Studi e Ricerche

Scienze sociali

Complexity of Social Phenomena

Measurements, Analysis, Representations and Synthesis

Leonardo Salvatore Alaimo

University Press

Sapienza Università Editrice

Collana Studi e Ricerche 127

Scienze sociali

Complexity of Social Phenomena

Measurements, Analysis, Representations and Synthesis

Leonardo Salvatore Alaimo

SAPIENZA
UNIVERSITÀ EDITRICE
2022

Sapienza Università Editrice
Piazzale Aldo Moro 5 – 00185 Roma

www.editricesapienza.it
editrice.sapienza@uniroma1.it

Iscrizione Registro Operatori Comunicazione n. 11420
Registry of Communication Workers registration n. 11420

ISBN 978-88-9377-251-8

DOI 10.13133/9788893772518

Pubblicato nel mese di novembre 2022 | *Published in November 2022*

Impaginazione/layout a cura di: Leonardo Salvatore Alaimo

In copertina | *Cover image*: Andrea Vizzini, *La mela di Magritte*, acrilico ed olio su tela, cm. 140 x 140 (2022).

Dedicated to
Laura and Alfio

Contents

List of Figures

List of Tables

List of Formulas

Preface

Leonardo Salvatore Alaimo

During the three years of my Ph.D., I analysed and studied phenomena often very different and apparently distant from one another: well-being; sustainable development; gender inequalities; the Brexit vote. The aim has always been to understand these *facets* of reality, to give them an explanation based on a *quantitative* point of view. My interest was to provide a measure of concepts often considered *difficult* to deal with and to understand.

When I finished my Ph.D., I tried to put together the experience I had gained. As mentioned, the research interests were many and, therefore, it was necessary to conceptualise them within a framework that would highlight the elements in common. This thesis is the result of such an attempt at conceptualisation. The title itself highlights the concepts common to my research work over the years.

The first concept I deal with is *complexity*. I have realized that all different socio-economic phenomena have in common their complex structure, often mistakenly exchanged with complication and difficulty. Nowadays, complexity is a concept that characterises all the natural and social sciences and defines our relationship with knowledge. The first Chapter examines precisely the theme of complexity, presenting different approaches and definitions to this issue. I have tried to reconstruct the way in which complexity became central in the relationship with knowledge, together with its qualifying concepts such as subjectivity, the concept of system and circular causality.

The second guiding concept of this research work is *measurement*. Understanding the world requires a sort of *translation*, a shift from the plane of reality in which we observe phenomena to the plane of numbers in which we try to encode them. This translation must be *meaningful*, it must reproduce as faithfully as possible in the world of numbers the phenomenon observed in the plane of reality. Measurement is a *need* for the knowledge of reality, which speaks to us with the *language of numbers*. This issue is the subject of the second Chapter, in which I address the question of the definition of this process. Subsequently, the measurement is contextualised within sociology, presenting the essential contribution on this theme offered by Paul F. Lazarsfeld with the operationalisation. Finally, the concept of *indicator* is explored, by analysing their crucial importance in the measurement of social phenomena. The Chapter presents all the main aspects through which it is possible to obtain a system of indicators, a tool for measuring complex social phenomena.

The way in which we can measure complex socio-economic phenomena is dealt with in the third Chapter. Synthesis is presented by a methodological point of view, considering both aspects of a system, units (rows) and indicators (columns). I focus on the synthesis techniques that allow a *dynamic analysis of phenomena* in order to obtain comparable measures not only in space, but also in time. Only in this way, a synthesis is meaningful. In the Chapter, I define the object of study, the three-way data array $\mathbf{X} \equiv \{x_{ijt} : i = 1, \ldots, N;\ j = 1, \ldots, M;\ t = 1, \ldots, T\}$, where x_{ijt} represents the determination of the j-th indicator in the i-th unit at the t-th temporal occasion. The methods of clustering these objects and summarising the indicators are addressed, considering both the aggregative and the non-aggregative approach (in particular, I propose an approach to apply posets to systems of indicators over time).

In the last two Chapters, I present two applications to real data. Both applications concern regional data. The choice was made because of the importance that the regional dimension has for a country like Italy, characterised by strong territorial disparities. The first one (the fourth Chapter) concerns the concept of well-being and, from a methodological point of view, the synthesis of statistical units. In particular, using the time series of regional composites produced by the Italian National Institute of Statistics for the Equitable and Sustainable Well-being project (BES), we classify the Italian regions according to different domains. We use a time series fuzzy clustering algorithm, particularly suitable for that type of data. The fifth Chapter deals with sustainable development and the issue of synthesis of statistical indicators over time. In particular, an aggregative method, the Adjusted Mazziotta-Pareto Index (AMPI), and a non-aggregative procedure based on posets will be compared.

1. Complexity: definitions and conceptual approaches

I think the next century will be the century of complexity. We have already discovered the basic laws that govern matter and understand all the normal situations. We don't know how the laws fit together, and what happens under extreme conditions. But I expect we will find a complete unified theory sometime this century. There is no limit to the complexity that we can build using those basic laws.

Stephen Hawking, 2000

In this interview on January 23, 2000 to *The San Jose Mercury News*, Hawking clearly stated that the 21st would be the century of complexity [62]. In the recent years, the topic of complexity has been a subject of great debate and interest becoming a mainstream issue both in natural and social sciences. This term is used in different fields (e.g., physics, chemistry, biology, engineering, sociology, psychology), even though it is relatively new in science (in particular in social sciences). But what is complexity? What does complex mean? In this chapter, I will try to answer these questions.

Complex is, sometimes, an abused term, used instead of other more appropriate terms, like large, difficult or complicated. First of all, it must be clarified that complex is not the same as *complicated*. The two terms are often used as synonyms to indicate the difficulty and incapacity of managing a situation. The difference lies in the etymology itself [6]. Complicated comes from the Latin term *cum plicum*, where *plicum* means paper crease. Complex comes from the Latin term *cum* *plexum*, where *plexum* means knot, weave. Complicated refers to the linear plicum, while complex refers to the interwoven plexum. It indicates something with interweaving, composed of many interconnected parts; compound; composite. An intricate association or assemblage of related things, parts, or units; an interrelated system [78, 77, 164]. Dealing with complicated problems requires the adoption of an analytic approach: one finds the right solution by *unfolding* problems in their *creases*, by breaking them down into their elementary components. One analyses the individual parts and finds a solution for each of them. As difficult as the problem may seem, it is always possible to find a solution. Dealing with complexity requires a synthetic or systemic approach. It is not possible to understand the *plexum* by analysing the individual components, because one would lose the whole. "If you split up the fabric weave in its threads or basic components, you obtain a group of threads whose analysis does not help recreate the original system, of the original fabric" [78, 16]. The solution to the complex problem must be found by trying to understand it as a whole. In brief, we understand the single elements and the structure of a complicated system; nothing prevents us from fully understanding it with time. On the contrary, we can only have a global perception of a complex system, through which we can understand it without understanding its elements [21].

	Complicated	Complex
Etymology	*cum plicum*	*cum plexum*
Approach	Analytic	Synthetic
Solution	Unfolding in its creases	Understanding as a whole

Fig. 1.1. Main differences between *complicated* and *complex*.
Source: Personal elaboration based on Alberto F. De Toni and Luca Comello [77].

We use complex often as a synonym for *difficult*. A complex problem is sometimes considered difficult because we are unable to understand or explain it. But actually, this difficulty is not inherent in the complex nature of the problem, but in trying to study it with an analytical approach, to break it down into its essential components, rather than to understand it as a whole. It should also be highlighted that complexity is different from *completion*. Having a complex view of reality does not mean having a complete view of it. In understanding complexity, everything is interdependent, we cannot isolate the elements from one another. Having the sense of complexity means having the sense of solidarity, the sense of the multidimensional nature of reality [205].

From what has been written above, we need different approaches to the complicated and the complex. According to Fritjof Capra [57], the transition from analysis to synthesis represents one of the most important advances in the 20$^{\text{th}}$ century science: "The properties of the parts can be understood only within the context of the larger whole…Systems thinking is *contextual*, which is the opposite of analytical thinking. Analysis means taking something apart in order to understand it; systems thinking means putting it into the context of a larger whole" [57, 29:30]. Contextuality is one of the main characteristics of complex systemic thinking: we must search for the sense of things, their meaning, within the context in which they are observed, in relation to the reality that surrounds them. The transition to systemic thinking coincides with the awareness of understanding systems by means of analysis. "In the shift from mechanistic thinking to systems thinking, the relationship between the parts and the whole has been reversed. Cartesian science believed that in any complex system the behaviour of the whole could be analysed in terms of the properties of its parts. Systems science shows that living systems cannot be understood by analysis. The properties of the parts are not intrinsic properties but can be understood only within the context of the larger whole" [57, 37]. From Fritjof Capra's thought, it is clear that the synthetic approach does not aim at reducing complexity. A meaningful synthesis must be able to *stylise* reality, presenting those characteristics that arise from that particular and often unique interconnection between its elements. The importance of synthesis for understanding complex phenomena will be deeply analysed in Chapter 2 regarding the measurement of social phenomena.

Complexity in science has no precise meaning and no unique definition [102]. We cannot approach the study of complexity through a preliminary definition: there is no such thing as one complexity, but different *complexities* [203]. Therefore, it seems more appropriate to talk about definitions of complexity. This term can assume profoundly different meanings because it has been influenced by the contribution of many disciplines. Complexity does not belong to a particular theory or discipline, but rather to a *discourse about science*. As we will see in paragraph 1.1, its importance coincides with a transformation in the relationship with knowledge. This concept has had an evident impact on all fields of knowledge, leading some scholar to consider it a new paradigm, although there are conflicting opinions about this (I analyse this question in paragraph 1.3). As shown by Alberto F. De Toni and Luca Comello [78], the journey of complexity starts off from the awareness of knowing very little about the phenomena around us. The multi-disciplinary progress made by science in the 20th century has given rise to a coherent system of knowledge called *complexity theory*. "Instability, non-equilibrium, irreversibility, chaos and disorder are some of the keywords of this new science" [78, 14]. Complex systems are the objects of complexity theory; each one of us is a complex system, made up of many different elements which are interconnected. Even though no unique definition exists, a consensus is being gained on the main characteristics which complex systems share (paragraph 1.2) and on the principles of complexity theory (paragraph 1.1.2).

1.1. Complexity and knowledge

The concept of complexity is closely linked to that of knowledge. Humanity has always had the aim of reflecting on its existence and investigating the possibilities and limits of knowledge. The importance of complexity in sciences begins when one realises the lack of human understanding of phenomena: each increase in knowledge corresponds to an increase in ignorance and inability to know. Thus, the growing attention to complexity coincides with a real evolution in science. Paraphrasing Isaac Newton and Robert King Merton [192], this is a *journey on the shoulders of giants*, given the stature of the scholars who have marked this path. The concept of complexity is the cornerstone for understanding the transition from classical to modern science.

1.1.1. Classical science: principles and criteria

The scientific revolution of the 17th century, which had as its protagonists scientists like Copernicus, Galileo, Descartes, Bacon, Newton led to the so-called *classical* science. The main objective of this new scientific approach is to simplify things and make phenomena widely predictable reducing them to their simple elements. In this way, it is believed possible to achieve an *objective* knowledge. The true and correct understanding of phenomena should seek their stability and unchangeability, considered essential characteristics of their objective nature.

The *causal explanation* is one of the main criteria of classical science. It is based on the assumption that by finding the cause of a phenomenon we explain its behaviour. Thus, an object is explained if its cause can be assigned. The latter is the *factor* of an object. If the cause occurs, inevitably determines the occurrence of the object. Consequently, the object is perfectly predictable. Therefore, causal explanation is based on two cornerstones:

- the cause is an irresistible productive factor, to which the effect necessarily follows;
- consequently, given the cause, the effect is infallibly predictable.

This view does not admit uncertainty. Complex objects, being by their nature unpredictable, are labelled as *non-scientific*. Determinism is certainly a crucial issue. It corresponds to the idea that it is possible to predict the future from the present. Newtonian mechanics is based on this principle: if we know the position and speed of an object at a given time, then we can know them at any other time.

The method used by classical science is the *experimental* one introduced by Galileo Galilei. This is based only on what is expressly observable and measurable. It marks the overcoming of any metaphysical reference in the knowledge of Nature. The experimental method focuses on the observation of phenomena, on the use of mathematics and on the reproducible experiment. Through the observation and repeated experimentation we can interpret the mathematical relationships that underlie and determine natural phenomena. Scientific hypotheses are then formulated and subject to the control of the experimental method. The confirmed hypotheses become scientific laws. *Linearity*[1] is the characteristic of the laws that describe reality: to certain causes correspond certain effects, which vary following linear laws. Time loses its meaning in classical science: it is an ideally reversible series of homogeneous events referable to quantitative laws. Another fundamental principle of classical science is *equilibrium*, on which reality is based. Everything is in equilibrium; if something does not seem to be in equilibrium, it is because of human limits. Objects are *closed systems*, isolated from the environment. They must be studied independently from the environment in which it is placed. A closed system is in a state of equilibrium.

[1] Linearity refers here only to the causal explanation.

A deterministic, in equilibrium, linear reality is also an *ordered* one, that is, governed by precise rules and laws, in which there is no room for chance and uncertainty.

Classical science is based on a separation between subject and object, between beings and Nature, considered the only possible way to an objective knowledge. The aim is to search for a *model*, an ideal representation of the phenomena that encloses all their characteristics. It is necessary to achieve the *Platonic Hyperuranium*, the perfection in terms of generalisation and immutability. Reality is inaccessible to complete knowledge; so, we can construct a representation of it based on information gained from experience.

Reductionism is one of the main principles of classical science. It has been strongly criticized; an element common to the various scholars of complexity is precisely the rejection of this principle. Nevertheless, it has had undoubted merits and has influenced our approach to knowledge. Reductionism entered scientific thinking between the 17th and 18th centuries, linked to the spread of the Newtonian mechanistic model, according to which reality can be *reduced* in terms of elementary particles and their movements. The importance of this principle for classical science lies in the idea that all phenomena can be explained rationally through mathematical models and laws. Knowledge is achieved by searching for a two-way correspondence between reality and a mathematical model capable of grasping its order. Therefore, it must be searched for objective coincidence between what Baruch Spinoza in *De Intellectus emendatione* defined *ordo idearum* (the order of knowledge) and *ordo rerum* (the order of nature). This correspondence between the structure of the real and the mathematical form can only be obtained through a *reduction* of the heterogeneous to the homogeneous. The complex and qualitative aspects of phenomena must be reduced to a purely quantitative and measurable level. This approach inevitably leads to the mechanistic construction of the reality and to determinism. In ancient philosophy the word *cause* had a general meaning. In Physics II 3 and Metaphysics V 2, Aristotle developed a theory of causality [132, 103]. It is commonly known as the doctrine of the four causes, according to which there are four causes behind all the change in the world [233, 852]:

- the material cause (ὕλη) is what an object is made of, its actual physical properties. It's what we can see, touch, taste, and so on (for instance, the wood is the material cause of a table);

- the formal cause (εἶδος) is the structure or design of an object, what makes it one thing rather than another (going back to the example of table, the original carpenter could have chosen to make the very same wood into a chair, but he didn't. Instead, his plan, or design, called for putting the wood together as a table: this is the formal cause);

- the efficient cause (ἀρχή τῆς κίνήσεως) is what the change and movement of things comes from. This is the thing or agent which actually brings something about (for the table, the carpenter who made it);

- the final cause is the end or purpose of things or actions; it constitutes that in function of which (τὸ οὗ ἕνεκα everything is or becomes; this is the good (ἀγαθόν) of each thing (the purpose for which the table is used: eating, studying, writing, etc.)

