}_{r_{ijt}}\ \underset{it}{\sigma}_{r_{ijt}}\ \underset{it}{cv}_{r_{ijt}}$ are respectively the arithmetic mean, the standard deviation and the coefficient of variation of the values of all *M* basic indicators in the i-th unit at the t-th temporal occasion. The sign ± depends on the type of phenomenon measured. If the composite is positive, i.e. increasing values of the index correspond to positive variations of the phenomenon considered, then *AMPI* with negative penalty ($AMPI^{-}$) is used; otherwise, we compute $AMPI^{+}$. This index is characterized by the combination of a *medium effect*
($\mu_{rijt}$) and a *penalty effect* ($\sigma_{rijt} * cv_{rijt}$), which allows penalizing units with unbalanced values of standardised indicators. The penalty wants to favour units which, mean being equal, have a greater balance among the various indicators. All values will be approximately within (70,130). The composite often has values outside this range. This could be either a limit of AMPI and a quality, as it allows highlighting the presence of a strong variability in the time series of the basic indicators.
This partially non-compensatory composite indicator bas been used for the synthesis of many different phenomena and is the one used by Istat for the construction of Equitable and Sustainable Well-being (BES) composite indicators since 2015 (Italian National Institute of Statistics, 2015).

– Mean-Min function (*MMF*):

First developed in (Tarabusi and Guarini, 2013), the Mean-Min Function (*MMF*) is a two-parameter function that incorporates two extreme cases of penalization of unbalance: the zero penalization represented by the arithmetic mean (compensatory approach) and the maximum penalization represented by the minimum function (noncompensatory approach). The composite index is defined as:

$$MMF_i = \mu_{r_i} - \alpha \left( \sqrt{\left( \mu_{r_i} - \min_j \{r_{ij}\} \right)^2 + \beta^2} - \beta \right)$$

$$(0 \leq \alpha \leq 1; \beta \geq 0) \quad (22)$$

where $\mu_{ri}$ is the mean of the normalized values (through any method of normalization) for unit i, and the parameters $\alpha$ and $\beta$ are respectively related to the intensity of penalization of unbalance and degree of complementarity between indicators. The function reduces to the arithmetic mean for $\alpha = 0$ and to the minimum function for $\alpha = 1$ and $\beta = 0$. The main advantages of this method are that it is independent from the choice of the normalization method and, by choosing the values of parameters appropriately, it could be obtained the aggregation function that best suits the specific theoretical approach. The main drawback is that there is not a general rule for choosing these values.

– Benefit of the Doubt (*BoD*):

The Benefit of the Doubt (*BoD*) approach is an aggregative method for composite indicators construction first developed in (Rogge et al., 2006), based on the Data Envelopment Analysis (DEA) [3]. The efficiency of a set of indicators can be adapted to construct a synthetic indicator using an input-oriented DEA. The synthetic measure is obtained as the weighted sum of the normalized indicators relatively to a benchmark; more precisely, it is defined as the performance of the single unit divided for the performance of the benchmark:

$$BoD_i = \frac{\sum_{j=1}^{M} r_{ij} w_{ij}}{r_{ij}^*} \quad (23)$$

where $r_{ij}$ is the normalized value of the j-th indicator for the i-th statistical unit according to Equation 9 and $w_{ij}$ is the corresponding weight. The benchmark $r_{ij^*}$ is defined as follows:

$$r_{ij}^* = \max_{r_{i \in [1 \dots N]}} \sum_{j=1}^{M} r_{ij} w_{ij} \quad (24)$$

The identification of the optimal set of weights guarantees that each unit is associated to the best possible position compared to all the others. The optimal weights are obtained by solving the equation:

[3] DEA is a linear programming technique, useful to measure the relative efficiency of decision making units (DMU) on the basis of multiple inputs and outputs (Charnes et al., 1978).

$$BoD_i^* = \max_{w_{ij}} \frac{\sum_{j=1}^{M} r_{ij} w_{ij}}{\max\limits_{k \in [1...N]} \sum_{j=1}^{M} r_{kj} w_{kj}}, \forall i = 1...N \tag{25}$$

under the constraints that the weights are non-negative and the result is bounded [0,1]. The most favorable weights are always applied to all observations. The main advantages of this method are related to the DEA solution: since the weights are specific for each unit, cross-unit comparisons are not possible and the values of the scoreboard depend on the benchmark performance. Another drawback is the multiplicity of equilibria. Hiding a problem of multiple equilibria makes the weights not uniquely determined (even if the composite indicator is unique). The optimization process could lead to many 0-weights if no restrictions are imposed on the weights. This method has been widely used in several fields like the European labor market analysis (Storrie and Bjurek, 2000); the European social inclusion policies evaluation (Cherchye et al., 2004); the internal market policies evaluation (Cherchye et al., 2007).

Table 9: Most used normalization and aggregation methods

| | |
|---|---|
| Normalization methods | Ranking |
| | Re-scaling (or Min-Max) |
| | Standardization (or Z-scores) |
| | Distance from a Reference (or Indicization) |
| Aggregation methods | Arithmetic mean |
| | Geometric mean |
| | Benefit of the Doubt (BoD) approach |
| | Mean-Min Function |
| | Mazziotta - Pareto Index (MPI) or Adjusted-MPI (for time series) |

The aggregation methods above the thick line are compensative and the ones under are only partially compensative. As example of usage, in the following table we present the results of the analyzed aggregation methods for the four countries data of the HDI dimensions previously used:

Table 10: Aggregation methods results for four countries data of the HDI dimensions

| | Arithmetic mean | Geometric mean | MPI | Mean-Min Function | BoD |
|---|---|---|---|---|---|
| ITA | 97,66 | 97,44 | 96,97 | 95,23 | 1,00 |
| DEU | 104,71 | 103,84 | 102,26 | 98,59 | 1,00 |
| FRA | 98,51 | 98,47 | 98,40 | 94,45 | 0,98 |
| ESP | 99,12 | 98,88 | 98,41 | 95,54 | 1,00 |

Arithmetic mean, Geometric mean and MPI have been computed on the data normalized using Equation 16, while Mean-Min Funtion and the BoD on the data normalized using Equation 9. The values of $\alpha$ and $\beta$ chosen in the Mean-Min Function are $\alpha$ = 0.5 and $\beta$ = 1. It is easily understandable that values achieved using different methods can not be compared. However, in order to compare the performance of each country between the different methods, we can transform the values obtained into rank using Equation 8 and then compare the position that each country achieves according to the method used:

| | Arithmetic mean | Geometric mean | MPI | Mean-Min Function | BoD |
|---|---|---|---|---|---|
| ITA | 4 | 4 | 4 | 3 | 1 |
| DEU | 1 | 1 | 1 | 1 | 1 |
| FRA | 3 | 3 | 3 | 4 | 4 |
| ESP | 2 | 2 | 2 | 2 | 1 |

As already highlighted, the aggregative-compensative approach is the dominant framework. In Table 9, the most used normalization and aggregation methods are reported. This approach to synthesis have been used for the construction of some of the most *"popular"* synthetic indicators for studying socio-economic issues and inequalities in modern societies, that will be analyzed in Section 3. The topics listed above are related to the subjective choices involved in different steps of composite construction. Subjectivity is an essential element in any measurement process, but its presence does not make the process arbitrary (Alaimo, 2020). The researcher should make choices based on the conceptual framework, but this is often not the case. This can lead to arbitrariness.

## 2.2 Non-aggregative approach

Despite its success, the aggregative-compensative approach has been deeply criticized as inappropriate and often inconsistent, from both conceptual and methodological point of view (Freudenberg, 2003; Maggino, 2017). Indicators are rarely homogeneous in many respects; the aggregating technique might introduce implicitly meaningless compensations and trade-offs among indicators; it is not clear how to combine ordinal variables and use numerical weights. This leads to a fundamental question: is the aggregation the only road to synthesis? To answer this question and to overcome, or at least diminish, the limitations of aggregative procedures, statistical research has focused on developing alternative procedures to synthesis, based on non-aggregative methods (Alaimo, 2020). In simple terms, the synthesis in these methods is achieved without the aggregation of the elementary indicators. One In particular, we consider one of the most used procedures in the synthesis field, that based on the application of Partially Ordered Set (poset). This methodology supplies concepts and tools that appropriately adapt to the needs of synthesis. This method is particularly suitable for the treatment of ordinal data, but it can also be applied to systems of mixed indicators (see, for instance: Annoni and Bruggemann (2009); Bru¨ggemann and Patil (2011); Fattore et al. (2015); Carlsen and Bru¨ggemann (2017); Di Brisco and Farina (2018); Arcagni et al. (2019); Rimoldi et al. (2020); Alaimo and Conigliaro (2021); Alaimo et al. (2021)). This is certainly a first advantage over the traditional aggregative methods, which can only be applied to cardinal data. Synthesis is obtained by means of *profiles* (i.e., the combinations of scores of each statistical unit in the basic indicators considered) and reflects the relational position of the statistical unit's profile with respect to all the others. Therefore, synthetic measures are obtained without any normalisation and aggregation of the basic indicators (Brüggemann and Patil, 2011; Fattore, 2017) and this is undoubtedly another advantage. Obviously, poset presents limitations and problems. In particular, as with other ordinal methods, it is highly computationally demanding. Indeed, as the number of observations and/or variables increases, the computation complexity and time increases. There is no perfect approach (just as there is no best method), but each of them has its strengths and weaknesses. The main advantages and disadvantages of the two approaches are presented in Table 11.

Table 11: Aggregative-compensative approach and poset: main strengths and weaknesses.

| | **Benefits** | **Drawbacks** |
|---|---|---|
| **Aggregative-compensative approach** | Simple implementation<br>Easy-to-read results | Different normalisation methods<br>Different aggregative methods<br>Not suitable for ordinal indicators |
| **Poset** | No normalisation and aggregation of basic indicators<br>Suitable for indicators of different scaling levels | Highly computationally demanding |

The procedure for applying posets to ordinal indicator systems or mixed systems is slightly different. However, there are many toy examples in literature which describe the different steps in detail (for instance, see: Alaimo et al. (2021,)). In the following Section, we present the main concepts and definitions, by using a toy example reported in Alaimo et al. (2021).

## 2.3 Poset: basic concept and definitions

Before describing the basic concepts of poset, we propose a small example useful to understand it better. Suppose we have 5 objects on which the presence or absence of 3 properties or attributes is observed. For simplicity, the absence of a property will be encoded with 0, the presence with 1. The result is the Table 12.