According to Aristotle, all the four kinds of causes may enter in the explanation of phenomena. After the scientific revolution of the 17th century there was a reduction in the concept of causality. Only the efficient cause corresponds more or less to what the classical science means by the word cause. This new interpretation of causality is closely linked to the reduction of phenomena to the purely material aspect. The principle of causality is reduced to become synonymous with *expectation*. This means that natural events are clearly determined, that it is possible to know them exactly and objectively and, consequently, to predict their future.

This new approach affects all fields of knowledge. There is a *schism* between classical science and the humanities. The former, with its quantitative, mathematical and demonstrable methods, contrasts with the latter, with their qualitative, imprecise and sometimes bizarre methods. This separation is an example of the more general rejection of any metaphysical explanation in favour of the search for the purely empirical, the demonstrable through the detached and autonomous analysis of reality.

1.1.2. The complexity theory: origins and principles

The importance of classical science is also recognised by scholars of complexity, who will nevertheless demonstrate the insufficiency of its principles and criteria. The rejection of the concept of complexity by classical science is due to three of its fundamental principles [204]:

1. the principle of **universal determinism**, according to which it is possible not only knowing all past events, but also predicting all events in the future,

2. the principle of **reduction**, according to which we can understand any phenomenon simply from the knowledge of its elements,

3. the principle of **disjunction**, that consists in separating cognitive difficulties from one another, leading to the separation between disciplines.

Based on these assumptions, complexity is absolutely rejected. Conceived as a synonym for uncertainty and confusion, it concerns only superficial or illusory appearances, since the criterion of truth of classical science is expressed by simple laws and concepts.

In the 19^{th} century, the birth of thermodynamics marks the beginning of the decline of classical science. Edgar Morin [204] states that the affirmation of the concept of complexity is linked to that of the second law of thermodynamics[2]. According to the concept of *entropy*, introduced by Rudolf Clausius in 1865, in every mechanical process part of (or all) the energy is dissipated in the form of heat (in other words, entropy can be considered as the impediment to the transformation of all the energy contained in a system). If a system has a limited amount of energy and is isolated (*closed system*), it is destined to exhaust the amount of transformable energy. Considering the Universe as an isolated system, we conclude that its entropy increases. Since each phenomenon involves a transformation of energy and the amount of energy of the Universe is finished, it follows that a day will come when each phenomenon will be impossible[3]. In short, all phenomena have a specific trend (*principle of irreversibility*), which is the one that tends to increase entropy. The latter grows until it reaches the state of *thermal equilibrium*, where changes in

[2] The second law of thermodynamics highlights the impossibility of transforming a certain amount of work entirely into energy. The fraction that can be transformed depends on the difference in temperature between the hot source supplying energy and the cold source that receives it: the higher this difference, the greater the quantity of work transformed into energy.

[3] According to the second law of thermodynamics, the day when the temperatures of the Universe will be equal.

the system are no longer possible. Entropy is also interpreted as the amount of disorder in a system, because heat is the random movement of system' elements. Therefore, the second law of thermodynamics describes the irreversible movement of closed systems towards a state of disorder. In every closed system, energy is constant while entropy tends to a maximum. Based on the work of Rudolf Clausius, in 1877 Ludwig Boltzmann elaborated a definition of entropy in statistical form as a measure of the degree of disorder of a system. According to Ludwig Boltzmann, there is a precise relationship between entropy and probability expressed in the formula:

$$S = K_b * lnP \tag{1.1}$$

where S is the entropy, K_b is the Boltzmann constant and P is the thermodynamic probability of a system' state, depending on the number of all possible configurations that generate the same thermodynamic state. The increase of entropy in an isolated system can be considered as the effect of the tendency of that system to evolve from a less probable to a more probable state. Ludwig Boltzmann introduces into physics a new way of interpreting reality, re-dimensioning the deterministic view through a probabilistic approach. Determinism remains, however, understood as the system's irremediable tendency to disorder (the more probable state). Therefore, thermodynamics is fundamental to understand the complexity theory because it introduces the concepts of disorder, irreversibility and instability. It focuses on irreversible and disorderly phenomena, while, classical science has as its object ordered and reversible phenomena.

Other important contributions to the advent of complexity are provided by Albert Einstein's theory of relativity (1905, 1915) and Werner Heisenberg's uncertainty (or indeterminacy) principle (1927). The first, although still based on a deterministic conception of reality, states that the classical mechanics cannot explain all phenomena, in particular the macrocosmic ones. Time and space are no longer considered absolute quantities and the importance of the active role of the researcher is recognised[4]. The principle of indeterminacy introduces the impos-

[4] In the theory of special relativity (1905), Albert Einstein determined that the laws of physics are the same for all non-accelerating observers, and that the speed of light in a vacuum was independent of the motion of all observers. In the theory of general

sibility of representing everything through a model[5]. Those theories underline how the laws of Newtonian mechanics, although not wrong, are insufficient and do not allow the explanation of all phenomena[6].

Several disciplines developed in the 20th century formed the basis for what would later become the theory of complexity[7]. Common to all of them is the introduction of principles different from those of classical science and which will be the basis of the complexity theory (Table 1.1 reports the main contributions).

Tab. 1.1. Fundamental contributions to complexity theory by some disciplines of the 20th century. Source: personal elaboration based on [77].

Discipline	Fundamental contributions
System theory	Open systems; system thinking
Cybernetics	Circularity; feedback; open systems
Chaos theory	Butterfly effect; order and disorder; chaos and determinism

The main concepts of complexity derive from the research work of Ilya Prigogine, the best known scholar in that field. He examines in depth the principles elaborated in other sciences (Table 1.1) and systematises them by an unified vision, creating *de facto* the complexity science. The starting point of Ilya Prigogine's thought are the concepts of *open system* and *entropy*. In contrast to classical science and thermodynamics, systems cannot be conceived as closed[8], because each of them is con-

relativity (1915), he determined that massive objects cause a distortion in space-time, which is felt as gravity.

5 According to this principle, we cannot measure exactly both the position and the velocity of an object, at the same time, even in theory. The very concepts of exact position and exact velocity together, in fact, have no meaning in nature. This clearly contradicts the laws of Newtonian mechanics, which do not apply in the microcosmic field.

6 For a more detailed analysis of the contributions of these disciplines to the complexity field, please see: Ilya Prigogine and Isabelle Stengers [229].

7 The importance of complexity was highlighted for the first time in psychology. Starting from the early 20th century, in Austria and Germany a new school of psychology (among its exponents Max Wertheimer, Wolfgang Köhler, Kurt Koffka) arises: the *Gestalt* psychology. Gestaltists affirm that the conscious experience must be considered by taking into account all the physical and mental aspects of the individual simultaneously, because each component is part of a system of dynamic relationships. We must understand objects as an entire structure rather than the sum of its parts. Thus, the concept of Gestalt in psychology is a good example of something complex that cannot be reduced to its elements; functionalism and structuralism in anthropology and sociology can be regarded as necessary approaches to a complex reality [248, 252].

8 Universe is the only system of which there is no empirical evidence whether it is open

tained within other systems and can exchange energy and information with them. For such open systems, the second law of thermodynamics is not always valid without limitations; the variation of entropy is given by

$$\Delta S = \Delta_i S + \Delta_e S \tag{1.2}$$

where $\Delta_i S$ is the entropy produced within the system and $\Delta_e S$ is the one that the system receives from the outside. The latter can be null (in a closed system), positive or even negative, the last situation is defined *negentropy* [243], i.e. the reverse concept of entropy, which describes the order that can emerge from chaos. This does not conflict with the second law of thermodynamics. In practice, systems tend to evolve between two opposing tendencies: entropy (disorder) and negentropy (order). They tend towards entropy, but they can also tend to a state of minimal entropy by importing energy from the outside[9]. So order and disorder, structure and change are linked together. Ilya Prigogine's work shifts the attention from stability to instability, *from being to becoming*, like the title of one of his famous books [227]. "We will need to associate the antagonist principles of order and disorder, and associate them making another principle emerge that is the one of organization. Here is in fact a complex vision, which one has refused to consider during a very long time, for one cannot conceive that disorder can be compatible with order, and that organization can be related to disorder at all, being antagonist to it" [204, 3]. Systems can, therefore, be in equilibrium, but also in *non-equilibrium*. Determinism and fate coexist. This means that when equilibrium prevails there is determinism, while when non-equilibrium prevails the fate has an essential role. Uncertainty enters science, not in opposition to determinism: some phenomena can be predicted (determinism) and others cannot. Periods of linearity are followed by periods of *non-linearity*, where small changes can generate great effects (the so-called *butterfly effect*). Reality is therefore characterised by the presence

or closed.

[9] This is what happens in the so-called *dissipative structures*, open systems that not only maintain a state of stability away from equilibrium, but can also evolve. In fact, when the flow of energy or information through them increases, they can evolve into new more complex structures, through new phases of instability. This is thanks to the import of negentropy [226].

of concepts, which only apparently seem to exclude each other[10]. Contradiction is a purely complex concept. "In the classical view, when a contradiction appears in reasoning it is a sign of error. You have to back up and take a different line of reasoning. However, in a complex view, when one arrives via empirical rational means at contradictions, this points not to an error but rather the fact that we have reached a deep layer of reality that, precisely because of its depth, cannot be translated into our logic" [205, 45]. Complexity does not contradict classical science, but can be considered a complement to it, adding principles and concepts to those already present, as shown in Figure 1.2.

	Classical Science	Complexity
Equilibrium	X	X
Non-equilibrium		X
Closed systems	X	
Open systems		X
Determinism	X	X
Fate		X
Linearity	X	X
Non-linearity		X
Reveribility	X	X
Irreversibility		X
Order	X	X
Disorder		X

Fig. 1.2. Main concepts of classical science and complexity.
Source: Personal elaboration based on Alberto F. De Toni and Luca Comello [77].

One of the main differences between these two approaches to science is the concept of *time* and its interpretation. Time in classical science is conceived as a theoretically reversible set of homogeneous states, explainable by mathematical laws and connected by causal links. Irreversibility becomes a key element of complexity: the past does not

10 These concepts coexist in nature, without contradictions, thanks to the instrument of *bifurcation* [229, 160–170], which marks the evolution of phenomena. There are periods of stability and equilibrium. When the system reaches the point of bifurcation there is discontinuity and rupture of the equilibrium. At this point the behaviour of the system will follow a non-linear trend.

imply a certain future; the latter cannot be known from the analysis of a series of conditions. In short, the future is open. Reductionism is profoundly reshaped, because complexity requires a new way of looking at the world, a complex, non-intuitive and non-linear causality, alongside the simple one derived from classical science. In order to understand complex phenomena, it is necessary to accept that there is a *circular relationship* between causes and effects (or, more precisely, between different and interconnected aspects of reality).

The emergence of complexity makes clear the impossibility of a complete and objective knowledge of Nature. Its static and immutable image, unjustified and wrong, is lost. It becomes clear that it is dynamic, temporal, in perpetual becoming. An example of this new thought is the analysis of the concept of Nature proposed by Ilya Prigogine and Isabelle Stengers [228]. The authors state that the presumed dialogue *detached*, established by classical scientists with Nature, has no theoretical consistency, but only operational. "Experimental dialogue with Nature does not imply passive observation but rather practice" [228, 41]. It is impossible to study phenomena isolated from their context, conceptualising them as ideal entities (as clearly stated by Fritjof Capra [57]). Modern science is characterised by the encounter of theory and practice: "the only possible way to knowledge is the systematic alliance between the ambition to model the world and that of understanding it" [228, 40]. This new conception emerges when time and dynamism break into classical physics, upsetting it. Time enters into areas from which it was traditionally excluded, where it was believed that there were *eternal laws* (at the microscopic and macrocosmic level). Physical or social phenomena are all characterised by an intersection of times and different speeds that make simplification absolutely ineffective. They are complex and made up of a plurality of times that give rise to articulated results. It should be pointed out immediately that the multiplicity of time and its relevance have always been known, but were practically ignored and denied. It is not possible to conceive of any form of knowledge that is not oriented in time, that does not have a *before* and an *after*. Nature is, therefore, an entity that grows and develops over time, not a static object regulated by immutable laws. Each phenomenon manifests itself in an articulated way and presents the fundamental characteristic identified by Aristotle: from the interaction of the parts emerge new properties not present in the single parts. Edgar Morin [202] defines them *emergencies*, qualities and properties of a system that present a character of novelty with respect to those of the individual parts taken alone or linked by different interactions in another system. "Emergency is a product of the organization that, although inseparable from the system as a whole, not only appears on a global level, but can appear at the level of the components" [65, 208]. According to Edgar Morin, a system is both more and less than the sum of its individual components, since it binds the parts themselves, making them different from what they were originally or could have become in a different system.

We cannot photograph reality as it is: the researcher builds a series of levels of reality, the result of his cultural preferences and cognitive abilities [182]. The idea of an objective and immutable knowledge and of the researcher distinct from the object of his investigation collapses. The only way to know reality is a dialogue between the researcher and Nature, a dialogue that necessarily presupposes the subjective component. Knowledge is a dialectical path between beings and reality. The myth of isolating phenomena to understand them falls. Each of us is an integral part of reality and moves in an environment in which it is conditioned and which inevitably conditions. In this sense, we can affirm that complexity is *subjective*: the observer, on the basis of his knowledge and experience of phenomena, establishes whether reality is more or less complex. In this vision, our knowledge is always relative to and is conditioned by a point of view. It is a product of our mind.

The science of complexity, as we have pointed out, is based on the contributions of many disciplines. In this, it is different from classical science (based on a clear separation between the various fields of knowledge, also considered of different importance). The essence of complexity is to be a *multidisciplinary thought*, to conceive all disciplines as different aspects of the same reality.

1.2. The concept of complex system

Complex is often associated with the term system. Before analysing the complexity of a system, we must define what a system is. This term is used both in the common language and in that of many scientific disciplines and its relevance is such that it has given rise to a *systems theory*. We could trace back the *official* birth of this theory to the foundation in 1954 at Palo Alto of the *Society for General System Research*. It was a group of researchers of different disciplinary fields, led by the father of systemic theory, biologist Ludwig von Bertalanffy. This group wanted to develop a theory that could relate traditionally separate fields of knowledge. The concept of system seemed perfectly suited to this purpose. According to Ludwig von Bertalanffy [29], deterministic explanations are insufficient in the analysis of complex phenomena. It is not the individual causalities, independent from one another, that determine the evolution of the systems, but entire interrelated causal complexes. A system is able to reach the same final state of dynamic equilibrium regardless of the intervention of individual causal factors (the so-called principle of *equifinality*).

System has many meanings. Ludwig von Bertalanffy [29] defines it as a set of elements standing in interaction. This definition does not formally clarify which are the elements themselves. Furthermore, there is no reference to the criterion for choosing either objects or relations that are given a systemic character, i.e. there is no observer of the system. The criterion of choice, specific to the observer, appears in the definition of James G. Miller [196]: a system is a region delimited in space-time, where the term *delimited* evidently refers to an observer who delimits and then chooses. In contemporary systemic theory, no one refuses to introduce this *observer-dependence*, considered a fundamental component. A more precise definition is given by Edgar Morin [202]: an organized global unit of interrelationships between elements, actions or individuals. A set of elements, to be defined as a system, must be governed by an organisational principle that establishes the rules of interaction between the elements.

An essential contribution to the theory of systems and the development of systemic thinking has been made by Donella H. Meadows. She defines a system as "an interconnected set of elements that is coherently organized in a way that achieves something" [189, 11]. This definition identifies the three main components of a system: *elements*, *interconnections* and *functions*. A system is not just a collection of things; they must be interconnected and have a purpose, i.e. they must be aimed at achieving an objective. The purpose of a system is often difficult to understand. "The best way to deduce the system's purpose is to watch for a while to see how the system behaves" [189, 14]. From this Donella H. Meadows' statement, it can be deduced that a system has its own behaviour, different from its parts and that, like any behaviour, it can change over time. Each system is based on a *stock*, i.e. the elements that constitute it in a given time. These stocks change over time due to the effect of *flows*. "Flows are filling and draining, births and deaths, purchases and sales, growth and decay, deposits and withdrawals, successes and failures" [189, 18]. Donella H. Meadows highlights the dynamism of the systems, their adaptation over time. One cannot understand them without understanding their dynamics of stocks and flows. Obviously, the change can concern both the system as such and one or even all of its essential components. Change can also be traumatic and unexpected. Most of systems are able to withstand the impact of drastic changes thanks to one of their fundamental characteristics, *resilience*. It is a "measure of a system's ability to survive and persist within a variable environment" [189, 76]. It is both the ability to adapt to change by evolving and the ability to resist it by restoring its initial state. Resilience presupposes change: it is not static being, but becoming.

A system is, therefore, an organic, global and organized entity, made up of many different parts, aimed at performing a certain function. If one removes a part of it, its nature and function are modified; the parts must have a specific architecture and their interaction makes "the system behave differently from its parts" [115, 17]. Systems evolve over time and most of them are resilient to change.

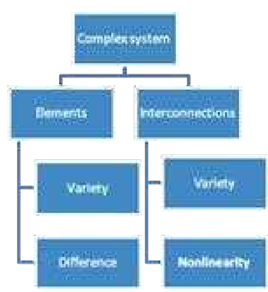


Fig. 1.3. Components and characteristics of complex systems.

At this point, we must ask ourselves what characteristics make a system *complex*. Complex systems (Figure 1.3) are made up of a great variety of elements, which have specialised functions. Therefore, elements are different from one another and it is precisely this diversity that makes it difficult their understanding [5]. The elements of most complex systems are not simple, but are other systems. Such mini-systems are in turn formed by systems and so on. This progressive encapsulation forms a *systemic hierarchy*, an essential characteristic of complex systems. The presence of a hierarchical structure allows the control of the elements, ensuring that they act in a coordinated and harmonious way. This type of structure is governed by the *slaving principle* [125, 126], according to which the elements at a lower hierarchical level are *slaved* to the upper level, to the overall (holistic) behaviour of the system. The term *slaving* underlines how the birth of a hierarchy presupposes a limitation of the freedom of the system's elements. A real contradiction is generated: the system is at the same time *more* and *less* than the sum of its parts. More because new qualities emerge from the whole, which would not exist without that specific organization. The whole is greater than the sum of its parts (the so-called system holistic principle). It is less because it creates one of the many possible organizations. In fact, "the organization imposes constraints that inhibit some of the potentiality found in its parts" [203, 51]. Finally, the complex system is *different* from the sum of its parts, because its properties and behaviours are unexpected. In order to understand a complex system, it is not necessary to analyse all its hierarchical levels, all the subsets that constitute it. Obviously this does not mean that we have to neglect the internal dynamics between the elements. It means that in a complex system the *interconnections* between the elements are more important that the elements themselves. The high density of interconnections is typical: the various elements are connected by a great variety of links. The interactions between the elements are non-linear. This is a fundamental characteristic. In simple systems the whole is strictly equal to the sum of its parts, the connections do not bring any added value. Non-linear connections are important in the definition of the structure and the organization of the system.

A crucial point for the study of complex systems is the so-called phenomenon of *self-organization*. This concept, closely related to hierarchy, expresses the possibility that highly organized behaviours arise from the circularity of relationships even in the absence of a planning or a project. In simple terms, it is an organization that emerges without having been projected by anyone. We must clarify that the concept of self-organization is different to that of selection. "Self-organization only considers the spontaneous movement of the system from one attractor to another attractor, movement that is not caused by any thrust coming from outside the system…Selection, on the other hand, is a choice between different stable states, therefore of equilibrium, that are in competition with one another; this choice takes place with reference to criteria that are external to the system" [30, 271–272]. At the initial level, each element acts according to relatively autonomous rules: the system appears unconnected and uncoordinated. As the system evolves, the interactions between the elements take on increasing importance, becoming more and more articulated, structured and varied. Thus, the system becomes more and more complex. Self-organization occurs when it exceeds a *critical threshold of complexity*. Structures involving and organising the elements in a harmonious way are formed. A hierarchical level is developed. New properties are generated (the *emergencies* of Edgar Morin [202]), which are not present in the elements and, consequently, cannot be understood from the analysis of their behaviours[11]. Therefore, self-organization consists in the passage from a myriad of individual and chaotic behaviours to a global and ordered one. We are faced with a paradox: the birth of a new hierarchical level makes the system simpler, reducing its complexity. The paradox is only apparent. In fact, as soon as the new hierarchical level is stabilised, a new evolution begins which will lead to an ever increasing complexity. Once the critical threshold is exceeded, a new process of self-organization begins. It is a process that continues indefinitely. A new model of organisation involves a simplification of the systemic structure, but also the beginning of a process of progressive complexification [160].

Simple systems are characterised by few elements and few linear relationships between them; they can be analysed analytically. Complex systems, on the contrary, are made up of many elements and many relations, linear and non-linear; they can only be understood in a synthetic way. In a complex system, elements and connections, besides being numerous, are various and different. As shown by Péter Érdi [102], simple systems are based on the principles of *simple cause and a single effect* and suggests that small changes in the cause imply small changes in the effect. It means that there is a linear relationship between the cause and the effect and that the system's behaviour is predictable. At the opposite, complex systems are based on circular causality (analysed in paragraph 1.1.2) and on the principle *small change in the cause, dramatic effects*. They need continuous flows from the environment in order to exist and function. The non-linearity implies two consequences:

1. the behaviour of complex systems is not necessarily proportional to the input they receive;

2. they are organized on *networks*, not on sequential processes.

---

11 Because they depend not on the nature of the elements but on their *relations*.

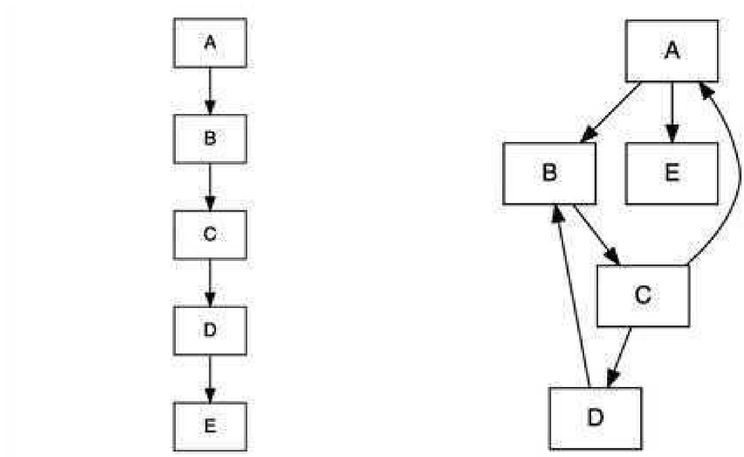


Fig. 1.4. Systems organization: sequential process on the left and network process on the right.

Figure 1.4 shows the two different organisational processes. On the left, there is an example of a sequential process: once phase A is over, phase B begins; once phase B is over, phase C begins and so on. It is a *linear chain*. On the right, a *network process* is shown. The various phases are no longer arranged in a linear chain: one phase can influence 2 or 3 others and can be influenced by them. The deep ones at the end of a process can also be included as new inputs in a process at the beginning. This creates *feedback*, which forms cycles because the final outputs return to the sequence, influencing it.

A particular type of complex system is the *Complex Adaptive System* (CAS). It can be defined as an open system made up of numerous elements interacting with each other, in linear and non-linear way, that constitute a unique and organic entity capable of evolving and adapting to the environment. We can deduce this definition from Mitchell M. Waldrop [264], who, reporting an analysis of John H. Holland[12], identifies the main characteristics of CASs. They add to the other characteristics typical of complex systems the ability to adapt and *learn*. CASs

[12] John H. Holland was one of the main researchers in the field of non-linear and complex systems. He played a fundamental role in creating Michigan's Center for the Study of Complex Systems. Outside of his Michigan activities, hee became an active member of the Santa Fe Institute in New Mexico, a novel research institution established in 1984 to further the study of complexity and non-linear phenomena.

are able to adapt to the world around them by processing information and building models capable of assessing whether or not adaptation is useful. The elements of the system have the main purpose of adapting and, in order to achieve this purpose, they constantly look for new ways of doing things and learning, thus giving rise to real dynamic systems. These systems challenge our ability to understand and predict. As stated by Arthur Battram [24], they are perfectly placed in the middle of *simple* systems, in which the connections between the elements are fixed and the behaviour is easily predictable, and *chaotic* systems, whose components are dispersed and free to interact, generating absolutely unpredictable behaviour. Compared to the latter, the CASs have a hierarchical structure that maintains a certain control. The behaviour is emerging from the interactions between the elements and oscillating between predictability and unpredictability.

It is evident that the main characteristics of complex adaptive systems are typical of social organizations and phenomena. Each of them is made up of a network of elements, which interact both with one another and with the environment. They are multidimensional and their different elements or dimensions are linked together in a non-linear way. They evolve over time, modifying both their dimensions and the links between them. The measurement and analysis of social organizations and phenomena requires the definition of systems of indicators capable of capturing their different aspects. As can be easily understood, these systems are dynamic, since they have to adapt to the changes in the measured phenomena. In simple terms, they are CASs and can be monitored and measured through systems of indicators that are CASs themselves.

1.3. Does a paradigm of complexity exist?

The issue of complexity cannot be approached by means of a preliminary definition. There is no such thing as one complexity, but different complexities. There is no one single way to complexity, but multiple ways to it. In this sense, Isabelle Stengers [249] states that there cannot be a *paradigm of complexity*, since it does not have an epistemological status comparable to that of other scientific notions. Complexity does not belong to a particular theory or discipline, but rather to a discourse about science.

The word *paradigm* has multiple meanings and an ancient origin.

Plato used it as a synonym for model; for Aristotle, however, a paradigm is an example. Margaret Masterman [181] identifies 21 different meanings associated with this term, often in contradiction from one another. Therefore, we have to clarify what Isabelle Stengers means by paradigm. It is the author herself who defines the concept, referring to the work of Thomas Kuhn: "a systematic articulation between a set of practical and conceptual tools and an *a priori* definition of the object and its rules of experimental manipulation" [249, 62]. In *The structure of scientific revolutions*, Thomas S. Kuhn [155] brings back to the top of the scientific debate the concept of paradigm. He wonders how the sciences progress over time. In the traditional conception, scientific progress is considered as a linear and progressive accumulation of new knowledge in addition to that already acquired. However, this process is sometimes interrupted by revolutionary moments, which mark a break with the past and the beginning of the construction of a new knowledge. After these scientific revolutions, the problems that are object of scientific investigation and the criteria for assessing these problems and proposing potential solutions change. There is a re-orientation of the discipline that consists in the transformation of its conceptual structure. Thus, a new paradigm consists in the passage from an old cognitive structure to a new one. It is a guide that defines an orientation and the criteria of a discipline.

As Isabelle Stengers states, complexity is not affirmed in the context of a scientific revolution leading to profound changes in its conceptual structure. It is not a new concept. It is not a revolution. On the contrary, complexity is affirmed in the context of a science that precisely questions a pair of concepts that guides the evaluation of reality: the *simple/complicated* pair. This pair is linked to the concept of paradigm as a model that represents the relationship between concepts and the possibility of experimenting with them. Complicated often has a negative meaning: it identifies the limits of the human capacity to know, the impossibility of using tools that allow the perfect understanding of simple systems. If we cannot understand a phenomenon through these tools, we define it as complicated and implicitly give it a negative meaning. Closely linked to this idea is the rejection of everything that we cannot understand, according to the canon of simplicity, and its labelling to *non-scientific* and *only subjective*. Some phenomena seem complicated because we observe them from our exclusive point of view, which is stable, whereas reality is a perpetual becoming. As previously written, complexity removes the idea of a simple Nature regulated by immutable rules, making the myth of objective knowledge disappear. "Our method seeks to envelop the phenomenon (observation), to recognize the forces within it (praxis), to provoke it at strategic points (intervention), to penetrate it by individual contact (interview), to question action, speech, and things. Each of these methods poses the fundamental methodological problem: the relationship between the research worker and the subject. It is not merely a subject-object relationship. The *object* of the inquiry is both object and subject, and one cannot escape the inter-subjective character of relations between men" [201, 259].

As Isabelle Stengers argues, the notion of complexity overturns the perspective between objective and subjective, questioning the objective categories derived from the simple model. Classical science favours the simplification of systems and, consequently, develops tools suitable, or rather adaptable, to these systems. In this sense a paradigm of complexity does not exist. The discovery of complexity does not correspond to the answer to a problem, but rather to the *awakening of a problem*. Complexity theory is not in antithesis with classical science, but is complementary to it. Classical science is not wrong, but insufficient. Complexity is a change of point of view. The description of systems based on simplification is poor, suitable for borderline cases. Systems must be analysed from multiple points of view. We can consider complexity a new lens to observe reality in order to grasp its multiplicity and dynamism.