Table 12: Poset example: system of 3 ordinal attributes for 5 objects.

| Objects | X | Y | Z |
|---|---|---|---|
| A | 1 | 1 | 1 |
| B | 1 | 0 | 1 |
| C | 0 | 1 | 1 |
| D | 1 | 1 | 0 |
| E | 0 | 0 | 0 |

We want to determine if it is possible to establish a rank between the objects considered, that is, if it is possible to say that one object is better than another one. In fact, it is often the final ranking that is the goal of a synthesis, rather than the exact scores (Fattore, 2017). Looking at data reported in Table 12, object A can be classified as "better" (whatever this means in specific contexts) than all the other ones, because it presents all the attributes considered. For the same reason, we can classify object E as the "worst", since it has no attributes. What about the other objects? They present similar situations, since they have 2 attributes of the 3 considered. However, it is not possible to establish a rank between these three objects: we cannot say, for example, that B is better than C since they have different combinations of attributes. They have conflicting achievements and, consequently, are not comparable. This exactly means dealing with a *partially ordered set*. Addressing the synthesis of such a system of indicators using the aggregative approach involves conceptual and methodological limitations. The use of aggregative methods presupposes that the indicators are cardinal, that is, the modalities they assume are numbers. These methods are, therefore, not suitable for ordinal variables, whose modalities are not numerical, even though they are often coded using numbers (as in the example in Table 12). Despite being conceptually wrong, however, the use of aggregative methods to synthesize systems with ordinal indicators is common practice in literature. This leads to misleading results and conclusions. For example, applying the arithmetic mean [4] to synthesize data in Table 12, objects A and E would have, respectively, the best and worst rank. The other three objects would all obtain the same score (0.67) and, consequently, the same rank, although, as mentioned above, they have different combinations in the basic indicators. Thus, the application of an aggregative method *makes comparable incomparabilities* among statistical units. Poset gives analytical tools to better deal with system presenting ordinal indicators, allowing the

[4] These considerations are independent of the aggregation method used.

construction of a synthesis that is not the result of an aggregation of the scores of basic indicators.

Given a finite object set $X$ consisting of several units of analysis $x_i$, $X = \{x_i\}$, if we can compare those units using a binary relation E the set is equipped with a *partial order* and we can call it a poset (partially ordered set). More precisely, a poset $\Pi = (X, \trianglelefteq)$ is a set $X$ equipped with a partial order relation $\trianglelefteq$ satisfying three main properties (Davey and Priestley, 2002):

- the first property is called *reflexivity* and indicates that an object can be compared with itself, i.e. $x \trianglelefteq x$ for all $x \in X$;
- the second property, *anti-symmetry*, states that, given two generic elements $a$ and $b$ belonging to the set $X$, if $b$ is better than $a$ and, at the same time, $a$ is better than $b$, then the two elements are identical; i.e. if $a \trianglelefteq b$ and $b \trianglelefteq a$ then $a = b$, $a,b \in X$;
- *transitivity* is present if the units are, at least, ordinal scaled and stated the possibility of defining an order among them. i.e. if $a \trianglelefteq b$ and $b \trianglelefteq c$, then $a \trianglelefteq c$, $a,b,c \in X$.

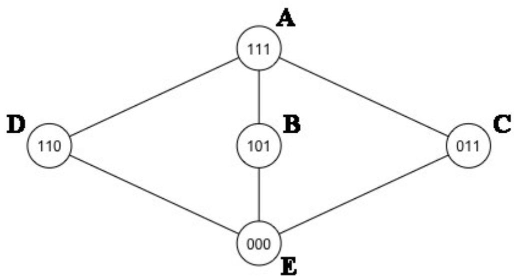


Figure 4: Hasse diagram of the system in Table 12.

If $a \leq b$ or, alternatively, $b \leq a$ then they are comparable, otherwise incomparable. The structure of comparabilities is defined by a matrix, called *incidence matrix*, $Z_P = (z_{ij}) \in Z^{k \times k}$ where $|X| = k$ is the cardinality of $X$ and $z_{ij}$ is equal to 1 if $x_i \trianglelefteq x_j$, 0 otherwise, with $x_i, x_j \in X$. Given two elements $x_i, x_j \in X$, $x_j$ covers $x_i$ ($x_i \prec x_j$) if $x_j$ dominates $x_i$ ($x_i \trianglelefteq x_j$) and there is no other element $x_s \in X$ that jointly dominates $x_i$ and is dominated by $x_j$ ($x_i \trianglelefteq x_s \trianglelefteq x_j$). Dealing with a multi-indicator system, the elements of the poset correspond to the combinations in the basic indicators for each statistical unit, the *profiles*. Given two profiles, $x$ and $y$, we will say that $x$ covers $y$ only if it has a profile with values in all the indicators equal to and at least one greater than those of $y$. Looking at Table 12, we can say that A covers all other elements of the set. If $x$ has a higher value in one indicator than $y$ and the latter has a value in another indicator higher than $x$, regardless of the values assumed in the other indicators, the two profiles are incomparable, since they actually express situations not akin with each other. In the reported example, $B$ and $C$ are incomparable, because $B$ has a value in the indicator $X$ higher than $C$ but $C$ presents a value in the indicator $Y$ higher than $B$. The corresponding incidence matrix is the following:

$$\mathrm{Z}_P = \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} . \tag{26}$$

A partially ordered set can be represented by means of a directed graph without cycles called *Hasse diagram*, in which the nodes are the elements of the sets. In the case of a system of indicators, each edge represents a specific profile. It graphically summarises the information in the incidence matrix. This diagram should be read from top to bottom and two elements are comparable $\trianglelefteq$ if an edge connects them in the diagram. Hasse diagram provides a vertical information regarding the comparabilities within the poset and a horizontal one about the incomparabilities among nodes, expressing the uncertainty in the set. Obviously, nodes connected by a path are comparable by transitivity. Figure 4 reports the Hasse diagram of the example presented in this work.

We must introduce two other crucial concepts. An *extension* of $\Pi = (X, \trianglelefteq)$ is a poset $\Pi_e = (X, \trianglelefteq_e)$ defined on the same set $X$ but equipped with a relation $\trianglelefteq_e$ that extends the relation $\trianglelefteq$. The consequence is that all the pairs of elements comparable in $\trianglelefteq$ are comparable in $\trianglelefteq_e$, while some pairs comparable in $\trianglelefteq_e$ are not comparable in $\trianglelefteq$. An extension of a poset is defined *linear* if all the elements of the set $X$ are comparable; in other words, it is a linear order obtained extending the starting poset so that all elements of the set $X$ are comparable. A poset generally has a set of linear extensions, $\Omega_\Pi$. An interesting property (Fattore, 2017) is that a poset is uniquely identified by a set of linear extensions that is different from that of any other poset and it is the result of the intersection of its linear extensions. Thus, we can study properties of a poset starting from the analysis of the set of its linear extensions. The latter, being linear, are easier to study and examine. Linear extensions, therefore, dissolve the incomparabilities present in the poset: given two generic incomparable elements, *a* and *b*, in some linear extensions *a* dominates *b*, while in others *b* dominates a. The *mutual ranking probability (MRP) matrix* of Π is a $k \times k$ (where $k$ is the number of elements of the set) matrix $M_\Pi = (m_{ij})$, where $m_{ij}$ is the fraction of linear extensions in $\Omega_\Pi$ such that the element $x_i$ is dominated by the element $x_j$.

In using poset for analysing multi-indicators systems, we define the structure of comparabilities among the units of the systems and analyse it by means of some mathematical tools. First, we want to give a *score* to each element of the set, in order to reduce the complexity. This is obtained by means of the *average rank*. Generally, the *rank* of an element $x_i$ in a linear extension ` is 1 plus the number of elements which dominates $x_i$ in `. Consequently, the *average rank* of an element $x_i \in \Pi$ is the average over $\Omega_\Pi$ of the ranks of $x_i$ in the linear extensions. The vector of average ranks of the poset elements $\mathbf{h}_\Pi$ is equal to the vector of row sums of the *MRP* matrix[5]. The MRP matrix and the average ranks vector (AvR) of the example reported in Table 12 are the following:

$$
Z_P = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 1.0 & 0.5 & 0.5 & 0.0 \\ 1.0 & 0.5 & 1.0 & 0.5 & 0.0 \\ 1.0 & 0.5 & 0.5 & 1.0 & 0.0 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix} \end{array} \begin{array}{c} AvR \\ 1 \\ 3 \\ 3 \\ 3 \\ 5 \end{array} . \tag{27}
$$

Average rank is bounded between a minimum, equal to 1, corresponding to the element with no others above it in the linear extensions (the best one) and a maximum, equal to the number of all elements of the poset. It represents the position of each element in the general order. We can integrate this information with that expressing the situation in terms of evaluation of satisfaction of each profile. To do this, we need to define a criterion capable of determining whether a profile belongs to the satisfied or dissatisfied class. Thus, we identify one or more *threshold profiles*, compared to which we identify the satisfied profiles in the poset. The identification of threshold profiles is a crucial and critical point. Although it is based on objective criteria (e.g. analysis of the literature, opinion of experts, etc.), there is no doubt that this step is strongly pervaded by subjectivity. This could be considered a weakness. However, subjectivity is an unavoidable element in measurement, which, however, does not make it arbitrary, since it always involves a relationship with the reality (Alaimo, 2020). Once the threshold(s) has been identified, a series of mathematical functions can

[5] For a more detailed analysis of the average height, please see Fattore (2017); Alaimo et al. (2020).

be used to describe the satisfaction levels of the profiles in relation to the threshold(s) identified. The so-called *identification function* expresses the number of events in which the profile falls into the area of dissatisfaction, considering the different linear extensions, assigning to each profile a score in [0,1] as follows:

- the scores of the threshold profiles are 1 (they are classified as dissatisfied);
- the scores of profiles below at least one element of the threshold are 1;
- the scores of profiles above any element of the threshold are 0 (they are classified as absolutely satisfied);
- the scores of all other profiles are in [0,1] (they are classified as *fuzzy* satisfied profiles).