Contrary to the position of Isabelle Stengers, other authors have argued that complexity can be considered a new paradigm of science. Edgar Morin [205] describes a paradigm of complexity as opposed to that of simplicity. "The paradigm or simplicity puts order in the universe and chases out disorder. Order is reduced to one law, one principle. Simplicity can see either the one or the many, but it can't see that the One is perhaps at the same time Many. The principle of simplicity either separates that which is linked (disjunction), or unifies that which is diverse (reduction)" [205, 39]. In this perspective, the aim of scientific knowledge is to reveal the simplicity hidden behind the apparent multiplicity and disorder of phenomena. For Edgar Morin, a re-organization of the structure of knowledge is fundamental, a switch from the simplistic-reductionist paradigm of classical science to that of complexity. This real epistemological transition will take place precisely in those natural sciences, which wanted to impose their laws and methods for the study of social phenomena. As the French scholar points out, this is a real paradox. While the social sciences take as a model the method proper to the natural sciences, within the latter emerges the criticism of the investigative model based on reductionism and on the principle of causality. What Thomas Kuhn called a *paradigm shift* is realized: there is a change in the type of logical relationship between master notions, key notions, key principles. A series of discoveries in the world of physics (principle of energy degradation, developments in thermodynamics, quantum mechanics, terror of the expansion of the universe) undermine the idea of an order that would unify all physical and social phenomena under the same universal laws. In particular, it is the principle of entropy that undermines the concept of order by introducing disorder and probability. The most probable configurations of the systems are the disordered ones: entropy indicates the direction of events and is a measure of the disorder to which all systems and organized beings tend. The organization of the systems is born at the expense of the entropy of the surrounding environment. The articulation between closed systems and open systems leads to the conclusion that the decrease in entropy of a subsystem, that is organized starting from disorder, occurs at the expense of the overall entropy of the universe, which increases. According to Edgar Morin, order and organization can arise from disorder: this is especially true for living systems. Complexity is a weave of heterogeneous constituents inseparably associated: it poses the paradox of the one and the multiple. Approaching knowledge by rejecting disorder and uncertainty, by selecting the elements of order and certainty and by removing ambiguity, risks not allowing us to grasp the true nature of phenomena. A method is not valid if it does not include complexity. We need a method that helps us think about the complexity of reality, instead of dissolving it. Simplification is wrong: we must think that the simple and the complex are linked. The paradigm of complexity requires the understanding of the relations between the whole and the parts, not the reduction. "The knowledge of the parts is not enough, the knowledge of the whole as a whole is not enough, if one ignores its parts; one is thus brought to make a come and go in loop to gather the knowledge of the whole and its parts" [204, 6].

Fritjof Capra and Pier Luigi Luisi [59] state that a full understanding of the main problems of our time (like energy, environment, climate change, food security) requires a new conception of life, a change of worldview in science and society. We cannot understand those problems in isolation, because they are systemic, i.e. all interconnected and interdependent. The problem is that most people have an obsolete perception of reality, linked to the classical approach to knowledge and inadequate to deal with complex social phenomena. Starting from the Kuhnian definition, Fritjof Capra defines the *social paradigm* as "a constellation of concepts, values, perceptions, and practices shared by a community, which forms a particular vision of reality that is the basis of the way the community organizes itself" [57, 6]. To analyse those social issues, it is necessary a paradigm shift, a radical shift in our perceptions, our thinking, our values, from a mechanistic to a holistic and ecological worldview. It is a shift from seeing the world as a machine to understanding it as a *network* [58].

Beyond the different positions on the issue, there is no doubt that the concept of complexity has led to a number of important innovations in the relationship with knowledge. In particular, the need for a new way of looking at reality emerges; the importance of going beyond empirical evidence, trying to grasp at the same time the whole and the individual components that make it up. We can discuss whether these transformations coincide with an actual paradigm shift or not; however, their importance is not in question at all.

2. Measurement of complex phenomena

If you haven't measured something, you really don't know very much about it.

Karl Pearson, *attributed*

I wanted to open the chapter on measurement with this statement attributed to Karl Pearson [261], who emphasises the importance of this process for the knowledge of the world. Measurement is a topic often ignored by researchers, who consider it residual compared to others. However, statistics is the science that studies social phenomena through mathematical methods. In other words, statistics aims to produce scientific knowledge by measuring reality. Scientific knowledge develops as a *dialogue between logic and evidence*, through two levels of scientific analysis, linked together even if analytically distinct [167]:

- a *theoretical-formal level*, in which theories and hypotheses are developed and abstract concepts with their mutual relations are specified;
- an *empirical level*, in which hypotheses are verified through empirical data.

Knowledge is, therefore, the result of a complex interaction between theory and observations represented and realized by measurement. This interaction is necessary and unavoidable. Each observation evaluated within a theoretical framework represents a datum. Any empirical observation can be used to generate many different types of data, according to different theoretical frameworks. The framework within which each observation is evaluated is a system for comparing an observation with one or more models. These models are identified by a dimensional system based on an unambiguously defined unit. The relationship between the model and the observed is the product of the measurement.

Thus, we can consider measurement as the application of a formal model to a property of a series of empirical objects. That model can represent reality at different levels of accuracy. If it provides a faithful image of an empirical system, then the logical implications must be comparable with the observable behaviour of the objects. If empirical observations are consistent with model-based predictions, then it can be concluded that the model provides an acceptable description of that segment of reality. Measurement represents a verification of theories as it involves the analysis of the goodness-of-fit of an abstract model to a property owned by a series of empirical objects. Consequently, it can be said that measurement is never immutable: it is always an attempt to define an affirmation about the nature of reality. In this sense, measurement systems are *falsifiable*[1] and it is not possible to prove the truth of the measurement systems, because there will always be a context in which the defined system can be inconsistent with the real phenomenon.

The topic of measurement is addressed in this Chapter starting from its general definition and the identification of its characterizing aspects (paragraph 2.1). Then, I analyse this process within the framework of sociology, focusing in particular on *operationalisation* (paragraph 2.2). Paragraph 2.3 describes the main steps involved in the creation of a *system of indicators*, focusing in particular on the hierarchical design (paragraph 2.3.1), the analysis of different models of measurement (paragraph 2.3.2) and the scale of measurement (paragraph 2.3.3).

[1] Falsification occurs when the specified properties do not correspond to the real properties investigated.

2.1. Measurement: definitions and main aspects

Generally speaking, measurement can be defined as the evaluation of the extension of something (an object, a property, etc.) in relation to a certain standard, i.e. the unit of measurement. However, there is no univocal definition of measurement in the literature. The concept can be traced back to Book V of Euclid's *Elements*. "Euclid presents a theory about ratios of magnitudes of a quantity and about relations of proportion between such ratios" [195, 25]. This is a purely quantitative definition of measurement, from which it seemed that measurement could only be of quantities. The concept evolves over time. According to Joel Michell [194], we can identify three intellectual strands dominated the theory of measurement:

- the *axiomatic approach*, which specifies as economically as possible the conditions sustaining measurement;
- the *operational approach*, which attempts to define those conditions in terms of directly observable manipulations upon the kinds of entities measured;
- the *representational approach*, which construes measurement as the numerical representation of facts about the entities measured.

All different definitions can be ascribed to these macro-categories. Figure 2.1 reports some of the most relevant definitions in the literature.

In the analysis of that process, we can start from Stanley S. Stevens' definition, according to which measurement is "the assignment of numerals to objects or events according to rules" [250, 677]. From this definition, we can highlight some important aspects of this process. Measuring involves a sort of *translation*, a shift from the plane of reality in which we observe phenomena to the plane of numbers in which we try to encode them. This translation must be *meaningful*, that is, it must reproduce as faithfully as possible in the world of numbers the phenomenon observed in the plane of reality. At the same time this translation is *necessary* for the knowledge of reality, which speaks to us with the *language of numbers*. "To those who do not know mathematics it is difficult to get across a real feeling as to the beauty, the deepest beauty, of nature . . . If you want to learn about nature, to appreciate nature, it is necessary to understand the language that she speaks in" [108, 102–104]. The Richard P. Feynman's statement is equally valid for the knowledge of the social reality. The rules of Stanley S. Stevens' definition must

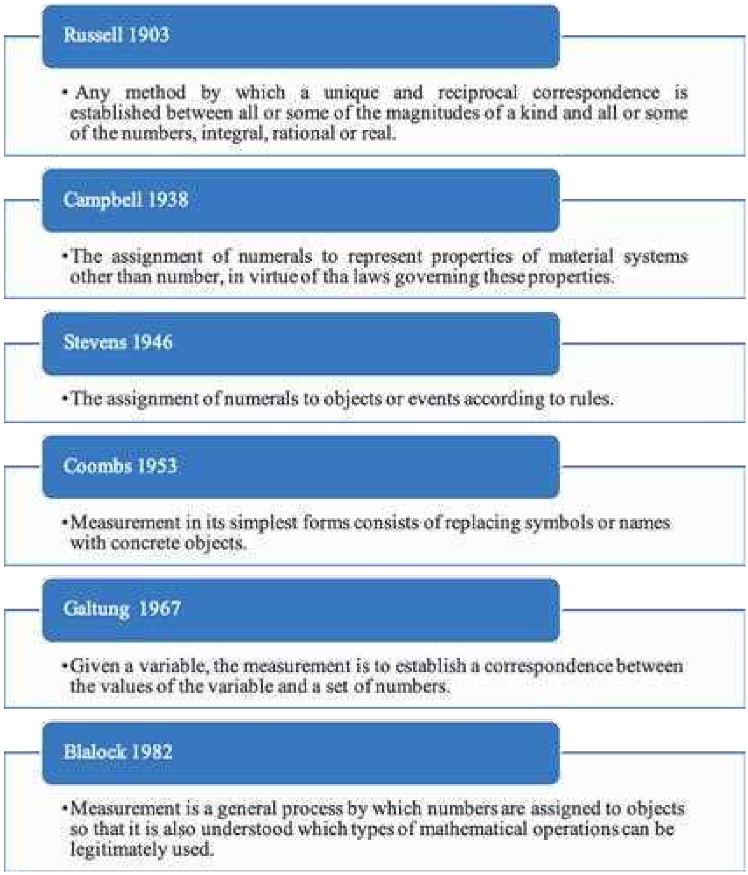


Fig. 2.1. Main definitions of measurement.

ensure that the translation is as faithful as possible. Thus, measurement is the basis of understanding reality: it consists of a selective description using rules consistent with specific purposes [218]. This is another fundamental aspect of the measurement process, being aimed at pursuing theoretical and practical purposes.

Measures must be *standardised*[2], i.e. they must be based on uniform procedures to collect, score and report numeric results. Those procedures must be subject to a verification of its proper functioning. In practice this means that every scientific observation must be preceded

[2] There is not the statistical concept of standardisation that will be described below. In this context, the concept refers to rigor, accuracy and reproducibility of the measurement procedures.

by a series of studies that allow to isolate or at least to minimise the foreign components. These components are not all independent but they are all assumed to be small in size and to compensate for the increase in the number of observations; these elements represent the *error of observation*. Measurement is mainly influenced by two different types of error. The *random error* refers to all those factors that confuse and disturb the measurement of any phenomenon. The higher the random error the lower the level of reliability of the measuring instrument. Variables always contain a random error at different levels; this means that it is the same measurement process that introduces the error component to different extents and the effect of this type or error on reliability can only be estimated. The effects of random errors are totally a-systematic; an instrument affected by such an error may overestimate or underestimate the size measured in a certain object. *Systematic error* influences the ability of the variable to measure what one wants to measure (the theoretical concept); in this sense, systematic error is at the centre of the validity problem, just as random error is at the centre of the reliability problem. The higher the systematic error, the lower the validity of the measuring instrument[3]. Given the presence of measurement errors, it is necessary to verify that the procedures defining a standardised measure have some specific characteristics, the most important of which have been reported in Figure 2.2.

Any measurement process starts with a definition. Measurement definitions can belong to two groups. *Nominal* or *conceptual definitions* relate to the attribution of a meaning to the phenomenon we want to measure. This is a crucial issue, especially in social sciences. As shown by Filomena Maggino [172], almost all measures in social sciences are developed through a *defining process*[4], namely "achieved as a consequence of a definition confirmed through the relationship observed between observations and the concept to be measured" [172, 87]. This is because those phenomena are not directly observable, but they derive theoretically from observations. Phenomena can have different nominal

3 There can be two types of systematic error: the methodological error, i.e. the error of definition/detection of the attribute to be observed; the specific error introduced by the observer in the use of the observation procedure.

4 In addition to the defining process, it is possible to identify two other approaches to measurement: the *fundamental process*, in which the measurement is not derived from other measurements (for instance, length); the *derived process*, in which the measurement is indirectly derived by means of other measures (for instance, velocity) [172, 173, 174].

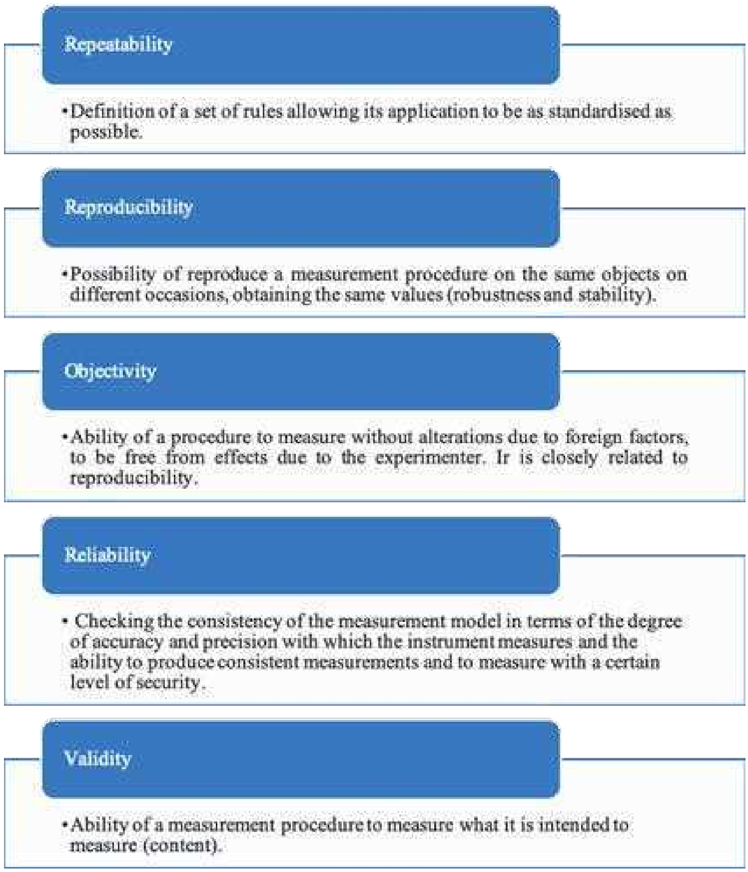


Fig. 2.2. Main characteristics of measurement process.

definitions according to some aspects:

- the nature of phenomena: those phenomena are complex and complexity involves *multidimensionality*; consequently, different definitions can take into account some dimensions of the phenomenon and not others;
- the specific field of study: the same concept can be defined differently by different sciences, considering it from different perspectives or highlighting different aspects;
- the spatio-temporal perspective in which the phenomenon is studied: concepts evolve over time and also change according to the territory where they are defined or respect to which are measured;

- the researcher's point of view: any description of reality is subjective, related to how the researcher views the reality; any conceptual definition is a *window* through which the researcher observe only some aspect of the reality.

No meaning can be attributed without *subjectivity*. However much quality is expected, there will always be the influence of the subject's point of view. As highlighted in the first Chapter, the role of the subject in knowledge production is now clearly recognised. This is particularly evident for socio-economic phenomena. Subjectivity represents one of the dimensions inevitably involved in defining concepts. We can summarise this concept by reporting the famous statement of Protagoras of Abdera: "Of all things the measure is Man, of the things that are, that they are, and of the things that are not, that they are not"[5]. Different researchers analysing the same phenomenon using the same nominal definition and the same indicators may arrive at different conclusions. Nominal definitions do not give any information about activities and operations to measure phenomena. *Operational definitions* indicate exactly how to measure phenomena: what is to be observed; how the observations are to be made; the assumptions governing the process of acquiring observations. The distinction between these two types of definition shows that in the measurement process there are two different *mindsets*. The first is the conceptual mindset, through which researchers define concepts representing phenomena and their relationships. The second one is the empirical mindset, which allows the development of indicators and research steps to investigate phenomena. As Hubert M. Blalock [34] states, the measurement process involves the knowledge of two languages for the researcher: the language of theory and that of research. Both those *languages* are required in scientific knowledge; the conceptual definition and the operational one need to be mutually linked. The relationships between them is complex, even if measurement problems often seem to be simple, especially when we deal with concepts of everyday language. This is misleading. Any measurement process has a complex nature that directly depends on the complexity of the phenomena to be measured.

All measurements rely on a set of assumptions of different kind. Those assumptions concern the connections between empirical obser-

[5] πάντων χρημάτων μέτρον ἐστὶν ἄνθρωπος, τῶν μὲν ὄντων ὡς ἔστιν, τῶν δὲ οὐκ ὄντων ὡς οὐκ ἔστιν [84].

vations, data and methods of analysis and must be developed in order to make measurement useful. We can identify three different kinds of assumptions involved in a measurement process. As the name itself suggests, *theoretical assumptions* relate to the meanings given to the phenomenon measured: appropriateness of definitions, completeness and adequacy of the indicators, variability in the meanings associated with indicators, and many others. As analysed in paragraph 2.3.2, one of the main theoretical assumptions involves the specification of the *model of measurement*. *Procedural assumptions* underlie the rules of correspondence used in assigning numbers to observations. These rules correspond to the operations through which we translate observations into categories or degrees of an attribute representing a concept. Indicators are developed through a step-by-step process that we call *hierarchical design* (paragraph 2.3.1). *Statistical assumptions* deal with the main characteristics of indicators and the statistical methods that can be used for their analysis. One of the main assumptions concerns the *level of scale of measurement* of the indicators considered. This determines what statistical methods can be used and the levels of precision in measurement, i.e. the amount of information available. That question will be examined in paragraph 2.3.3. All those assumptions must be sufficiently general to be applied to any situation a researcher may encounter. The violation of these assumptions generate problems with *validity*, i.e. the risk of measuring something different from what we are supposed to measure (see Figure 1.4). This is a very crucial issue in measurement. According to Richard A. Zeller and Edward G. Carmines [276], even if the concept seems to be simple, there are many different types of validity:

- *Face Validity*

  It is determined by the apparent and external significance that a measure presents. To assess it, expert judgements are required regarding the validity that the measures seem to have.

- *Content validity*

  The extent of which a set of items cover the content of some concept of interest. It refers to the ability of indicators to accurately represent the universe of measured content; even this type of validity requires expert judgement. For example, a test to verify the mathematical preparation of a group of students composed only of questions regarding the sum, is unlikely to be recognized as valid in its content.

- *Construct validity*

  It is given by the level in which the indicators accurately measure the theoretical constructs that interest to measure. For example, we want to build a tool to measure the bullying tendency of a particular group of individuals; according to the hypothesis, the index will be a good measure of bullying (construct validity) if it will be inversely correlated with self-esteem; in order to proceed in this way it is necessary to assume that the theory is correct, i.e. that there is a relationship between the self-esteem and the tendency to bullying.

- *Criterion-related validity*

  It concerns the correlation between a measure and some criterion variable of interest. It is determined by observing how much the indicator correlates with another measure that the researcher considers valid in measuring the same construct. The verification of such validity is done starting from its adequacy in correlating with an external criterion. The presence of a statistically significant relationship is considered a verification of validity. The external criterion can be measured at the same time (*concurrent validity*) or subsequently (*predictive validity*).

- *Concurrent Validity*

  It is determined by observing how much the instrument correlates with other instruments that the researcher considers valid in measuring the same characteristic; the observation of a statistically significant relationship is considered a verification of validity. It is a specification of the criterion-related validity. For example, if we want to verify whether a particular measure is a valid instrument to measure self-esteem, it is possible to correlate the scores obtained by the subjects by using this instrument, with the scores obtained by the same subjects by using a different scale considered a good measure of self-esteem. A high correlation highlights the validity of the new test.

- *Convergent Validity*

  It is determined by comparing and correlating the scores obtained with the measurement to be validated with those obtained with the measurement of another construct, theoretically related to the first. The possibility of verifying the convergent validity, therefore, depends on the existence of constructs, and relative measures, linked with the one measured.

In social sciences, it is particularly important to focus on what Lee J. Cronbach and Paul E. Meehl [71] define as *construct validity*[6]. We must investigate it when there is no criterion of content accepted as fully adequate to define a concept we want to measure. "Construct validation takes place when an investigator believes his instrument reflects a particular construct to which are attached certain meanings. The proposed interpretation generate specific hypotheses, which are a means of confirming or dis-confirming the claim" [71, 290]. This type of validation aims at the assessing if a specific measure relates to others consistently with a theoretical framework. For construct validation, first we must specify the theoretical relationship between the concepts themselves. After, we must analyse the empirical relationship between the measures of the concepts. Finally, we must interpret the empirical evidence in terms of how it clarifies the construct validity of the particular measure. The only way to assess construct validity is *within a specific theoretical context*.

2.2. Measurement in sociology

This research work deals specifically with measurement in sociology. In that field, dealing with measurement means dealing with indicators. Almost all social phenomena are complex, their measurement will also have to consider this complexity. The aim of measuring social phenomena should be to understand them in their nature, as something different from the simple sum of their parts. Understanding each of them as a *whole*. As previously written, knowledge is the result of a complex interaction between theory and observations represented by and realized through the measurement. This interaction is the basis of *scientific research*, which we can define as a creative process of discovery that develops according to an established itinerary and procedures that are consolidated within the scientific community. There is no contradiction between creativity and the presence of established procedures.

[6] Constructs can be defined as "complex concepts that are inferred or derived from a set of interrelated attributes of people, objects or events, typically embedded in a theory and often not directly observable, but measurable using multiple indicators" [218, 2].

The attribute *creative* refers to the personal capacities of the researcher, to that subjectivity which is a relevant aspect of any scientific research. Hans Reichenbach [234] introduced the classical distinction between *context of discovery* and *context of justification*. The first stage is not subject to rules and procedures: it is not possible to define logical and invariable rules that allow to carry out the creative function. Scientific work is not just about producing new theoretical hypotheses. The scientist must also test them. The phase of the context of justification consists precisely in the empirical verification of theories, which must be done following specific rules. Knowledge is made up of "empirically confirmed and logically consistent statements of regularities" [191, 270]. The first rule of empirical research is that it must develop within a collectively shared framework. "Science is public, not private" [193, 26]. This public-collective nature of science responds to a twofold need. On the one hand, it implies *control*: concepts and procedures must be standardised and the results obtained must be reproducible. On the other hand, it implies *cumulability*, the systematic accumulation of knowledge.

An essential contribution to the study of measurement in sociology was made by Paul F. Lazarsfeld. In his famous article *Evidence and inference in social research* (1958), he starts from the consideration of the specificity of measuring social phenomena. "When social scientists use the term measurement it is in a much broader sense than the natural scientists do" [162, 100]. The author, in practice, points out that in the social sciences measurement has a typical character, not comparable to that of the natural sciences where measurement processes are all fundamental or derived (see paragraph 2.1). It is therefore necessary to examine the procedure used by social scientists to characterize their object of study. Paul F. Lazarsfeld identifies four steps in the process by which concepts are translated into empirical indices:

1. *Imagery of the concept*

   The definition of a measuring instrument usually begins with this step. By immersing himself in a theoretical problem, the researcher creates a rather vague image of phenomenon (construct). The construct may often be the result of the perception of many heterogeneous phenomena having some underlying characteristic in common; the researcher tries to account for them. "In any case, the concept, when first created, is some vaguely conceived entity that makes the observed relations meaningful" [162, 101].

2. *Concept specification*

   In this phase, imagery is divided into components, called *dimensions*. They are specifications of the original construct that are achieved through an elaborate analysis of the phenomena. They can be derived logically from the overall concept or one aspect can be deduced from another. The concept is, therefore, a complex combination of phenomena, rather than a simple and directly observable item. "...every concept we use in the social sciences is so complex that breaking it down into dimensions is absolutely essential in order to translate it into any kind of operation or measurement" [162, 102].

3. *Selection of indicators*

   The third step is the selection of indicators for each dimension identified. At this stage, the researcher has to address some problems. The first problem is understanding what an indicator is. Paul F. Lazarsfeld affirms that indicators are directly suggested to researchers by common experience and "each indicator has not an absolute but only a probability relation to our underlying concepts" [162, 103]. The relation depends on the definition of the concept. In this sense, an indicator is a *purposeful statistics* [133]: it is not simple crude statistical information but represents a measure organically connected to a conceptual model. A statistical index becomes an indicator only when its definition and measurement occur in the ambit of a conceptual model and is connected to a defined aim[7]. A measure can be defined an indicator only if it is within a theoretical framework[8]. "Indicator is what relates con-

---

[7] Even though the terms indicator and index are often used in an interchangeable way, they have different meaning. Index comes from the Latin word *index*, which means "anything that is useful to indicate". In statistics, it represents historically a very generic term applied with multiple meanings. Indicator comes from the Latin word *indicator*, which means "who or what indicates"; in statistics, it represents a more recent term indicating indirect measures of economic or social phenomena not directly measurable.

[8] Kenneth C. Land [159] states that a statistical index can be considered an indicator when:

- it represents a component in a model concerning a social system;
- it can be measured and analysed in order to compare the situations of different groups and to observe the direction (positive or negative) of the evolution along time;
- it can be aggregated with other indicators or dis-aggregated in order to specify the

cepts to reality" [172, 92]. Consequently, a wide variety of possible indicators can be identified to measure a specific dimension of a concept. This raises another question: how many indicators should we consider? There is no *correct* answer to that question. Generally speaking, we need to choose a number of indicators that allow us to adequately represent the desired conceptual dimension, avoiding redundancy and ensuring the reduction of error. The issue will be better addressed in the following pages.

4. *Combination of indicators into indices*

   This is the last step; we must "to put Humpty Dumpty together again" [162, 104]. The concept needs to be reconstituted. All the indicators that we have collected and used have produced data; at this point, a synthesis of the indicators must be made. As shown in the first Chapter, synthesis is the only way that allows us to have a meaningful view of social phenomena. We will deal in detail with this topic in the third Chapter. However, it should be pointed out immediately that the synthesis, in this context, is the result of the application of statistical methods to data. The choice of methods will have an influence on the procedure and on the results obtained.

This process allows the empirical translation of the theory, the so-called *operationalisation*, i.e. the process through which (abstract) concepts are translated into (measurable) variables. The variable is, therefore, an operationalised concept; more precisely, it consists of the operationalised property of an object, since the concept, in order to be operationalised, must be applied to an object. Between concept, property and variable there is the same link that exists between the weight (concept), the weight of an object (property) and the weight of an object measured through the balance (variable). The language of social research is the language of variables; they are the *social sciences vocabulary* [48]. It is important to underline the extremely arbitrary nature of the operational definition: the way in which the researcher decides to operationalise a concept is absolutely questionable. There is no absolutely correct definition: the decision on how to operationalise depends only on the researchers choices. Moreover, each operational definition always entails

model.

a limitation of the concept. In fact, there will always be a gap between the variable and the concept. Finally, it should be noted that the process described by Paul F. Lazarsfeld is typical of quantitative sociological research. In fact, in qualitative research there is no equivalent to the operationalisation of concepts. Qualitative research moves in a different way. The concept is not operationalised, but used as a *sensitising concept* [39], i.e. as an orientation towards research. Sensitising concepts are considered by researchers as interpretive devices and as a starting point for a qualitative study, drawing attention to important features of social interaction and providing guidelines for research [49]. According to Herbert Blumer, the concepts must all be sensitising and not definitive, i.e. they do not provide prescriptions of what to see, but merely suggest directions along which to look. "The metaphor that I like is that of lifting the veils that obscure or hide what is going on. The task of scientific study is to lift the veils that cover the area of group life that one proposes to study. The veils are not lifted by substituting, in whatever degree, preformed images for firsthand knowledge. The veils are lifted by getting close to the area and by digging deep into it through careful study" [40, 39]. Herbert Blumer's statement can be extended and generalised from the specific perspective of symbolic interactionism to the general one of qualitative sociological research.

2.3. Developing indicators to measure complexity

Investigating different aspects related to social phenomena requires the definition of basic indicators representing what is actually measured with reference to the corresponding dimension [168, 172]. In social field, the measurement process is associated with the development of indicators. The latter is a *normative exercise* since:

- indicators are related to a conceptual definition;
- a phenomenon can be defined in different ways.

Consequently, in order to describe a phenomenon, different group of indicators can be selected. The normative nature of the selection of indicators is an aspect of the subjective component involved in any measurement process that, as previously written, cannot be excluded. Moreover, in the selection of indicators many *subjectivities* are involved. "…term *subjective* changes its value with reference to the context in which it is used" [172, 89]. The definition of phenomena is subjective.

Describing reality always depends on the researchers' point of view. Conceptual frameworks represent *small windows* through which only some aspects of the reality can be observed. The definition of the hypotheses on reality is pervaded by subjectivity: researchers, through the dialogue with the working hypothesis, can change perspective in a path of knowledge in continuous evolution. Subjectivity refers also to the kind of information which has been defined in the ambit of a conceptual framework and subsequently observed. We can distinguish between *objective* information, collected by observing reality and *subjective* information, collected only from individuals and their assertions. Thus, we can have *objective indicators*, based on explicit criteria shared by external observers, and *subjective indicator*, based on subjective evaluations and criteria which can vary from one individual to another. In summary, this process cannot be considered arbitrary, since it always involves a relationship with the reality. Many times, in the name of "objectivity", technical choices are done in an arbitrary way. Given the complexity of such a reality, we can consider data as a *fragmented text*; the researcher must read this text looking for a sense. This *sense structuring* process is not an arbitrary one, but necessarily involves some subjectivity [172].