In each linear extension, a profile is clearly below at least an element of the threshold or it is above all elements. Thus, it can be reliably classified as satisfied or not. Thus, we can define a function $idn_\delta(\cdot)$, which assigns in each linear extension $\delta$:

$$\begin{cases} 1 & \text{if the profile is classified as dissatisfied in } \delta \\ 0 & \text{if the profile is classified as satisfied in } \delta \end{cases}$$

The count of linear extensions where a profile is classified as dissatisfied makes it possible to quantify such ambiguities and obtain a non-linear identification function *idn*(·) that assigns scores in [0,1] to each profile. The mathematical formalization of this function for a profile $\pi$ of the poset $\Pi$ is the following:

$$idn(\pi) = \frac{1}{|\Omega_\Pi|} \sum_{\delta \in \Omega_\Pi} idn_\delta(\pi) \tag{28}$$

This function gives information about the ambiguity of the set in terms of dissatisfaction. This information can be integrated with that expressing the intensity of such dissatisfaction by means of the so-called *severity function*. Severity is the arithmetic mean of the graphical distance of the profile from the first profile above all threshold ones (its score is 0 for profiles above the threshold). Given a deprived profile $q$ in a linear extension $\delta$ and a profile $s$ nearest to $q$ in $\delta$ as the first profile ranked above all the elements of the threshold, the severity of $q$ in $\delta$ is defined as the graph distance of $q$ from $s$ in the Hasse diagram of $\delta$ (Fattore et al., 2015, 422). Severity is equal to 0 for non-deprived profiles in $\delta$. It is formalized as follows:

$$svr(\pi) = \frac{1}{|\Omega_\Pi|} \sum_{\delta \in \Omega_\Pi} svr_\delta(\pi) \tag{29}$$

As stated by (Fattore, 2016, 845), we can define a relative severity function by using the maximum value (deprivation severity reaches the maximum on the bottom profile) as benchmark:

$$svr_{rel}(\pi) = \frac{1}{|\Omega_\Pi|} \sum_{\delta \in \Omega_\Pi} \frac{svr_\delta(\pi)}{svr_\delta(\pi_{max})} \tag{30}$$

# 3 Inequality indices literature review

According to (Koh, 2020) Inequality refers to the phenomenon of unequal and/or unjust distribution of resources and opportunities among members of a given society. The term inequality may mean different things to different people and in different contexts. Moreover, inequality encompasses distinct yet overlapping economic, social, and spatial dimensions. Measuring inequality is crucial to understanding and assessing every aspect of it and then trying to limit its development and reduce it. However, due to the complex and multidimensional nature of these phenomena, it is non trivial to give a representation that captures every aspect. To this end, a number of composite indicators have been created in recent years to provide synthetic measures of inequality in its entirety, or by focusing on certain aspects of it, such as: gender, economic and social inequality. In this section we review the main inequality indices created by international organisations, exploring their intent, methodology and the data sources they refer to.

## 3.1 Human Development Index

The Human Development Index (Undp, 1997) was developed by United Nations Development Programme (UNDP) and is based on the assumption that there is no coincidence between economic development and human development. Since 1990, UNDP has published the Human Development Report (HDR), the annual report on the development dimension, in which it analyses the level of human development of international countries. Conceptually, HDI is based on Amartya Sen's capability approach (Sen et al., 1999), a theoretical framework that starts from two assumptions: the freedom to achieve well-being is of primary social and moral importance; well-being must be understood in terms of people's capabilities and functions. Capabilities are the actions and objects that people can achieve if they choose, such as being well fed, getting married, being educated and travelling; functionings are capabilities that have actually been realised. This realisation depends crucially on certain personal, socio-political and environmental conditions which, in the capability literature, are called conversion factors. HDI takes into account three fundamental dimensions that allow for the full realisation of an individual:

- The possibility of leading a long and healthy life (Health dimension);
- The possibility of acquiring knowledge (Education dimension);
- The possibility of having access to a level of income that can guarantee a decent standard of living (Income dimension).

The elementary indicators identified to measure each of these dimensions are:

- Life expectancy at birth for the health dimension
- The average number of years of education received and the average number of years of education expected for the education dimension • Gross national product per capita, at purchasing power parity for the income dimension (taken as a logarithm).

### 3.1.1 Methods:

The elementary indicators are standardised using the Min-Max method, which brings the indicators into a range [0,1]; the minimum and maximum values of each indicator are chosen according to the theoretical framework of departure and act as “natural zeros” and ”aspirational targets”. After normalisation, the arithmetic mean of the two standardised indicators is calculated for the Education dimension, thus creating a single dimensional index as for the other two dimensions. HDI is an index that takes values between 0 and 1 (1 maximum development - 0 minimum development) and was first calculated as the arithmetic mean of the three dimensions but after the 2010 methodological review (Kovacevic et al., 2010) as the geometric mean of three dimensional indicators, which are given the same weight, to reduce the *compensability* effect. The HDI for each country is obtained as follow:

$$HDI = \sqrt[3]{I_{Health} * I_{Education} * I_{Income}}$$

Where $I_{Health}$ is the Health dimension, $I_{Education}$ is the education dimension and $I_{Income}$ is the income dimension.

### 3.1.2 Data sources:

The human development data are sourced from international data agencies, such as: Barro-Lee Educational Attainment Dataset, CEDLAS (Center for Distributive, Labor and Social Studies), CRED EM-DAT (Centre for Research on the Epidemiology of Disasters), Eurostat, FAO (Food and Agriculture Organization), Gallup, GCP (Global Carbon Project), Pep Canadell, ICF Macro, Demographic and Health Surveys (DHS), IDMC (Internal Displacement Monitoring Centre), IHME (Institute for Health Metrics and Evaluation), ILO (International Labour Organization), IMF (International Monetary Fund), IPU (Inter-Parliamentary Union), ITU (International Telecommunication Union), LIS (Luxembourg Income Study), OECD (organization for Economic Co-operation and Development), UNCTAD (United Nations Conference on Trade and Development), UNDESA (United Nations Department of Economic and Social Affairs), UNECLAC (United Nations Economic Commission for Latin America and the Caribbean), UNESCO (United Nations Educational, Scientific and Cultural Organization) Institute for Statistics, UNESCWA (United Nations Economic and Social Commission for Western Asia), UNHCR (United Nations High Commissioner for Refugees), UNICEF (United Nations Children’s Fund), UN Inter-agency Group for Child Mortality Estimation (UN IGME), UN Maternal Mortality Estimation Group (World Health Organization, United Nations Children's Fund, United Nations Population Fund and World Bank), UNODC (United Nations Office on Drugs and Crime), UNOHCHR (United Nations Office of the High Commissioner for Human Rights), UNSD (United Nations Statistics Division), UN Women, UNRWA (United Nations Relief and Works Agency for Palestine), WHO (World Health Organization), World Bank and World Inequality Lab.

## 3.2 Gender Development Index (GDI)

The Gender Development Index (GDI) (UNDP, 2021) developed by UNPD since 1995, is directly linked to Human Development Index, in fact it is given by the ratio of the HDI calculated separately according to gender and represents the HDI for the female gender as a percentage of the HDI for the male gender. As HDI, GDI is conceptually based on Sen's capability approach and takes into account the same dimensions and the same indicators of HDI.

### 3.2.1 Methods:

The estimate of income produced is obtained by calculating the wage share for each gender. The female wage share is obtained as follow:

$$S_f = \frac{\frac{W_f}{W_m} * EA_f}{\frac{W_f}{W_m} * (EA_f + EA_m)}$$

Where $\frac{W_f}{W_m}$ is the ratio of female to male wages, $EA_f$ represents the female share of the economically active population, while $EA_m$ is the male share. The male wage share is given by $S_m = 1 - S_f$. The estimated women's per capita income ($GNI_{pcf}$) is obtained from the gross national income per capita ($GNI_{pc}$), first multiplied by the female wage share, $S_f$, and then divided by the female share of the population, $P_f = \frac{N_f}{N}$:

$$GNI_{pc_f} = GNI_{pc} * \frac{S_f}{P_f}$$

Equally to the HDI the indicators are normalized using the Min-Max method, the Education dimension is achieved by aggregating the two indicators that belong to it using the arithmetic mean and for the calculation of the Income dimension, the natural logarithm of the individual values is used to account for a marginally decreasing effect of higher income values. The values of the female and male HDI indices are given by the geometric mean of the three dimensions for each gender:

$$HDI_f = \sqrt[3]{I_{S_f} * I_{I_f} * I_{R_f}}$$

$$HDI_m = \sqrt[3]{I_{S_m} * I_{I_m} * I_{R_m}}$$

Then the GDI is expressed as the ratio between the HDI of the female gender and the HDI of the male gender:

$$GDI = \frac{HDI_f}{HDI_m}$$

### 3.2.2 Data sources:

As part of the Human Development Reports, the data sources of the GDI are part of the same list of the HDI, listed in Subsection 3.1.2.

## 3.3 Global Multidimensional Poverty Index (MPI)

The Global Multidimensional Poverty Index (MPI) (UNDP, 2021) with the methodological note published in (Alkire et al., 2021) is an index produced by the United Nations Development Programme (UNDP) in collaboration with the OPHI - Oxford Poverty and Human Development Initiative to provide a synthetic measure of poverty that takes into account various forms of deprivation experienced by people in their daily lives, including poor health, inadequate education and low living standards. The report, published annually, currently examines the level and composition of multidimensional poverty in 109 countries covering 5.9 billion people. The MPI is based on the same dimensions considered for HDI, but the indicators selected are different:

- For the health dimension, two indicators are taken into account:
  - A Nutrition index
  - A child mortality index
- For the education dimension, two indicators are used:
  - Years of schooling
  - An indicator of school attendance
- For the income dimension, 6 indicators are included:
  - Cooking fuel
  - Sanitation
  - Drinking water
  - Electricity
  - Housing
  - Assets

### 3.3.1 Methods:

In the global MPI, a person is identified as multidimensionally poor or MPI poor if it is deprived in at least one-third of the weighted MPI indicators. In other words, a person is MPI poor if the person's weighted deprivation score is equal to or higher than the poverty cutoff of 33.33%. After the poverty identification step, they aggregate across individuals to obtain the incidence of poverty or headcount ratio (H) which represents the proportion of poor people. They then compute the intensity of poverty (A), representing the average number of weighted deprivation experienced by the poor. They then compute the adjusted poverty headcount ratio ($M_0$) or MPI by combining H and A in a multiplicative form ($MPI = H * A$).