Now, we describe the main steps that allow to develop a measurement system based on indicators.

2.3.1. Hierarchical design

Indicators should be developed through the so-called *hierarchical design*, an implementation of the Paul F. Lazarsfeld' model described in paragraph 2.2. The hierarchical design is a step-by-step process and requires the definition of different components

The starting point is the definition of the concept. We have already analysed that all social measurement processes are based on a robust conceptual definition (almost all measures in social sciences are developed through a defining process; see paragraph 2.1). In this step, the researcher must ask himself: *what is the phenomenon to be studied?*. It is not a simple question. In fact, it may not be easy to define phenomena such as well-being, quality of life, poverty, gender inequality and so on. It is also necessary to remember the subjectivity inherent to any definition. A good starting point is not to rely on common sense, but to seek out what other researchers have done. However, evaluating the objectivity and quality in the definition of the phenomenon only by considering the reference to the literature selected by the researcher is not completely correct. In fact, also the literature selection is a subjective activity. By means of conceptualisation, it is possible to define models to construct and evaluate data and defining their temporal ambit and territorial disaggregation level. Concepts (and their variables and dimensions) could be observed within *domains*. A domain represents a facet of the reality in which the concepts should be monitored and assessed[9]. Obviously, there is not a list of domains valid for all socio-cultural contexts, because the list depends on societal values, even if some researchers highlighted that certain domains are always present in different studies (see, for instance, the so-called Stiglitz-Sen-Fitoussi report [251]).

The second step is the identification of *latent variables*. Each of them represents an aspect to be observed and confers an explanatory relevance onto the corresponding defined concept. Latent variables reflect the nature of the phenomenon consistently with the conceptual model. Their identification is founded on theoretical assumptions requiring a fundamental analysis of the literacy review, also about its *dimensionality*. Based on its level of complexity, the variable can be described by one or more factors. The different factors of each variable are referred as dimensions. Thus, we can have two different situations [172, 91]:

1. *uni-dimensional*, if the definition of the considered variable assumes a unique underlying dimension;

2. *multidimensional*, if the definition of the considered variable assumes several underlying factors.

This step influences the selection of indicators. The selection of latent variables and dimensions can be a particularly demanding exercise, especially in the case of social phenomena.

The last step is the selection of basic indicators to measure the defined variables. Each latent variable could be defined and measured by a single indicator. This *single indicator approach* is weak and assumes the existence of a direct correspondence between one latent variable and one indicator (in other words, the possibility of measuring one dimension with just one indicator). It is preferable to adopt a *multi-indicator*

[9] Typical examples of domains are housing; health, transport, environment, leisure and culture, social security, crime and safety, education, labour market, working condition, and so on.

approach, consisting in using several indicators for each conceptual dimension to cover the conceptual dimension's variability. This approach allows the overcoming (or, at least, the reduction) of problems produced by the single indicator approach. In fact, using multiple indicators increases measurement accuracy and precision, allowing to compensate the random error.

By applying accurately the hierarchical design, we define a *system of indicators*. This is not a simple collection of measures. It is a complex system and as such it has all its fundamental characteristics (see paragraph 1.2). Indicators within a system are interconnected and new properties typical of the system and not of its constituent elements emerge from these interconnections. In particular, a system of indicators allows the measurement of a complex concept that would not otherwise be measurable by taking into account the indicators individually. Obviously, indicators change over time as the concept to be measured changes and evolves. Therefore, a system of indicators is *Complex Adaptive System* (CAS) used to measure a *Complex Adaptive System* (a concept).

2.3.2. Measurement models: formative and reflective

One of the main theoretical assumptions involves the specification of the *model of measurement* referring to the relationship between constructs 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 [211, 88][10]. 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* [33, 43, 83, 82, 81, 172]. 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 [166, 46]. Therefore, causality is from the construct to the measures. Specifically, the latent variable η represents the com-

[10] We must clarify that measurement models are different from structural models. A measurement model describes relationships between a construct and its measures (indicators), while a structural model specifies relationships between different constructs. "The reason for drawing a distinction between the measurement model and the structural model is that proper specification of the measurement model is necessary before meaning can be assigned to the analysis of the structural model" [17, 453].

mon 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 2.1:

$$x_i = \lambda_i \eta + \epsilon_i \tag{2.1}$$

where x_i is the i-th indicator of the latent variable η, ϵ_i is the measurement error for the i-th indicator and λ_i is a coefficient capturing the effect of η on x_i. Measurement errors are assumed to be independent (i.e., $\mathrm{Cov}[\epsilon_i, \epsilon_j] = 0$, for $i \neq j$) and unrelated to the latent variable (i.e., $\mathrm{Cov}[\eta, \epsilon_i] = 0$, 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 [42]. Each indicator has a specific error component. Typical examples of reflective scenarios include measures of attitudes and personality. Let's suppose we want to measure the intelligence of a group of individuals using the results obtained by each of them in a series of tests. In this hypothesis, it is quite evident that the intelligence of each individual influences the result of the tests and not vice versa. As a consequence, we expect that the results of an individual to the different tests are quite the same and, from a statistical point of view, correlated with each other (because they are determined by the same latent variable). If a test gives a completely different result, it does not measure that specific construct. Figure 2.3 shows the main components of reflective models and their relationships.

We can say that the concept of formative models owes much to the analysis of the factors underlying the *puzzling variables* such as discrimination, prejudice, personal feelings. Such variables inevitably force a conscientious scholar to face questions related to their origin and nature. For example, let's consider of a concept such as discrimination against immigrants. Where does it originate from? An aversion to immigrants can be linked to (and *caused by*) several factors, from ideas and attitudes received from the primary group (a person's relatives and childhood contacts) to personal experiences, the impact of secondary group's members or the influence of mass-media. All these factors would concur to shape the attitudes towards immigrants, but they are not necessarily

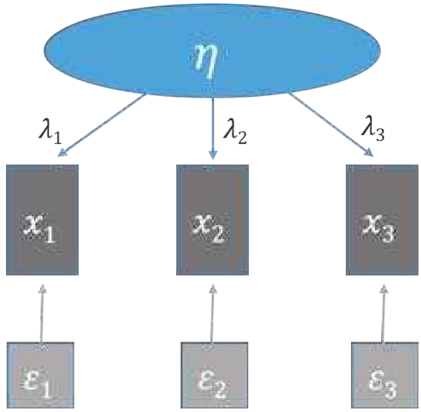


Fig. 2.3. Reflective measurement model.

independent of each other. For instance, different attitudes received from the primary group would affect the responsiveness to the attitudes of the secondary group, and so on and so forth. Émile Durkheim can be considered a pioneer of the use of formative models, although he did not pose the theoretical problem of measurement. Reading his study on the suicide [11] [89], one would identify some of the situations concerning these puzzling variables [248]. For instance, observing increase in the number of suicides in summer, Enrico Morselli had advanced the hypothesis that there was a link between the suicide rate and temperature. Émile Durkheim did not accept this explanation and, moving from the evidence of the seasonality of suicides, suggested that there were other factors – associated with both the season and its high average temperatures – affecting the seasonal variance in the suicide rates. These factors regarded, according to Émile Durkheim, the concentration of economic activities and social contacts (with all their consequences for the individual's balances), in summer. Émile Durkheim, as mentioned above, does not conceptualise the problem of measurement, but com-

11 In this research work, Émile Durkheim reports the results of a secondary analysis carried out on a series of statistics collected in the main European countries between 1841 and 1860 or on material already published on the subject, such as the work of Enrico Morselli [206]. Émile Durkheim applied the term *suicide* to all cases of death resulting directly or indirectly from a positive or negative act of the victim himself, which he knows will produce this result. He identified four types of suicide (altruistic, egoistic, fatalistic, anomic) based on the degrees of imbalance of two social forces: social integration and moral regulation. He concluded that the more socially integrated and connected a person is, the less likely he or she is to commit suicide. As social integration decreases, people are more likely to commit suicide.

pared the variables by examining their relationships in a systematic way and under different conditions (think, for example, of his suggestions about the combined effect of education and belonging to a (religious) community on suicide rates). In doing so, he points out that of his suggestions about the combined effect of education and belonging to a (religious) community on suicide rates. Officially, the formative measurement model was proposed for the first time by Richard F. Curtis and Elton F. Jackson [73]. 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 [33, 34, 158] 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. Let's suppose we want to measure the gender inequality. We must start with its definition: we can say that it refers to systematic differences in the outcome of men and women on a variety of issues ranging from economic participation and opportunity, political empowerment, and educational attainment to health and well-being [235, 14]. In this case, by means of the definition, we already identify the components that form the concept and, consequently, the indicators to be selected. According to this definition, a measure of the gender inequality must take into account economic participation and opportunity, political empowerment, and educational attainment to health and well-being and use at least one indicator to measure each of them. If one of these dimensions is not taken into account, the concept of gender gap changes. The model is specified as follows:

$$\eta = \sum_{i=1}^{n} \gamma_i x_i + \zeta \tag{2.2}$$

where γ_i is a coefficient capturing the effect of indicator x_i on the latent variable η, and ζ 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., $\mathrm{Cov}[x_i, \zeta] = 0$)[12]. Indicators are not interchangeable; thus, omitting an indicator is omitting part of the construct (this changes

[12] Equation 2.2 represents a multiple regression equation and, in contrast to equation 2.1, the latent variable is the dependent variable and the indicators are the explanatory variables.

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 [42]. Indicators have no specific measurement error terms [99]; in formative models, we only observe disturbance term (ζ) un-correlated with x_i [98]. Almost all measurement processes of socio-economic phenomena adopt a formative model. Figure 2.4 shows the main components of formative models and their relationships.

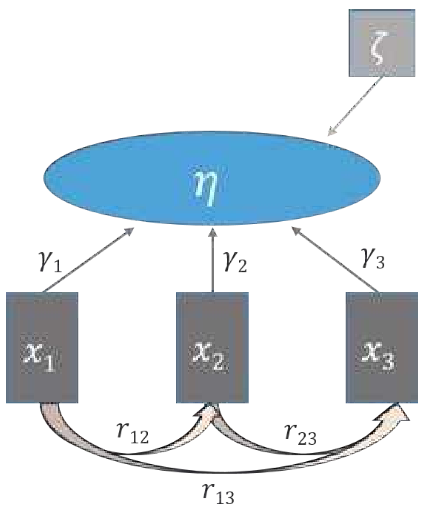


Fig. 2.4. Formative measurement model.

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 [135, 271, 98, 2] 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. At the same time, the result of an incorrect specification of the model is also evident. Despite the growing attention and use of formative models " ...researchers in the social sciences assume that indicators are effect indicators. Cause indicators are neglected despite their appropriateness in many instances" [43, 65]. The main reason is that there are some controversial issues regarding the conceptualisation, the estimation and the validation of formative measures. For instance, the treatment of multicollinearity is a fundamental topic. Indeed, the presence of multicollinearity is undesirable in formative models as it causes estimation difficulties [81]. By looking at the equation 2.2, the consequence of correlations among formative indicators in unstable estimates for the indicator coefficients γ_i and, consequently, it becomes difficult to separate the distinct influence of individual indicators on the latent variable η. Moreover, multicollinearity can lead to validity problems [83]. Generally speaking, multicollinearity is particularly worrying when the scientist is interested in estimating the contribution made by the single indicator to the phenomenon that must be explained; it is less worrying when the scientist is interested in estimating the contribution made by the whole set of indicators. In the literature, different approaches for dealing with multicollinearity are proposed. Starting from the consideration that indicators highly correlated quite likely contain redundant information [46], a possible approach is the indicator elimination based on the variance inflation factor (VIF), which assesses the degree of multicollinearity[13]. However, using a purely statistical criterion to consider multicollinearity and eliminate some indicators can lead to change the meaning of construct. For this reason, other approaches have been proposed to treat multicollinearity[14]; the most common strategy is not to eliminate correlated indicators (so as not to influence the construct). Important scholars support the validity and effectiveness of the formative models [22, 44, 81, 45]. The debate continues in the 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 [99]. 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 [64]. The choice of the measurement model should be guided only by its appropriateness to the phenomenon one intends to study.

13 Studies adopting this approach usually apply the commonly accepted cut-off value of VIF $N > 10$ or its tolerance equivalent.

14 For instance, Sönke Albers and Lutz Hildebrandt [16] propose combining formative indicators into an index (using an arithmetic or geometric mean) and using the latter as a single-item construct in the subsequent analysis.

2.3.3. Levels of scale of measurement

Indicators within a system can be different, i.e. they can belong to different levels of scale of measurement. As highlighted by Brian E. Perron and David F. Gillespie [218], the latter refers to the amount of information available and the relationships between the modalities of an indicator. This is a crucial issue, since the properties of the indicator determine the type of statistical tool that can be used to study it and consequently influence, as we will see in the third Chapter, the choice of method of synthesis for a system of indicators. This subject is often underestimated. We often see studies that deal with nominal or ordinal variables as if they were cardinal variables, using for their synthesis tools that are inappropriate to their level of scale (for instance, the arithmetic or geometric mean).

The main theory of measurement scales was proposed by Stanley S. Stevens [250][15]. All measurements in science are conducted using four different types of scales: *nominal*, *ordinal*, *interval* and *ratio*. These four levels correspond to increasing degrees of precision in measurement and are based on a set of assumptions about how precisely the numbers describe the measured attribute and what statistical tools and mathematical proprieties are appropriate for each level.

- *Nominal scale level*

  A variable at nominal scale level occurs when the property to be recorded assumes:

  - discrete states, i.e. a series of finite states;
  - un-orderable states, i.e. it is not possible to establish a hierarchy between these states.

  The only relations that can be established between the modalities are of type $=$ or $\neq$. The operation of dividing a property into unordered categories (classification) consists in simply associating a name to each category. Nominal measurement is defined by mutual exclusive and exhaustive classification. *Mutually exclusive* means that any rating of an object automatically excludes other alternatives. *Exhaustive classification* means we have the entire set of

[15] Although Stanley S. Stevens' theory is widely adopted, it has been and continues to be contested by other scholars [230, 263]. Alternative scale taxonomies were proposed in the literature [207, 61].

categories that defines our concept completely. Nominal measurement is a primitive level of classification and is not to be confused with nominal definitions. Classification is the first and most basic operation in all sciences. The categories for nominal measurement, whether labelled with words or numbers, are simply different. The nominal scales assume the properties of symmetry and transitivity.