### 3.3.2 Data sources:

The data sources they refer to are: ICF Macro Demographic and Health Surveys, United Nations Children's Fund Multiple Indicator Cluster Surveys and for several countries national household surveys with the same or similar content and questionnaires.

## 3.4 Genuine Progress Indicator (GPI)

The Genuine Progress Indicator (GPI) (Cobb et al., 1995) was created in 1995 by the US organisation Redefining Progress with the aim of redefine progress by developing an economic indicator that tries to come as close as possible to the economic reality that people experience; in fact, GPI is a metric used to measure a country's economic growth. GPI takes into account everything that GDP uses, but adds other indicators that represent the cost of negative effects related to economic activity, such as the cost of crime, the cost of ozone depletion and the cost of resource depletion. It attempts to measure whether the environmental impact and social costs of economic production and consumption in a country are negative or positive factors in overall health and well-being. The Genuine Progress Indicator consists of 26 separate time series data columns spanning the 1950-2004 period, with a two year time lag, which take into account economic, social and environmental aspects:

- Social dimension
  - Automobile accidents
  - Commuting
  - Services of highways and streets
  - Lost leisure time
  - Volunteer work
  - Personal pollution abatement
  - Crime
  - Family breakdown
  - Domestic labor
- Environmental dimension
  - Non-renewable resource depletion
  - Ozone depletion
  - Climate change
  - Net loss of forest cover
  - Net loss of farmland
  - Net loss of wetlands
  - Noise pollution
  - Air pollution
  - Water pollution

- Economic dimension
  - Personal consumption
  - Income inequality
  - Adjusted personal consumption
  - consumer durables services
  - Consumer durables costs
  - Underemployment
  - Net capital investment

The countries covered by GPI are: Australia, Austria, Belgium, Chile, China, Germany, India, Italy, Japan, Netherlands, New Zealand, Poland, Sweden, Thailand, United Kingdom, United States and Vietnam

### 3.4.1 Methods:

The GPI formula is:

$$GPI = C_{adj} + G + W - D - S - E - N$$

Where: $C_{adj}$ = personal consumption with income distribution adjustments; $G$ = capital growth; $W$ = unconventional contributions to welfare, such as volunteerism; $D$ = defensive private spending, $S$ = activities that negatively impact social capital; $E$ = costs associated with the deterioration of the environment; $N$ = activities that negatively impact natural capital.

### 3.4.2 Data sources:

The main data sources of the indicators are: National Income and Product Accounts (NIPA) tables published by the Bureau of Economic Analysis, U.S. Census Bureau, Michigan Survey Research Center, Bureau of Labor Statistics (BLS) American Time Use Surveys (ATUS), Bureau of Justice Statistics National Crime Survey, Statistical Abstract of the United States, BEA's National Income and Product Accounts, World Health Organization, National Agricultural Statistics Service, USDA's National Agricultural Lands Study, Farm Information Center and Oak Ridge National Laboratory.

## 3.5 Better Life Index (BLI)

The Better Life Index (BLI) (OECD, 2021) first released by the OECD (Organization for Economic Co-operation and Development) in 2011, with the aim of providing a measure of the quality of life for OECD countries. It is an interactive index that takes into account 11 thermometers, covering as many areas of socioeconomic interest, divided into internal indicators. The eleven thermometers (dimensions) considered are:

- Housing (number of rooms per person, presence of basic sanitary facilities, expenditure incurred in maintaining the home)
- Income (net wealth and disposable income per household)

- Employment (job security, income from employment, employment rate and long-term unemployment rate)
- Social relations (quality of social networks)
- Education (duration, level and skills of students)
- Environment (water quality and air pollution)
- Civic engagement (voter turnout and stakeholder participation in the legislative process)
- Health (perceived health status and life expectancy)
- Satisfaction (perceived level of happiness and satisfaction)
- Safety (number of homicides and perceived level of peace of mind when walking alone at night)
- Work-life balance (number of employees with a working week of more than 50 hours and daily minutes of leisure and personal care).

### 3.5.1 Methods:

For each country tested, all the necessary data is collected and a score is obtained for each theme, ranging from 1 to 10. Graphically, the result achieved by each country is represented by a flower with 11 petals: each petal is larger the higher the score achieved by the country in that area. The synthetic index is the graphical representation obtained by taking into account for the different countries the different levels reached in the eleven thermometers considered.

### 3.5.2 Data sources:

The data sources used for the BLI are: European Union Statistics on Income and Living Conditions (EU-SILC), National Statistical Offices, OECD National Accounts Database, OECD Wealth Distribution, OECD Employment and Labour Market Statistics, OECD Average annual wages, Gallup World Poll, OECD Education at a glance, PISA at a glance, OECD Exposure to air pollution, OECD Indicators of Regulatory Policy and Governance (iREG), International Institute for Democracy and Electoral Assistance (IDEA), OECD Health Statistics and Time Use Surveys.

## 3.6 Equitable and Sustainable well-being (Bes)

In 2010, the Italian National Istitute of Statistic (ISTAT) launched a joint initiative with the Consiglio Nazionale dell'Economia e del Lavoro (CNEL) for the measurement of Equitable and Sustainable Well-being (BES) in Italy (ISTAT, 2020). The BES project was created with the aim of providing a conceptual and methodological framework for assessing the progress of a society not only from an economic, but also from a social and environmental perspective. Since 2013, ISTAT has been publishing annually the Report on Fair and Sustainable Welfare.

The BES is the result of a participatory process involving trade associations, trade unions, representatives of the third sector and academic experts, which started with a consultation on the importance of the dimensions of wellbeing carried out in February 2011 as part of the Istat multi-purpose survey "Aspects of daily life". Respondents were asked to rate the importance of some dimensions of well-being by giving a score from 0 to 10 to a list of fifteen conditions. The results of the consultation, together with a survey of international experiences, formed the basis for the definition of the reference framework for the measurement of wellbeing, based on twelve domains, i.e. the areas in which wellbeing is to be measured, and 130 indicators, which take into account both aspects that have a direct impact on human and environmental wellbeing and those that measure the elements that are functional to improving the wellbeing of the community and its surrounding environment. The domains identified are:

- Health
- Education and training
- Economic well-being
- Work and life-time balance
- Politics and institutions
- Social relations
- Security and safety
- The environment
- Subjective well-being
- Research and innovation
- Landscape and cultural heritage
- Quality of services

For each domain, ISTAT calculates a synthetic index considering the selected elementary indicators, available at the level of regional territorial disaggregation.

### 3.6.1 Methods:

The methodology used to aggregate the elementary indicators in each dimension is the AMPI method and the normalization process is the variant of Min-Max related to the AMPI explained in the previous section.

### 3.6.2 Data sources:

The data sources used for the Bes are: Istat, Italian Ministry of Education, University and Research, National learning surveys (Invalsi), Bank of Italy, Italian Ministry of the Interior, Individual regional councils, National Commission for Companies and the Stock Exchange (CONSOB), Italian Ministry of Justice, Higher Institute for Environmental Protection and Research (ISPRA), Terna S.p.A. and European Union's Earth Observation Programme (Copernicus).

## 3.7 Social Progress Index (SPI)

The Social Progress Index (SPI) (Imperative, 2021), first released in (Porter et al., 2014) is made by the global nonprofit Social Progress Imperative with the aim to inform on the social and environmental health of societies and helping them prioritize actions that accelerate social progress. Trough 53 social and environmental indicators, the SPI measures how well a society provides its people through 3 dimensions, each one composed by 4 components:

- Basic human needs
  - Nutrition and basic medical care
  - Water and sanitation
  - Shelter
  - Personal safety
- Foundations of wellbeing
  - Access to basic knowledge
  - Access to information and communications
  - Health and wellness
  - Environmental quality
- Opportunity
  - Personal rights
  - Personal freedom and choice
  - Inclusiveness
  - Access to advanced education

The SPI ranks 168 countries.

### 3.7.1 Methods:

The indicators are normalized using the Z-score method and to calculate component scores, the set of indicators within each component are aggregated into a factor using Principal Component Analysis (PCA), that is a data simplification technique used in multivariate statistics. PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. In the calculation of the SPI each principal component is than converted into a component score on a scale of 0 to 100 using the Min-Max formula times 100. Each dimension is the arithmetic average of the four components that make up that dimension and the overall Social Progress Index score is calculated as the arithmetic average of the three dimensions. In establishing country rankings for overall performance, country scores are divided into six tiers based on hierarchical clustering.

### 3.7.2 Data sources:

The sources of the data used comes from several institutions such as: UNDP, Institute for Health Metrics and Evaluation, World Bank, International Energy Agency, UN Educational, Scientific, Cultural Organization Institute for Statistics, International Labor Organization and Gallup World Poll.

## 3.8 Gender Social Norms Index (GSNI)

The Gender Social Norms Index (GSNI) (UNDP, 2020) was introduced in the 2019 Human Development Report by United Nations Development Programme (UNDP) starting from the assumption that gender disparities are a persistent form of inequality in every country. Despite remarkable progress in some areas, no country in the world, rich or poor, has achieved gender equality. All too often, women and girls are discriminated against in health, in education, at home and in the labour market, with negative repercussions for their freedoms. GSNI measures how social beliefs obstruct gender equality in areas like politics, work, and education, and contains data from 75 countries, covering over 80 percent of the world's population. The GSNI dimensions are:

- Political empowerment
- Educational empowerment
- Economic empowerment
- Physical integrity

GSNI is constructed based on responses to seven questions from the World Values Survey, which are used to create seven indicators:

- Men make better political leaders then women do
- Women have the same rights as men
- University is more important for a men than for a women
- Men should have more right to a job than women
- Men make better business executives than women do
- Proxy for intimate partner violence
- Proxy for reproductive rights

### 3.8.1 Methods:

The answer choices to the questionnaire vary by indicator. For indicators for which the answer choices are strongly agree, agree, disagree and strongly disagree, the index defines individuals with a bias as those who answer strongly agree and agree. For the political indicator on women's rights, for which the answer is given on a numerical scale from 1 to 10, the index defines individuals with a bias as those who choose a rating of 7 or lower. For the physical integrity indicators, for which the answer also ranges from 1 to 10, the index defines individuals with a bias using a proxy variable for intimate partner violence and one for reproductive rights. For each indicator a variable takes the value of 1 when an individual has a bias and 0 when the individual does not. Two methods of aggregation are then used in reporting results in the form of an Index. The core gender social norms index (GSNI) is based on the "union approach." It measures the percentage of people with bias(es), independent of the number of biases. In many instances, it might take only one bias from one person to block a woman's progress in society. A second gender social norms index (GSNI2) is based on a simple "intersection approach." It measures the percentage of people with at least two biases.