- *Ordinal scale level*

  In this case, the property to be registered assumes *discrete states that can be ordered*. The element that distinguishes this level of scale from the previous one is, therefore, the possibility of an ordering, which allows to establish not only a relation of type $=$ or $\neq$, but also order relations of type $<$ or $>$. However, the distance between the different modalities is unknown. The attribution of values to individual modalities is, therefore, not causal, but must follow a criterion that preserves the order of states. For this reason, natural numbers are almost always used. The numbers have, however, only an ordinal value: they are labels assigned to the modalities preserving the sequence. In addition to the assumptions of symmetry and transitivity, ordinal measurement assumes a single continuum underlying the classification or position of cases. This assumption introduces asymmetry between scale values.

- *Interval scale level*

  Interval scale measurement add to the properties of previous scale levels (mutual exclusive and exhaustive classification, ordering) the presence of *equal intervals between scale values*. While ordinal rating allows only for a comparison of serial position, interval scales allows us to calculate the magnitude of difference because the intervals between scale units have a *standard unit value*. In this case, the modalities have full numerical meaning, in the sense that numbers also have the main cardinal properties. In particular, at this level of scale it is possible to perform operations of subtraction and sum of values, but not of multiplication or division because of the absence of meaningful zero value. "With the interval scale we come to a form that is quantitative in the ordinary sense of the word. Almost all the usual statistical measures are applicable here, unless they are the kinds that imply a knowledge of a true zero point" [250, 679]. In the interval scale level variables there is a *conventional* *zero*; this means that between the modalities of the variable it is possible to perform operations of sum and/or subtraction. On the other hand, all four fundamental arithmetic operations can be performed on the intervals between the values, thus allowing the application of the vast majority of statistical methods. "The zero point on an interval scale is a matter of convention or convenience, as is shown by the fact that the scale form remains invariant when a constant is added" [250, 679].

- *Ratio scale level*

  Ratio measurement is achieved by adding a meaningful zero point to an interval scale. "An absolute zero is always implied, even though the zero value on some scales never be produced" [250, 679–680]. All arithmetic operations can be applied on the values for determining all four relations: equality, rank-order, equality of intervals and equality of ratios. All types of statistical measures are applicable.

The properties of the measuring scales are cumulative, i.e. properties and statistical tools used at a lower scale level can also be used at a higher scale but not vice versa.

2.4. Final considerations

Measuring in social sciences field requires a robust conceptual definition, a consistent collection of observations and a consequent analysis of the relationship between observations and defined concepts. Managing indicators introduces at the same time [172, 111]:

- a *challenge*, represented by the need of dealing with complexity;
- a *need*, given by the need of making indicators relative;
- a *risk*, given by the reductionism.

Indicators are the tools to understand complexity. They play a key role in describing, understanding and controlling complex systems. An indicator is, therefore, a tool for understanding reality. It is not necessarily a number. It can be an object, a map, an image. It is what allows us to grasp the complexity and guide us in understanding it [175]. There is a large amount of literature on the use of metaphoric images for the representation of phenomena, especially for complex ones [254, 165].

According to Theodore M. Porter [223], the *soft* power of numbers and indicators is characteristic of our time. If we hope to use indicators and other measures to make the world navigable in simpler terms, let us be careful what we wish for. It is essential that what we are going to build is an authentic representation of the reality, preserving the systemic characteristics of the phenomena defined by elements and their relationships. In this perspective, each indicator measures and represents a distinct constituent of the defined phenomenon and all of them do not represent a pure and simple collection of indicators but are part of a complex system, a multi-indicator system, in which. In other words, only a complex instrument (a multi-indicator system) allows a full and correct understanding of complexity.

3. Synthesis of multi-indicator systems over time: methodological aspects

The properties of the parts can be understood only within the context of the larger whole... Analysis means taking something apart in order to understand it; system(ic) thinking means putting it into the context of a larger whole.

Fritjof Capra, 1996

As analysed in the second Chapter, the correct application of the hierarchical design develops a system of indicators. The latter is a complex system, the analysis and understanding of which requires approaches allowing more concise views. The guiding concept is *synthesis*. Synthesising data responds to a range of cognitive and practical needs. For example, we can have the objective of knowing what is *behind* the data or how we can use it. Generally speaking, synthesising responds to a need for concreteness in the relation with things. It is justified by the fact that knowledge of complex phenomena involves some form of *reductio ad unum* [237]. The correct way of understanding those phenomena is to conceive them as a whole, adopting a synthetic approach as underlined in Fritjof Capra's statement. Getting in contact with reality always involves some process of synthesis, more or less conscious, consisting in the reduction of a multiple in units. This *reduction* could be a risk. Any synthesis should be a *stylisation* and not an over-simplification of reality. Dealing with systems of indicators, the synthesis must be a meaningful measure, capable of representing the complex system without trivialising or simplifying it.

From the methodological point of view, synthesis concerns different aspects of the system [171]:

- *Synthesis of units*

  The aim is to aggregate the units of observation in order to create macro-units 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 basic 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.

This Chapter focuses on the issue of the synthesis of statistical indicators from a methodological point of view. Before continuing, some clarifications need to be made. Although it has been and still is considered a *niche field*, the topic of synthesis of indicators has a rich and varied scientific literature. There are many approaches that have been developed in the literature, as well as many statistical methods and procedures for synthesising indicators. The aim in this thesis is not to present a review of all the methods, but rather to focus on some specific ones.

It was my precise choice not to deal explicitly with statistical models as a form of representation of the complexity of the systems. The concept of model is repeatedly referred to in this work. However, the term refers to measurement models (see paragraph 2.3.2) and not to statistical regression models. The reasons for not dealing with these statistical tools are different. First of all, they would merit a very broad discussion, which would be the subject of a specific thesis (as many previous works have already done). Even is they are not the subject of this thesis, some clarifications on statistical models need to be made. Although widely used, they are often badly used. For instance, statistical models are often applied for the analysis of complex phenomena without testing the nature of the response variable and other assumptions underlying these same models (for instance, in linear regression models, the presence of a linear relationship, multivariate normality, absence of multicollinearity and auto-correlation, homoscedasticity, etc.). No verification of assumptions can often lead to erroneous conclusions and, consequently, to a misunderstanding of the phenomenon under study[1]. The estimation of the coefficients of a model requires the presence of an appropriate number of units. In the study of social phenomena, we often faced with a limited number of statistical units (as, for example, in the Italian regional analysis covered in the fourth and the fifth Chapters of this thesis), which often makes it impossible to estimate a model. In general, the realistic possibility of knowing reality by using mathematical-statistical models is questionable. As George E.P. Box [50] stated, *all models are wrong*. The researcher cannot obtain a *correct* one by excessive elaboration, but he should seek an economical description of natural phenomena. In a book of 1987, he revisited his statement, affirming that, although all models are wrong, some are useful and less wrong than others [51]. "Modelling in science remains, partly at least, an art. Some principles do exist, however, to guide the modeller. The first is that all models are wrong; some, though, are better than others and we can search for the better ones. At the same time we must recognise that eternal truth is not within our grasp" [188, 8]. The study of reality always involves a loss of information; in this sense, we must choose a model that allows us to limit this loss as much as possible.

In this research work, I take into account synthesis techniques that allow a *dynamic analysis of phenomena* in order to obtain comparable synthetic measures not only in space, but also in time. Only in this way, a synthesis has full meaning. Only this way, we can grasp not only the differences between one statistical unit and the others, but also how that same unit has changed over time and, consequently, how its relationship with other units has also changed over time. I deal with systems where

[1] Just to make an example, in a study on the territorial determinants of voting for Brexit in June 2016, Leonardo S. Alaimo and Luigi M. Solivetti [15] showed that most of the research works on the topic [121, 25] used linear regression models to analyse the effects of a series of explanatory variables on the response variable (the percentage of votes for Brexit in the Local Government Districts). The authors highlighted that the use of a linear model to fit a *fractional response variable* bounded in a range $[0, 1]$, although it is the common tool used by researchers, rarely provides the best description of the response variable and is based on erroneous assumptions. The authors chose the regression model according to the nature of the dependent variable, using a *fractional logit regression model* [214]. Using this model, based on correct assumptions about the nature of the dependent variable, the authors come to different conclusions than other studies.

all indicators are cardinal (according to Stanley S. Stevens' classification, interval and ratio scales, paragraph 2.3.3). There is a large literature on the treatment and synthesis of multidimensional systems of ordinal data using *non-aggregative methods*, allowing the construction of synthetic measures without the aggregation of the scores of basic indicators. Within this approach, the *Partially Ordered Set* (poset) has become a reference over the years, as demonstrated by many works in different fields of research [18, 107, 60, 56, 80, 19, 31, 10, 11]. This method perfectly fits the needs of ordinal data analysis; at the same time, it can also be suitable for quantitative data. Using poset with cardinal data also makes it possible to overcome some problems that often occur in synthesis, like, for instance, the absence of strong interconnections between the indicators considered, which prevents effective size reductions through aggregation procedures [105]. There are some examples of the application of poset on cardinal data in the literature. For instance, Marco Fattore [106] proposes a way to assess and compare the environmental sustainability level of countries by means of poset, using average height and the concept of embedded scales. By following this methodology, Leonardo S. Alaimo et al. [8] propose a well-being measuring procedure based on poset. In particular, using the framework and the system of indicators of the Italian *Equitable and Sustainable Well-being* (BES) released by Istat, the authors defined a synthetic measure (based on poset) for each BES domain, allowing the comparison of Italian regions. Another powerful approach has been proposed to deal with metric data, the so-called *object-based* approach [147, 146]. This approach is focusing on objects, which have a profile, corresponding to the set of indicators and is particularly suitable for metric data and infinite sets[2]. This thesis adds a new step to the analysis proposed by Marco Fattore [106] and Leonardo S. Alaimo et al. [8]. I define non-aggregative synthetic measures based on poset that allow the analysis of complex phenomena over time. To test the validity of the proposed procedure, I compare the results obtained with those of the main aggregative methods for the analysis of phenomena over time, highlighting the strengths of the new procedure.

The starting point is the characterisation of the object of this study, the system of indicators over time, from a conceptual and mathematical point of view (paragraph 3.1). After, I present the exploratory analysis,

[2] This approach will not be specifically addressed in this research work.

the preliminary step in any synthetic exercise (paragraph 3.2); I propose a temporal approach to this analysis. In the paragraph 3.3, I focus on a very important issue, often little addressed in the statistical indicators field, the synthesis of statistical units, focusing on the overview of some time series clustering methods. Paragraph 3.4 deals with the synthesis of statistical indicators, considering both the aggregative and the non-aggregative approach. In the paragraph 3.4.3, I present my new methodological proposal for the synthesis of multi-indicators systems over time based on poset.

3.1. Analysing multi-indicator system: the importance of *time*

As highlighted in the introduction to this Chapter, the objects of this research work's analysis are the system of indicators over time. This is a very crucial issue in the field of synthesis. From a conceptual point of view, we can affirm that all socio-economic phenomena present the typical features of the *Complex Adaptive Systems* - CASs (see paragraph 1.2). This means that those phenomena change over time; these changes may concern different aspects and, consequently, may influence the definition and measurement. Let's think about phenomena like *well-being* or *sustainable development*. Their conceptions not only change from one cultural context to another, but within the same cultural context they change over time. Dimensions not considered before are added, others are excluded. All these changes, of course, entail a consequent adjustment of the systems of indicators *designed* to study these phenomena. Indicators can be added and others can be excluded. These systems are, therefore, *adaptive*. The temporal dimension is conceptually pervasive. By excluding the temporal dimension and analysing the indicators from a cross-sectional point of view, it merely *photographs* the observed reality, making effective the reductionist risk. The simple photograph of reality provides only a partial explanation of it; we cannot know the causes of the situation observed, whether it is the result, for example, of a worsening or an improvement compared to the past. Basing our interpretation only on a punctual perspective can make it an oversimplification. Trend analysis not only allows a clearer view of the reality observed, but also allows, with the appropriate precautions, to define scenarios for future trends. Thus, the central point of the issue is that indicator systems are longitudinal by nature. As well as from a conceptual point of view, the temporal perspective obviously also has effects on the practical and methodological side. It is necessary to provide tools for analysis and synthesis of indicators that take account of time. The statistical literature on the analysis of time series is very rich and has developed in various fields. However, in the specific field of indicators and their synthesis, the temporal dimension has been little addressed and developed.

For this reason, my thesis focuses on temporal analysis. Not only to fill what in a certain sense can be considered a *void* in the literature, but especially because it makes no sense to speak of synthesis in a perspective that does not take into account the evolution of phenomena over time.

3.1.1. Multi-indicator systems over time: mathematical formalisation

In its simplest form, a system of indicators is a matrix of data, typical of multivariate statistics:

$$\mathbf{X} \equiv \left\{ x_{ij} : i = 1...N; j = 1...M \right\} = \left\{ \begin{matrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \ddots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{matrix} \right\} \tag{3.1}$$

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 indicator systems are in the form of a particular type of three-way data array[3]: the *three-way data time array*[4]. These data structures are characterised 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

3 A three-way data array is a specification of a multiway array [66, 153] in which there are three indices.

4 We need to specify that in this work we deal with three-way time data arrays, which are different from *three-way time data sets*. A set of multivariate data constitutes a three-way data time array if all the combinations of the values of the three indices are present: in practice, we have the same set of units on which the same set of indicators is observed in the different times. On the contrary, if not all the possible combinations of the three indices are present, we will have a three-way data time set. Obviously there can be different types of three-way data time set; for more information, please see: Pierpaolo D'Urso [90].

the multivariate information is collected [90]. The three-way data time arrays can be formally represented as follows:

$$\mathbf{X} \equiv \{x_{ijt} : i = 1, \ldots, N;\ j = 1, \ldots, M;\ t = 1, \ldots, T\} \quad (3.2)$$

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. In other words, the three-way data time array $\mathbf{X}$ can be seen as a collection of T matrices of order $(N \times M)$, each of which represents a slice of $\mathbf{X}$.

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. The latter is a fundamental aspect and cannot be excluded in the synthesis. A synthesis is meaningful, in fact, only if it allows an evaluation of phenomena not only in space, but also in time, grasping their evolution; the aim is to build synthetic measures that allow the analysis of complex phenomena over time.

3.2. Exploratory analysis

Before making any synthesis on the multi-indicator system, an *exploratory analysis* of the data should be carried out. This analysis focuses on basic indicators (columns) and aims to discover possible statistical relations among them in order to measure their reciprocal influence. Exploratory analysis practically consists in applying techniques of multivariate statistics to data aiming at studying the relationships between the basic indicators and verifying their contribution to the explanation of the variability. That we can define the *traditional approach*[5] to exploratory analysis consists in the study of correlations and the application of principal component analysis (PCA), in order to evaluate the complexity of the phenomenon [136, 3, 14, 4, 184]. As known, the two techniques considered assume independence of error among observations and, consequently, they are not appropriate to analyse repeated measures, in which each unit provides more than one data point. Thus, from the operational point of view, the exploratory analysis is carried out on the

[5] We describe the most common way of performing exploratory analysis. Although, many other statistical tools can be useful. For instance, *factor analysis* or *multidimensional scaling* can be used to test the hypothesised dimensional structure underlying the selected indicators.

basic indicators taking into account the last available year or each year separately. This is not a particularly good strategy. In both cases, the time perspective is not taken into account. For instance, analysing only the last year available we can observe correlations among the indicators that are different and not recorded in the other years.