### 3.8.2 Data sources:

The GSNI is based on data from the World Values Survey.

## 3.9 Global Gender Gap Index (GGGI)

The Global Gender Gap Index (GGGI) (WEF, 2021) was introduced by the World Economic Forum (WEF) in 2006 as a tool to measure the extent of gender inequality and track its evolution over time. The indicator is produced at country level for 149 countries and examines the gap between men and women through 14 variables organised into four key categories (sub-indicators):

- Economic participation and opportunity
- Educational attainment
- Health and survival
- Political emancipation

### 3.9.1 Methods:

The GGGI is constructed through a four-step process, described below:

1. *Conversion into ratios*. Initially, all data are converted into ”female to male” ratios.

2. *Truncation of data to the parity benchmark*. The ratios obtained are truncated at the “equality benchmark”. For all indicators (except for the two health indicators: in the case of the sex ratio at birth, the equality benchmark is set at 0.9445, and in the case of life expectancy, the equality benchmark is set at 1.06), this equality benchmark is taken as 1, i.e. an equal value between women and men.

3. *Calculation of sub-indicator scores*. The weighted arithmetic mean of the indicators within each sub-indicator is calculated to derive the corresponding summary scores. First, the sub-indicator scores are normalised to equalise their standard deviations (z-score standardisation). Subsequently, the scores of each sub-indicator are aggregated into a single value by means of a weighted average whose weights are determined by the ratio of 0.01 to the standard deviation of each indicator. This determines how much the indicator has to vary in relation to its standard deviation to result in a one percentage point change in the indicator. These four values are then expressed as weights which add up to one to calculate the weighted average of the four indicators.

4. *Calculation of final scores*. For all sub-indices, the highest possible score is 1 (perfect gender equality) and the lowest possible score is 0 (maximum inequality), thus linking the scores between inequality and reference equality. To calculate the GGGI, a simple arithmetic average of the scores each nation obtains in the different sub-indicators is used. This final value also varies between 1 and 0, thus making it possible to compare ideal standards of equality, as well as the relative rankings of countries.

### 3.9.2 Data sources:

The GGGI indicator’s data sources are: International Labour Organization (ILO), World Economic Forum, International Monetary Fund (IMF), World Bank, UNESCO, Inter-parliamentary Union, World Health Organization (WHO), United Nations, OECD, World Bank Enterprise Survey, Quotaproject.org and UNICEF.

## 3.10 Gender Equality Index (GEI)

The Gender Equality Index (GEI) (EIGE, 2021) is an indicator produced by the European Institute for Gender Equality (EIGE) and presented to the public in 2013. The index measures progress in gender equality, relative to the EU policy context. In particular, the GEI measures how far the EU and its member states have come in achieving gender equality. There are eight dimensions referred to in the GEI but only six are used to construct a summary indicator of gender equality.

- Work
- Money
- Knowledge
- Time
- Power
- Health

In addition to these, the Violence domain measures gender-based violence against women, while the Intersectional Inequalities domain studies gender inequality within specific population groups (people with disabilities, migrants, etc.). The GEI is composed of 31 indicators, divided into 14 sub-dimensions representing the six main dimensions.

### 3.10.1 Methods:

To *assign a weight to each variable, dimension and sub-dimension* 4 different methodologies were applied in order to choose the most appropriate one according to the unit of analysis. The four methods included weighting by means of equivalent weights, a variant of equivalent weights for the variable segregation in work, weights derived from Principal Component Analysis (PCA), and finally weights representing the judgement expressed by experts, members of the working group and the EIGE forum. The latter method is also referred to as the Analytic Hierarchy Process (AHP), i.e. a method for organizing and analyzing complex decisions that provides a rational framework for a needed decision by quantifying its criteria and alternative options, and for relating those elements to the overall goal. It starts by comparing the importance of criteria, two at a time, through pair-wise comparisons and converting the evaluations into numbers, which can be compared to all of the possible criteria. In the final step of the process, numerical priorities are calculated for each of the alternative options. These numbers represent the most desired solutions. In the third edition of the index, the robustness analysis confirmed the application of equivalent weights for the variables and sub-dimensions and weights derived from expert judgement for the dimensions. Specifically, to weight the dimensions, the experts were asked to compare the dimensions in pairs and to assign each a score between 1 (equal importance of the dimensions) and 9 (most important dimension). Finally, the average of the relative weights assigned by the experts was calculated in order to obtain an overall score for each domain. To *aggregate* variables to be grouped in order to create indices at the sub-dimension, dimension and overall GEI level, the arithmetic mean was used for the aggregation of variables, while the geometric mean was used for the aggregation of dimensions and sub-dimensions. The initial metric used to calculate the Gender Equality Index does not require any *normalisation*, since it adapts to the unit of measurement and corrects the range of variation of each variable by limiting it between [0,1]. Furthermore, it allows each variable considered to be interpreted in terms of its distance from the point of equality, set at 1, and allows variables to be compared within each country. To *proceed with the calculation of the Gender Equality Index*, all the variables within each sub-dimension are aggregated to create sub-dimensional indicators. These are then aggregated at the dimension level. Finally, all dimensional indicators are aggregated to construct the Gender Equality Index. The final metrics of the GEI are as follows:

$$I_i^t = \prod_{(d=1)}^{6} \left\{ \prod_{(s=1)}^{ns_d} \left[ \sum_{v=1}^{n_s} \frac{\Gamma(X_{itv})}{n_s} \right]^{\frac{1}{ns_d}} \right\}^{W_{AHP_d}}$$

With $i$ = 1,...,28 EU countries, $d$ = 1,...,6 dimensions, $s$ = 1,...,14 sub-dimensions, $v$ = 1,...,31 indicators, $n_s$ =number of indicators in the subdimension s, $ns_d$ =number of sub-dimensions in the dimension d and $W_{AHPd} \in [0,1]$. Where $I_i^t$ corresponds to the best Gender Equality Index for the i-th country during the period $t$, $\Gamma(X_{itv})$ is the metric initially considered expressed at the level of variable $v$, while the term $W_{AHPd}$ identifies the weights derived from the judgment expressed by peers at the dimension level.

### 3.10.2 Data sources:

The data sources of the indicators used for the calculation of the GEI are: Eurostat, Eurofound and EIGE Gender Statistics Database.

# 4 Cultural and Linguistic Diversity Indices

According to UNESCO "Universal Declaration on Cultural Diversity" (Torres, 2002), *each individual must acknowledge not only otherness in all its forms but also the plurality of his or her own identity, within societies that are themselves plural. Only in this way can cultural diversity be preserved as an adaptive process and as a capacity for expression, creation and innovation*. Primary importance in this context is attached to linguistic diversity. Languages, with their complex implications for identity, communication, social integration, education and development, are of strategic importance for people and the planet. There is growing awareness that languages play a vital role in development, not only in ensuring cultural diversity and intercultural dialogue, but also in attaining quality education for all and strengthening cooperation, in building inclusive knowledge societies and preserving cultural heritage, and in mobilizing political will for applying the benefits of science and technology to sustainable development. For these reasons in the last decades there has been a growing interest among researchers in measuring these phenomena. In this section we analyse some of the best known cultural and diversities indices in literature.

## 4.1 The Multiculturalism Policy Index (MCP)

The Multiculturalism Policy Index (MCP) first released in (Banting et al., 2006), is a scholarly research project that monitors the evolution of multiculturalism policies in 21 Western democracies. The MCP is designed to provide information about multiculturalism policies in a standardized format that aids comparative research and contributes to the understanding of state-minority relations. The project provides an index at three points in time: 1980, 2000, 2010 and for three types of minorities:

- one index relating to immigrant groups
- one relating to historic national minorities
- one index relating to indigenous peoples

The MCP Index for Immigrant Minorities examine the adoption of the following eight policies:

1. Constitutional, legislative or parliamentary affirmation of multiculturalism 2.

The adoption of multiculturalism in school curriculum

3. The inclusion of ethnic representation/sensitivity in the mandate of public media or media licensing
4. Exemptions from dress-codes, Sunday-closing legislation etc.
5. Allowing dual citizenship
6. The funding of ethnic group organizations to support cultural activities
7. The funding of bilingual education or mother-tongue instruction
8. Affirmative action for disadvantaged immigrant groups.

The MCP Index for immigrant minorities is available on annual basis, with scores for each MCP policy in each country from 1960 to 2020. The MCP Index for Indigenous Peoples examines the adoption of the following nine policies:

1. Recognition of land rights/title
2. Recognition of self-government rights
3. Upholding historic treaties and/or signing new treaties
4. Recognition of cultural rights (language; hunting/fishing)
5. Recognition of customary law
6. Guarantees of representation/consultation in the central government
7. Constitutional or legislative affirmation of the distinct status of indigenous peoples
8. Support/ratification for international instruments on indigenous rights
9. Affirmative action

The MCP Index for National Minorities examines the adoption of the following six policies:

1. Federal or quasi-federal territorial autonomy
2. Official language status, either in the region or nationally
3. Guarantees of representation in the central government or on constitutional courts
4. Public funding of minority language universities/schools/media
5. Constitutional or parliamentary affirmation of "multinationalism"
6. According international personality (eg., allowing the substate region to sit on international bodies)

### 4.1.1 Methods:

For each indicator, a qualitative assessment is provided along with the relevant evidence. A response of "yes" indicates that the country has met or exceeded the standard outlined in the indicator, a response of "no" indicates that the country has not met this indicator, while a response of "partial" indicates that the country has made some progress in this area.

### 4.1.2 Data sources:

For each indicator, policy documents, program guidelines, legislation, and government news releases were examined to assess the extent to which a country has met or exceeded the standard outlined in the indicator. Where official government documents were not available, secondary sources and other academic research have been used.

## 4.2 Migrant Integration Policy Index (MIPEX)

The Migrant Integration Policy Index (MIPEX) (Solano and Huddleston, 2020) was released in (Niessen et al., 2007) to evaluate and compare what governments are doing to promote the integration of migrants. The project informs and engages key policy actors about how to use indicators to improve integration governance and policy effectiveness. For this purpose, it identifies and measures integration policies and the links between the latter, outcomes and public opinion, drawing on international scientific studies. The aim of MIPEX is to measure policies that promote integration in both social and civic terms. The MIPEX includes 52 countries and collects data from 2007 to 2020, in order to provide a view of integration policies across a broad range of differing environments. It considers a system of 58 indicators, covering 8 policy areas that have been designed to benchmark current laws and policies against the highest standards through consultations with top scholars and institutions [6] using and conducting

[6] The highest standards are drawn from Council of Europe Conventions, European Union Directives and international conventions

comparative research in their area of expertise. The policy areas of integration covered by the MIPEX are:

- Labour Market Mobility
- Family Reunion
- Education
- Political Participation
- Long-term Residence
- Access to Nationality
- Anti-discrimination
- Health[7]

### 4.2.1 Methods:

For each area, a synthetic measure (dimensional) is calculated as an arithmetic mean of the elementary indicators, i.e. those selected for measuring each policy area. Each dimensional synthetic indicator is bounded between [0,100], in which the maximum of 100 is awarded when policies meet the highest standards for equal treatment.

### 4.2.2 Data sources:

The values of the elementary indicators of the MIPEX are chosen by experts from each country, by means of a questionnaire.

## 4.3 Worldwide Language Index

The Worldwide Language Index (WLI) (Preply, 2021) developed by Preply (a language learning app and e-learning platform), provides a detailed analysis of the countries in the EU, the U.S, and Canada, that provide the best environment for language learning. They analyse 18 factors split across seven dimensions, including:

- Official languages
- Foreign language learning
- Language learning at school
- Level of command of best-known foreign language
- Access to language learning through technology
- Subtitles, dubbing and voiceover
- Language diversity

### 4.3.1 Methods:

Each of the dimensions in the study was weighted equally (14.3%) as each dimension was deemed to contribute equally to an optimum environment for language learning. The final results were calculated by a process of normalization using the Min-Max method. The results were normalized by calculating the results for each dimension on a scale from either 0 to 25, 0 to 50, or 0 to 100, (depending

[7] Health data are only available for years 2014 and 2019

on the weighting of the dimension) and then calculating the overall score for each dimension. The results for each of the seven dimensions were then averaged using arithmetic mean to obtain a final overall score for each country. Finally the scores for each country were ranked from highest to lowest to reveal the best and worst countries for language learning.

#### 4.3.2 Data sources:

The values of the elementary indicators of the WLI comes from government website of each respective country.

## 4.4 Ljubljana Guidelines on Integration of Diverse Societies

The Ljubljana Guidelines on Integration of Diverse Societies (HCNM, 2012) published by Organization for Security and Co-operation in Europe (OSCE) are a set of High Commissioner on National Minorities (HCNM) Guidelines on integration of diverse societies. HCNM Guidelines seek to provide guidance to OSCE participating States on how best to integrate diverse societies. They cover structural principles without which good integration policies are difficult to conceive as feasible, in particular these Guidelines consist of four parts:

- Structural principles
- Principles for integration
- Elements of an integration policy framework
- Key policy

*The structural principles* are interlinked and necessary although never fully achievable. They are goals towards which all States should be aiming; *the principles for integration* provide basic theses and values that relate more specifically to integration or are necessary for formulating integration policy; *elements of an integration policy framework* sets out the framework for the elaboration and implementation of integration policies, including mechanisms, processes and cross-cutting themes; finally, *the key policy areas*, although not an exhaustive list of relevant policy areas, deals with the main thematic considerations and includes more specific examples within those policy areas, which have to be selected and adapted to each specific context.

## 4.5 European Index of Multilingual Policies and Practices

The European Index of Multilingual Policies and Practices (Extra and Yagmur, 2012) is part of the Language Rich Europe (LRE) project, carried out by a consortium of over 30 acknowledged institutions in Europe under the leadership of the British Council and co-financed by the European Commission. Babylon, Centre for Studies of the Multicultural Society at Tilburg University, has led on the research element of the project, developing draft indicators based on European Union (EU) and Council of Europe (CoE) resolutions, conventions and recommendations to examine language policies and practices in 25 countries and regions, constructing and administering the research questionnaire among their partner network, processing and analysing the data, and writing up the cross-national outcomes of data collection. The research partners in each country/region have complemented the data collected with their own analysis of the findings, supported by examples of good practice and promising initiatives. The overall objectives of the LRE project are:

- To facilitate the exchange of good practice in promoting intercultural dialogue and social inclusion through language teaching and learning •
- To promote European co-operation in developing language policies and practices across several education sectors and broader society •
- To raise awareness of the EU and CoE recommendations for promoting language learning and linguistic diversity across Europe.

Eight language domains are used by the LRE survey that are covered by a total of 260 questions of the questionnaire. The eight domains are:

- Languages in official documents and databases
- Languages in pre-primary education
- Languages in primary education
- Languages in secondary education
- Languages in further and higher education
- Languages in audiovisual media and press public spaces
- Languages in public services and press
- Languages in business

### 4.5.1 Methods:

The LRE is a survey study and includes 25 national and regional profiles. National profiles are provided for 15 countries, namely 12 European Union (EU) countries plus Switzerland, Bosnia and Herzegovina and Ukraine. Regional profiles are provided for four EU countries (the Netherlands, Spain, UK and Germany). Each profile provides both qualitative and quantitative data, and contains information on the national/regional context, on the eight language domains, on key findings overall, and on promising initiatives and/or pilots. The purpose is to help readers to interpret the national/regional profiles. The national/regional profiles are a combination of survey results and a commentary on these, written by the country/regional researcher.

### 4.5.2 Data sources:

The LRE is a survey study and the data are collected by means of a questionnaire. The collection of the primary data took place in cities of each country or region prompted.

## 4.6 Intercultural Cities (ICC) Index

The Intercultural Cities (ICC) Index (Europe, 2019) is part of the Intercultural Cities Programme developed by Council Of Europe, to support local authorities to design and implement inclusive integration policies. The programme is based on the "Intercultural integration policy model" which focuses on enabling communities, organisations and businesses to manage the diversity of people in a way which ensures the equal value of all identities, cohesion and competitive advantage. The Intercultural Cities programme is now being implemented by over 130 cities in Europe and beyond. The ICC Index is monitoring the efforts cities make to encourage participation, interaction, equality of opportunities and the mainstreaming of interculturalism and diversity advantage principles. Based on this, the Council of Europe sends back an analytical report with recommendations and examples of good practice from other cities. In a second step, an expert visit takes place with independent experts and a Council of Europe representative that will involves city officials and a wide range of local stakeholders, to review their policies through an intercultural lens. Local stakeholders are then guided through the development (or revision) of a comprehensive intercultural strategy to manage diversity positively and realise the diversity advantage. The ICC Index analysis is based on the answers to 83 questions grouped in 12 indices:

- Commitment
- Media and communication
- Intercultural lens
- International outlook
- Mediation and conflict resolution
- Intercultural intelligence and Language competence
- Welcoming newcomers
- Participation
- Leadership and citizenship
- Interaction
- Antidiscrimination

### 4.6.1 Methods:

Once the questionnaire is filled in satisfactorily, the data are verified and processed by BAK Economics, a Swiss research institute specialised measuring the effectiveness of regional and local policies. Questions are weighed according to their relative importance. For each index or sub-index, participating cities can attain a maximum of 100 points. The data is also analysed from a policy perspective and compiled into a report by experts from the Council of Europe. The report includes:

- The results of the city in the different governance/policy areas
- Charts that illustrate visually the scores attained by the city for each index and how they compare to the city average or to a cluster of cities sharing similar characteristics
- Information on the city's good practices that could inspire other cities • Recommendations based on examples of good practice from other cities that the responding city may consider to increase its score in one or several governance/policy areas.

The quantified data is also included into interactive Intercultural Cities Index Interactive chart.

#### 4.6.2 Data sources:

Data is collected through a questionnaire consisting of 90 questions on:

- The local setting and demographic context
- Intercultural policies, structures and actions
- Governance/policy areas which contribute to intercultural integration
- Additional information the responding city may like to provide

### 4.7 Canadian Index for Measuring Integration (CIMI)

The Canadian Index for Measuring Integration (CIMI) (for Measuring Integration, 2016), is an interactive tool that allows to measure the outcomes of immigrants in Canadian regions. It is a data-driven Canadian index that examines four dimensions of immigrant integration in Canada to assess the gaps between immigrants and the Canadian-born population. The dimensions are:

- Economic
- Civic and democratic participation
- Social
- Health

The CIMI identifies factors that underline successful immigrant integration, assesses changes and trends over time (currently from 1991 to 2020), enables detailed examination of four the dimensions of integration and provides rankings based on empirical evidence for Canadian geographies.

### 4.7.1 Methods:

The indicators are normalized using the Min-Max method and the overall rankings for the CIMI are based on the following weights: Economic Dimension 0,4, Social Dimension: 0,3, Civic Democratic Participation Dimension: 0,2 and Health Dimension 0,1. The weighting system was developed by CIMI researchers. To examine several integration-related outcomes while adjusting for socio-demographic differences between immigrants and Canadian-born population, allowing for equal comparisons between geographies and across time, they uses multiple regression analyses, including both Linear Regression for continuous dependent variables and Logistic Regression for binary (0,1) dependent variables. The reason they choose of using multivariate regression analysis is that it allows us to estimate the main effects of key independent variables (i.e., immigrant status, geography, and immigrant status x geography) on the dependent outcomes while holding constant several demographic factors (e.g., age, sex, ethnicity, language) and socioeconomic factors (income, occupation, education).

### 4.7.2 Data sources:

The CIMI uses three primary data sources in the analysis of economic, social, civic and democratic and health outcomes for immigrants and Canadian-born populations, which include the Census, Canadian Community Health Survey (CCHS), and General Social Survey (GSS).

## 4.8 Indices of Social Development (ISD)

The Indices of Social Development (ISD) (of Social Studies, 2021) has been developed by The International Institute of Social Studies (ISS) to measure social development across countries, as well as the links between social development and other development outcomes. The ISD brings together 184 indicators, synthesising them into a usable set of measures to track how different societies perform along six dimensions of social development:

- Civic activism
- Intergroup cohesion
- Clubs and associations
- Interpersonal safety and trust
- Gender equality
- Inclusion of minorities

The indices are composed from 21 reputable data sources for 195 countries, over the period from 1990 to 2015, and are updated as new data become available. The ISD allow estimating the effects of social development for a large range of countries on indicators like economic growth, human development, and governance.

### 4.8.1 Methods:

The indices are normalized using the Z-score method and aggregated using the method of ”matching percentiles”. In this approach, scores are assigned to countries based on ordinal rankings. The ranks of countries for variables included in the index are used to assign equivalent values to countries with equivalent ranks. This method uses a recursive process of matching observational ranks over pairs of variables: a master and an input variable. The initial master variable is a random variable and the input variables are the dimensions listed above. Taking each of the input variables in turn, the algorithm first determines which observations appear in both the master and input variables. Observations for this conjoint set are then ranked separately for the master and the input variables. Having obtained master and input variable ranks for each observation, they next create a match variable which rescales the input variable by assigning the cardinal value of the country in the master variable to the country with the same ordinal rank in the input variable. For example, if Albania, Burundi, Cameroon, and Denmark were to have master variable scores of 0,45, 0,61, 0,65 and 0,89, and input variable scores of 0,82, 0,94, 0,31 and 0,46, then they would receive match scores of 0,65, 0,89, 0,45 and 0,61. Each observation which has a value in the input variable will receive a matched value. This matching is done against the master variable for each of the K input variables (sub-indices) used in creating the index. Once the match values are assigned for each of the input variables, the K match variables are averaged to create the index score for each country. As the indexing process is obviously influenced by the draw of the random normal master variable (scaled to be roughly bounded between 0 and 1), the newly created index score is fed back through the indexing process as a new master variable. This process continues recursively until the index reaches convergence. The convergence parameter that they choose is $10^{-4}$ for the sum of the squared differences between the master variable and resulting index within a particular iteration.

### 4.8.2 Data sources:

The ISD data sources are: Afrobarometer, Asian Barometer, Cingranelli-Richards (CIRI), Human Rights Data Project, Civicus, Cross-National Time-Series Data Archive, Demographic and Health Surveys, Economist Intelligence Unit, European Social Survey, John Hopkins Comparative Nonprofit Sector Project, International Country Risk Guide, International Crime Victims Survey, International Labour Organisation, International Social Survey, International Telecommunication Union, Latinobarometer, Minorities at Risk, Nature Conservancy, OECD Factbook, UNESCO, World Development Indicators and World Values Survey.

# References


[1] M. I. Aguirre-Urreta, M. R¨onkk¨o, and G. M. Marakas. Omission of Causal Indicators: Consequences and Implications for Measurement. *Measurement: Interdisciplinary Research and Perspectives*, 14(3):75–97, 2016.

[2] L. S. Alaimo. Sustainable development and national differences: an european cross-national analysis of economic sustainability. *RIEDS-Rivista Italiana di Economia, Demografia e Statistica-Italian Review of Economics, Demography and Statistics*, 72(3):101–123, 2018.

[3] L. S. Alaimo. Complexity of social phenomena: Measurements, analysis, representations and synthesis. *Unpublished doctoral dissertation, University of Rome "La Sapienza", Rome, Italy*, 2020.

[4] L. S. Alaimo. Complexity of Social Phenomena: Measurements, Analysis, Representations and Synthesis. *Unpublished Doctoral Dissertation, University of Rome" La Sapienza", Rome, Italy*, 2020.

[5] L. S. Alaimo and P. Conigliaro. Assessing subjective well-being in wide populations. a posetic approach to micro-data analysis. In R. Bruggemann, L. Carlsen, T. Beycan, C. Suter, and F. Maggino, editors, *Measuring and Understanding Complex Phenomena: Indicators and their Analysis in Different Scientific Fields*, pages 243–263. Cham: Springer International Publishing, 2021.

[6] L. S. Alaimo and F. Maggino. Sustainable development goals indicators at territorial level: Conceptual and methodological issues—the italian perspective. *Social Indicators Research*, 147(2):383–419, 2020.

[7] L. S. Alaimo and G. Nanni. Gender inequality in italy: a territorial analysis. *Rivista Italiana di Economia Demografia e Statistica*, 72(4):41, 2018.

[8] L. S. Alaimo, A. Arcagni, M. Fattore, and F. Maggino. Synthesis of multiindicator system over time: A poset-based approach. *Social Indicators Research*, pages 1–23, 2020. doi: 10.1007/s11205-020-02398-5.

[9] L. S. Alaimo, A. Arcagni, M. Fattore, F. Maggino, and V. Quondamstefano. Measuring Equitable and Sustainable Well-being in Italian Regions. The Non-aggregative Approach. *Social Indicators Research*, pages 1–23, 2020. doi: 10.1007/s11205-020-02388-7.

[10] L. S. Alaimo, A. Ciacci, and E. Ivaldi. Measuring sustainable development by non-aggregative approach. *Social Indicators Research*, 157(1):101–122, 2021.

[11] L. S. Alaimo, E. Ivaldi, S. Landi, and F. Maggino. Measuring and evaluating socio-economic inequality in small areas: An application to the urban units of the municipality of genoa. *Socio-Economic Planning Sciences*, page 101170, 2021.

[12] L. S. Alaimo, F. Maggino, and E. Seri. An Application of Temporal Poset on Human Development Index Data. In C. Perna, N. Salvati, and F. S. Spagnolo, editors, *50th Scientific meeting of the Italian Statistical Society. Book of Short Papers SIS2021*, pages 211–216, 2021.

[13] S. Alkire, U. Kanagaratnam, and N. Suppa. The global multidimensional poverty index (mpi) 2021. 2021.

[14] P. Annoni and R. Bruggemann. Exploring Partial Order of European Countries. *Social Indicators Research*, 92(3):471–487, 2009.

[15] A. Arcagni, E. B. di Belgiojoso, M. Fattore, and S. M. Rimoldi. Multidimensional analysis of Deprivation and Fragility Patterns of Migrants in Lombardy, Using Partially Ordered Sets and Self-organizing Maps. *Social Indicators Research*, 141(2):551–579, 2019.

[16] K. G. Banting, W. Kymlicka, et al. *Multiculturalism and the welfare state: Recognition and redistribution in contemporary democracies*. Oxford University Press on Demand, 2006.

[17] W. Becker, M. Saisana, P. Paruolo, and I. Vandecasteele. Weights and importance in composite indicators: Closing the gap. *Ecological indicators*, 80:12–22, 2017.

[18] E. Bericat. The european gender equality index: Conceptual and analytical issues. *Social Indicators Research*, 108(1):1–28, 2012.

[19] A. Bianchi and S. Biffignandi. Workplace social environment indicator: A comparative analysis of european regions. *Social Indicators Research*, pages 1–20, 2020.

[20] H. M. Blalock. *Causal Inferences in Nonexperimental Research*. N.C.: University of North Carolina Press, 1964.

[21] H. M. Blalock. The Measurement Problem: A Gap between the Languages of Theory and Research. In F. Kerlinger, editor, *Methodology in Social Research*, pages 5–27. New York: McGraw-Hill, 1968.

[22] H. M. J. Blalock. *Conceptualization and Measurement in the Social Sciences*. Beverly Hills: SAGE Publications, 1982.

[23] H. M. Blalock Jr. *Causal inferences in nonexperimental research*. UNC Press Books, 2018.

[24] G. Boccuzzo and M. Gianecchini. Measuring young graduates' job quality through a composite indicator. *Social Indicators Research*, 122(2):453–478, 2015.

[25] K. A. Bollen. Multiple Indicators: Internal consistency or No Necessary relationship? *Quality and Quantity*, 18(4):377–385, 1984.

[26] K. A. Bollen. *Structural Equations with Latent Variables*. New York: Wiley, 1989.

[27] K. A. Bollen. Interpretational confounding is due to misspecification, not to type of indicator: Comment on howell, breivik, and wilcox (2007). 2007.

[28] K. A. Bollen and A. Diamantopoulos. In defense of causal-formative indicators: A minority report. *Psychological methods*, 22(3):581, 2017.

[29] K. A. Bollen and R. Lennox. Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin*, 110(2):305, 1991.

[30] F. Booysen. An overview and evaluation of composite indices of development. *Social indicators research*, 59(2):115–151, 2002.

[31] R. Bru¨ggemann and G. P. Patil. *Ranking and prioritization for multiindicator systems: Introduction to partial order applications*. Springer Science & Business Media, 2011.

[32] G. Brunello and L. Rocco. The effect of immigration on the school performance of natives: Cross country evidence using pisa test scores. *Economics of Education Review*, 32:234–246, 2013.

[33] L. Carlsen and R. Bru¨ggemann. Fragile State Index: Trends and Developments. A partial Order Data Analysis. *Social indicators research*, 133 (1):1–14, 2017.

[34] R. Castellano and A. Rocca. Gender gap and labour market participation: A composite indicator for the ranking of european countries. *International Journal of Manpower*, 35(3):345–367, 2014.

[35] A. Charnes, W. Cooper, and E. Rhodes. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6):429 – 444, 1978. ISSN 0377-2217. doi: https://doi.org/10.1016/ 0377-2217(78)90138-8. URL http://www.sciencedirect.com/science/ article/pii/0377221778901388.

[36] L. Cherchye, W. Moesen, and T. Van Puyenbroeck. Legitimately Diverse, yet Comparable: On Synthesizing Social Inclusion Performance in the EU. *JCMS: Journal of Common Market Studies*, 42(5):919–955, 2004.

[37] L. Cherchye, W. Moesen, N. Rogge, and T. Van Puyenbroeck. An Introduction to 'Benefit of the Doubt'Composite Indicators. *Social Indicators Research*, 82(1):111–145, 2007.

[38] M. Ciommi, C. Gigliarano, A. Emili, S. Taralli, and F. M. Chelli. 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, 2017.

[39] C. Cobb, T. Halstead, and J. Rowe. *The genuine progress indicator: summary of data and methodology*, volume 15. Redefining Progress San Francisco, 1995.

[40] R. F. Curtis and E. F. Jackson. Multiple Indicators in Survey Research. *American Journal of Sociology*, 68(2):195–204, 1962.

[41] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press, 2002.

[42] A. M. Di Brisco and P. Farina. Measuring gender gap from a poset perspective. *Social Indicators Research*, 136(3):1109–1124, 2018.

[43] A. Diamantopoulos and J. A. Siguaw. Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British journal of management*, 17(4):263–282, 2006.

[44] A. Diamantopoulos and H. M. Winklhofer. Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38(2):269–277, 2001.

[45] A. Diamantopoulos, P. Riefler, and K. P. Roth. Advancing Formative Measurement Models. *Journal of Business Research*, 61(12):1203–1218, 2008.

[46] O. D. Duncan. *Notes on Social Measurement: Historical and Critical*. New York: Russell Sage Foundation, 1984.

[47] P. D'Urso. Classificazione fuzzy per matrici a tre vie temporali. *Unpublished Doctoral Dissertation, University of Rome "La Sapienza", Rome, Italy*, 2000.

[48] J. R. Edwards. The Fallacy of Formative Measurement. *Organizational Research Methods*, 14(2):370–388, 2011.

[49] J. R. Edwards and R. P. Bagozzi. On the nature and direction of relationships between constructs and measures. *Psychological methods*, 5(2): 155, 2000.

[50] EIGE. Gender equality index 2021: Health. https://eige.europa.eu/publications/gender-equality-index-2021-health, 2021.

[51] C. O. Europe. Intercultural cities index. https://www.coe.int/en/web/interculturalcities/, 2019.

[52] G. Extra and K. Yagmur. *Language Rich Europe: Trends in policies and practices for multilingualism in Europe*. 11 2012. ISBN 9781107633865.

[53] M. Fattore. Partially Ordered Sets and the Measurement of Multidimensional Ordinal Deprivation. *Social Indicators Research*, 128(2):835–858, 2016.

[54] M. Fattore. Synthesis of indicators: The non-aggregative approach. In *Complexity in society: From indicators construction to their synthesis*, pages 193–212. Springer, 2017.

[55] M. Fattore, F. Maggino, and A. Arcagni. Exploiting Ordinal Data for Subjective Well-being Evaluation. *Statistics in Transition - New Series*, 3(16):409–428, 2015.

[56] C. I. for Measuring Integration. Canadian index for measuring integration. https://www.integrationindex.ca/, 2016.

[57] M. Freudenberg. Composite indicators of country performance: a critical assessment. 2003.

[58] X. Gan, I. C. Fernandez, J. Guo, M. Wilson, Y. Zhao, B. Zhou, and J. Wu. When to use what: Methods for weighting and aggregating sustainability indicators. *Ecological Indicators*, 81:491–502, 2017.

[59] E. Giannakis and A. Bruggeman. Exploring the labour productivity of agricultural systems across european regions: A multilevel approach. *Land use policy*, 77:94–106, 2018.

[60] E. Grieco and K. A. Hamilton. Realizing the potential of migrant 'earn, learn, and return'strategies: does policy matter? *Center for Global Development's*, page 1, 2004.

[61] F. H. Harbison, J. Maruhnic, and J. R. Resnick. *Quantitative analyses of modernization and development*. Number 15. Industrial Relations Section, Department of Economics, Princeton University, 1970.

[62] O. HCNM. The ljubljana guidelines on integration of diverse societies. *The Hague: HCNM Office*, 2012.

[63] R. D. Howell, E. Breivik, and J. B. Wilcox. Reconsidering Formative Measurement. *Psychological Methods*, 12(2):205–218, 2007.

[64] S. P. Imperative. 2021 social progress index. https://www.socialprogress.org/global-index-2021, 2021.

[65] ISTAT. Equitable and sustainable well-being (bes) report. https://www.istat.it/en/well-being-and-sustainability/ the-measurement-of-well-being/bes-report, 2020.

[66] I. Italian National Institute of Statistics. *Rapporto BES 2015: il benessere equo e sostenibile in Italia*. Rome: Istat, 2015. URL https://www.istat.it/it/archivio/175169.

[67] P. Jensen and A. W. Rasmussen. The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30(6):1503–1515, 2011.

[68] S. Y. Koh. Inequality. In A. Kobayashi, editor, *International Encyclopedia of Human Geography (Second Edition)*, pages 269–277. Elsevier, Oxford, second edition edition, 2020. ISBN 978-0-08-102296-2. doi: https://doi.org/10.1016/B978-0-08-102295-5.10196-9. URL https://www.sciencedirect.com/science/article/pii/B9780081022955101969.

[69] J. Kondyli. Measurement and evaluation of sustainable development: A composite indicator for the islands of the north aegean region, greece. *Environmental Impact Assessment Review*, 30(6):347–356, 2010.

[70] M. Kovacevic et al. Review of hdi critiques and potential improvements. *Human development research paper*, 33:1–44, 2010.

[71] D. Krajnc and P. Glaviˇc. A model for integrated assessment of sustainable development. *Resources, conservation and recycling*, 43(2):189–208, 2005.

[72] K. C. Land. On the Estimation of Path Coefficients for Unmeasured Variables from Correlations among Observed Variables. *Social Forces*, 48 (4):506–511, 1970.

[73] K. C. Land. On the Definition of Social Indicators. *The American Sociologist*, 6(4):322–325, 1971.

[74] F. M. Lord and R. M. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing Company, 1968.

[75] F. Maggino. Dealing with syntheses in a system of indicators. In *Complexity in society: From indicators construction to their synthesis*, pages 115–137. Springer, 2017.

[76] F. Maggino. *Complexity in Society: From Indicators Construction to their Synthesis*. Cham: Springer, 2017.

[77] J. M. M. Mart´ın, J. A. S. Fern´andez, and J. A. R. Mart´ın. Comprehensive evaluation of the tourism seasonality using a synthetic dp2 indicator. *Tourism Geographies*, 21(5):284–305, 2018.

[78] M. Mazziotta and A. Pareto. On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social indicators research*, 127(3):983–1003, 2016.

[79] M. Mazziotta and A. Pareto. Synthesis of Indicators: The Composite Indicators Approach. In F. Maggino, editor, *Complexity in Society: From Indicators Construction to their Synthesis*, pages 159–191. Cham: Springer, 2017.

[80] G. Munda and M. Nardo. Non-compensatory Composite Indicators for Ranking countries: A defensible Setting. *EUR Report, EUR*, 21833, 2005.

[81] M. Nardo, M. Saisana, A. Saltelli, and S. Tarantola. *Tools for Composite Indicators Building.* European Commission. Institute for the Protection and Security of the Citizen, JRC: Ispra, Italy, 2005.

[82] J. Niessen, T. Huddleston, and L. Citron. Migrant policy integration index. Technical report, Technical report, British Council and Migration Policy Group, 2007.

[83] OECD. Handbook on Constructing Composite Indicators. Methodology and User Guide, 2008.

[84] OECD. Oecd better life index. https://www.oecdbetterlifeindex. org/, 2021.

[85] I. I. of Social Studies. Indices of social development. https://isd.iss. nl/, 2021.

[86] V. Perez, F. Guerrero, M. Gonzalez, F. Perez, and R. Caballero. Composite indicator for the assessment of sustainability: The case of cuban naturebased tourism destinations. *Ecological Indicators*, 29:316–324, 2013.

[87] M. E. Porter, S. Stern, and M. Green. *Social progress index 2014*. Social Progress Imperative Washington, DC, 2014.

[88] Preply. Worldwide language index. https://preply.com/en/d/ global-worldwide-language-index--lp, 2021.

[89] S. M. Rimoldi, A. Arcagni, M. Fattore, L. Terzera, et al. Social and material vulnerability of the Italian municipalities: Comparing alternative approaches. *Social Indicators Research*, pages 1–18, 2020.

[90] N. Rogge, L. Cherchye, W. Moesen, and T. Van Puyenbroeck. 'Benefit of the Doubt'Composite Indicators. In *European Conference on Quality in Survey Statistics*, 2006.

[91] M. Saisana and S. Tarantola. *State-of-the-art report on current methodologies and practices for composite indicator development*. EUR 20408 EN, European Commission-JRC: Italy, 2002.

[92] A. Sen et al. Commodities and capabilities. *OUP Catalogue*, 1999.

[93] G. Solano and T. Huddleston. Migrant integration policy index (mipex). https://www.mipex.eu/, 2020.

[94] E. Stebbings, T. Hooper, M. C. Austen, E. Papathanasopoulou, and X. Yan. Accounting for benefits from natural capital: Applying a novel composite indicator framework to the marine environment. *Ecosystem Services*, 50:101308, 2021.

[95] D. Storrie and H. Bjurek. Benchmarking European Labour Market Performance with Efficiency Frontier Techniques. Technical report, WZB Discussion Paper, 2000.

[96] D. Suhartanto, A. Brien, I. Primiana, N. Wibisono, and N. N. Triyuni. Tourist loyalty in creative tourism: the role of experience quality, value, satisfaction, and motivation. *Current Issues in Tourism*, 23(7):867–879, 2020.

[97] B. Talukder, K. W Hipel, G. W vanLoon, et al. Developing composite indicators for agricultural sustainability assessment: Effect of normalization and aggregation techniques. *Resources*, 6(4):66, 2017.

[98] E. C. Tarabusi and G. Guarini. An unbalance adjustment method for development indicators. *Social indicators research*, 112(1):19–45, 2013.

[99] J. Tica and I. Koˇzi´c. Forecasting croatian inbound tourism demand. *Economic research-Ekonomska istraˇzivanja*, 28(1):1046–1062, 2015.

[100] M. Torres. Unesco: Universal declaration on cultural diversity. *Paris, France. Retrieved last*, 2002.

[101] Undp. *Human development report 1997*. Oxford University, 1997.

[102] UNDP. 2020 human development perspectives: Tackling social norms–a game changer for gender inequalities. http://hdr.undp.org/en/gsni, 2020.

[103] UNDP. Gender development index (gdi). http://hdr.undp.org/en/content/gender-development-index, 2021.

[104] UNDP. The 2021 global multidimensional poverty index (mpi). http://hdr.undp.org/en/2021-MPI, 2021.

[105] F. Vidoli and C. Mazziotta. Robust Weighted Composite Indicators by means of Frontier Methods with an Application to European Infrastructure Endowment. *Italian Journal of Applied Statistics*, 23(2):259–282, 2013.

[106] WEF. World Economic Forum. *Global Gender Gap*, 2020. Geneva: World Economic Forum.

[107] WEF. Global gender gap report 2021. https://www.weforum.org/ reports, 2021.

[108] J. B. Wilcox, R. D. Howell, and E. Breivik. Questions about Formative Measurement. *Journal of Business Research*, 61(12):1219–1228, 2008.