Leonardo S. Alaimo and F. Maggino [13] propose to perform exploratory analysis by using statistical techniques applicable to repeated measures over time (i.e. data of a three-way data time array). When we want to study the correlation between two indicators in which measurements of the same unit are repeated over time, we need to consider two aspects [35, 36]. The first one is whether observations with high values of one indicator also tend to have high values of the other one. For instance, if we want to know if high values in the life expectancy correspond to high values in the employment rate, we are interested in whether the first variable's average for an observation correlates to the second variable's average. In this case, we can average the repeated-measures data for each observation and perform a standard *Pearson correlation on average data*[6]. Given the array 3.2, we obtain the average data matrix:

$$\overline{\mathbf{X}} \equiv \left\{ \overline{x}_{ij} : i = 1...N; j = 1...M \right\} \tag{3.3}$$

where $\overline{x}_{ij}$ is the arithmetic mean of the T temporal determinations of the j-th indicator in the i-th unit. On the obtained matrix we compute the correlations for all the pairs of indicators considered. The Pearson's correlation coefficient is given by:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} \tag{3.4}$$

where $\text{Cov}(X, Y)$ is the co-variance of the two indicators X and Y, $\text{Cov}_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$; σ_x and σ_y are the standard deviation respectively of X and Y.

6 The Pearson correlation coefficient is probably the most widely used measure for linear relationships between two normal distributed variables. Related to it, the *Spearman correlation coefficient* can be understood as a rank-based version of Pearson's correlation coefficient, which can be used for variables that are not normal-distributed and have a non-linear relationship. Also, its use is not only restricted to continuous data, but can also be used in analyses of ordinal attributes. It is given by: $\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$.

The second aspect we must consider in analysing repeated measures correlation is if an increase in one indicator within the observation is associated with an increase in the other. For instance, we want to know if an increase in the life expectancy within the individual is associated with an increase in employment rate. To do this, we do not have to consider the differences between observations, but we have to look only at changes within them. In this case, we must use the *repeated-measures correlation coefficient* (r_{rm}). The repeated-measures correlation (rmcorr) accounts for non-independence among observations using the analysis of covariance (ANCOVA) to statistically adjust for inter-individual variability [23]. Rmcorr is an atypical application of (ANCOVA), with an opposite purpose. The analysis of co-variance is a statistical method used to test the effects of categorical variables on a continuous dependent variable, controlling for the effects of selected other continuous nuisance variables, which co-vary with the dependent. On the contrary, we can use "rmcorr to determine the relationship between two continuous variables, while controlling for the effect of the categorical variable, which in this case is the between-participants variance" [23, 3]. We can analyse variations *within the observation* using multiple regression. Let us consider two indicators, X e Y, one of which is the outcome variable (e.g. X), while the other one (e.g. Y) and the observations are the predictors (it does not matter which variable we regress on which). To isolate the correlation among the different measures of a specific observation, we treat each observation as a categorical factor using dummy variables (with $N-1$ degrees of freedom). Using ANCOVA, we can show how the variability in the outcome variable can be partitioned into components due to different sources represented, in this case, by the other variable, the observations and the residual. This method is equivalent to fitting parallel lines through the data of each observation and the residual sum of squares represents the variation about these lines.

We can estimate rmcorr using the equation describing an experimental design GLM for the single-factor independent measures ANCOVA with one co-variate [236, 216]:

$$Y_{ij} = \mu_Y + \tau_j + \beta(X_{ij} - \overline{X_j}) + \epsilon_{ij} \tag{3.5}$$

where Y_{ij} is the dependent measure for the i-th participant at the j-th factor level; μ_Y is the overall mean of the dependent variable; τ_j is is the effect of the j-th factor level; β is the overall slope coefficient of the co-variate; $(X_{ij} - \overline{X}_j)$ is the difference between the value of the covariate for the i-th participant at the j-th factor level (X_{ij}) and the mean of the covariate values for the j-th participant $(\overline{X}_j)$; ϵ_{ij} is the error term. We calculate rmcorr using formula 3.5 with an unusual model specification. "ANCOVA is typically used to assess the effects of different (treatment or factor) levels upon a dependent measure, while controlling for the effects of another continuous variable (the covariate). For rmcorr, the participant is the factor level and the covariate is the second measure" [23, 5].

Before calculating r_{rm}, we need to transform the three-way array into a bi-dimensional stacked matrix $\mathbf{X}_t$, by combining the indices i and t on the rows and assigning the index j on the columns. So, each unit will have data in multiple rows. From the operational point of view, we must *reshape* the three-way time data array in the formula 3.2 from *wide* format (in which, a unit's repeated measures are in a single row and each measure is in a separate column) to *long* one (in which each row is one time point per unit) as follows:

$$\mathbf{X}_t \equiv \left\{ \begin{array}{ccccc} x_{1_1 1} & \cdots & x_{1_1 j} & \cdots & x_{1_1 M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1_t 1} & \cdots & x_{1_t j} & \cdots & x_{1_t M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1_T 1} & \cdots & x_{1_T j} & \cdots & x_{1_T M} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x_{i_1 1} & \cdots & x_{i_1 j} & \cdots & x_{i_1 M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i_t 1} & \cdots & x_{i_t j} & \cdots & x_{i_t M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N_T 1} & \cdots & x_{i_T j} & \cdots & x_{i_T M} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ x_{N_1 1} & \cdots & x_{N_1 j} & \cdots & x_{N_1 M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N_t 1} & \cdots & x_{N_t j} & \cdots & x_{N_t M} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N_T 1} & \cdots & x_{N_T j} & \cdots & x_{N_T M} \end{array} \right\} \tag{3.6}$$

The equation 3.5 is rewritten for rmcorr to show one measure as a function of its mean value, participant, and the covaried value of the other measure[7]:

$$X_{ij} = \overline{X}_j + \mathrm{Obs}_j + \beta(Y_{ij} - \overline{Y}_j) + \epsilon_{ij} \tag{3.7}$$

where:

- X_{ij} is the value of variable X for the i-th measure at the j-th participant/unit;
- $\overline{X}_j$ is the mean of the variable X, in all i-th measures, for the j-th participant/unit;
- Obs_j is a unique identifier, used as a dummy coded variable;
- β is the slope coefficient of the co-variate;
- $(Y_{ij} - \overline{Y}_j)$ is the difference between the value of Y for the i-th measure at the j-th participant/unit (Y_{ij}) and the mean of Y values in all i-th measures for the j-th participant/unit $(\overline{Y}_j)$;
- ϵ_{ij} is the error term.

By removing the variation due to observations (and any other nuisance variables, which might be present), we calculate the r_{rm} as follows:

$$r_{rm} = \sqrt{\frac{\mathrm{SS}_Y}{\mathrm{SS}_Y + \mathrm{SS}_{Error}}} \tag{3.8}$$

where SS_Y is the sum of squares of the indicator Y; SS_{Error} is the residual sum of squares. The r_{rm} is bounded $[-1, 1]$ (the sign depends on the sign of β in the formula 3.7); it expresses the strength of the linear association between two variables and evaluates the overall or common intra-individual association between them. Rmcorr generally has much greater statistical power than a standard Pearson correlation using averaged data, and its power increases exponentially when either the value

[7] In the equation 3.7, i and j are exchanged for consistency: j = participant/unit and i = repeated measure.

of the number of repeated observations, or that of the total number of unique observations, increases.

Thus, when we deal with repeated measures, the correct analysis of the relationship between two variables must take into account both the correlation within (CW) and between (CB) observations.

Historically, PCA[8] has been carried out for repeated measures by reducing the problem to two dimensions. We may analyse three-way data either after aggregating over one of the three ways (in the case of the repeated measures, the time variable) or by analysing all two-way matrix contained in the three-way data array separately [148]. These approaches may lead to misleading conclusions, because they do not offer an explicit description of the three-way interaction in the data. For this reason, some specific techniques have been developed [152, 148, 153]. "The strength of three-way methods is that they summarise the entities of each mode through a few components and describe the relations between these components" [148]. For the objectives of exploratory analysis in the construction of composites, these techniques would probably be too sophisticated. In the exploratory analysis, PCA only has a descriptive purpose. In particular, if the variance explained by the first component is high[9], most of the indicators correlate, and they represent a single aspect of the phenomenon. This leads to the conclusion that we can consider only one latent factor and, then, we can construct a single composite. Otherwise, if the variance explained by the first component is not very high, there are several groups of indicators representing different aspects of the phenomenon and, consequently, this seems to highlight the presence of more than one latent factor and the necessity of constructing more than one composite. To perform PCA for exploratory analysis over time, we adopt a procedure similar of that adopted for correlation *between* observations analysis (CB), i.e. we perform a standard two-way PCA on matrix in formula 3.3.

We need to make some clarifications. The importance of the study of existing correlations among the indicators of a system is evident in the

---

8 Principal component analysis is a technique for reducing the dimensionality of datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximise variance. Finding such new variables, the *principal components*, reduces to solving an eigenvalue/eigenvector problem. It was invented in 1901 by Karl Pearson [217] and later independently developed and named by Harold Hotelling [134].

9 There is no precise threshold; in general, we can say that if the first component explains more than 50% of the total variance, we can consider only one latent construct present.

case of a *reflective model*. In fact, the indicators in this case are function of the latent variable to be measured; the correlation among indicators is explained by the measurement model and two un-correlated indicators cannot measure the same construct. Therefore, correlation analysis allows the exclusion of those indicators that are not effect of the latent variable. The analysis of correlations among the basic indicators is equally important for the *formative models*. In this case, the internal consistency of the indicators is of minimal importance, since two un-correlated indicators may both be relevant for the same construct. The correlations are not, therefore, of particular relevance; however, it is important to study them for an overview of the data structure. It is important to clarify that, although important, the correlations are not decisive[10]: in the construction of synthetic indicators, they are a *guide*. The first thing to evaluate is the measurement model, considering that it depends not on an arbitrary choice of the researcher, but on the definition of the phenomenon and the consequent nature of the latent variable. Therefore, correlations' analysis must not guide the selection of indicators, which must always be done according to the theoretical framework. Moreover, the analysis of the correlations must be carried out considering not only the strength, but also and above all the direction of the relations, in order to identify those unexpected. "Indicators should be selected on the basis of their analytical soundness, measurability, country coverage, relevance to the phenomenon being measured and relationship to each other" [213, 15].

Exploratory analysis is a preliminary phase to synthesis, which has a sort of *diagnostic purpose*. By applying a series of statistical techniques, information is collected that allows the evaluation of the system of indicators. In this sense, it is absolutely fundamental to use specific graphical representations useful to highlight the structure of the data and the relationships among the indicators.

3.3. Synthesising units in three-way data time arrays: an overview

As anticipated in the introduction to this Chapter, one of the aspects of the synthesis of multi-indicator systems concerns the aggregation of statistical units into macro-units. Dealing with this topic means dealing

10 That is also because, in most cases, only the presence of linear correlation is analysed, where the relationship among indicators could be non-linear.

with *cluster analysis*. Generally speaking, clustering is an unsupervised learning task that aims at decomposing a given set of objects into subgroups, called *clusters*, based on similarity. Objects belonging to the same cluster are as similar as possible, whereas objects belonging to different clusters are as dissimilar as possible [154]. In a very simple manner, given a dataset $X = \{x_1, x_2, \dots x_n\}$ with n data points, the aim of clustering is to classify them into K groups (clusters), which are disjoint subsets of X, not knowing in advance whether such clusters actually exist. A clustering is, in first approximation, the partition obtained, that is, $\mathbf{C} = \{c_1, c_2, \dots c_k\}$. This multivariate technique has been developed in several different fields and, consequently, there is a large range of approaches to cluster analysis in the scientific literature[11]. As specified in paragraph 3.1.1, this research work focuses on the statistical tools for synthesising three-way time data arrays (see formula 3.2). So, we need to concentrate on techniques suitable for those type of arrays. The general problem of clustering those objects is concerned with the separation of a set of time series data into clusters with the property that series in the same group have a similar structure and series in other groups are quite distinct.

We can geometrically represent a three-way time data array by indicating "the elements of one of the three classification modes as vectors of a vectorial space defined with regard to the other ones" [93, 15]. Let's consider the geometrical representation of the time data array $\mathbf{X}$ (formula 3.2) in the vectorial space of the units $\mathbb{R}^{M+1}$, where the first M dimensions correspond to the M variables and the last dimension is referred to the time [90]. In this vectorial space, each unit i-th for each time t-th is represented by the vector [91]:

$$\mathbf{y}_{it} = (x_{i1t}, \dots, x_{ijt}, \dots, x_{iMt}), i = 1 \dots N, t = 1, \dots, T \tag{3.9}$$

For fixed t, the generic matrix $\mathbf{X}_t$ is represented by the scatter:

$$S_N(t) \equiv \{\mathbf{y_{it}} : i = 1, \dots, N\} \tag{3.10}$$

11 For a systematic view of cluster analysis, for instance see: Christian Hennig and Marina Meila [130].

The set of scatters $\{S_N(t) \equiv \{\mathbf{y_{it}} : i = 1, \ldots, N\} : t = 1, \ldots, T\}$ is represented by T hyperplanes. For fixed i, the generic matrix $\mathbf{X}_i$ is represented by the scatter:

$$S_T(i) \equiv \{\mathbf{y_{it}} : t = 1, \ldots, T\} \tag{3.11}$$

that describes the multivariate time trajectory of the i^{th} unit during time. The set of scatters $\{S_T(i) \equiv \{\mathbf{y_{it}} : t = 1, \ldots, T\} : i = 1, \ldots, N\}$ represents the set of the time trajectories of the N units. Each time trajectory crosses the T hyperplanes.

For the classification task, the matrix $\mathbf{X}_i \equiv \{\mathbf{x}_{it} : \ t = 1, \ldots, T\}$ represents the i-th multivariate time trajectory [91, 93], where:

$$\mathbf{x}_{it} \equiv (x_{i1t}, \ldots, x_{ijt}, \ldots, x_{iMt}) \text{ with } i = 1, \ldots, N, t = 1, \ldots, T$$

This paragraph highlights the main features of techniques of clustering multivariate time trajectories, trying to provide a guide in choosing those that best fit the nature of the multi-indicator systems. This is not an easy task, also because in the field of synthesis researchers often focus on indicators, while the topic of the *synthesis of units* is very often neglected.

The literature on time series clustering has increased over the last two decades, with a large range of empirical applications in many different fields [90, 273, 247, 117, 141, 111, 219, 231, 93, 177]. Time series data are of interest because of its pervasiveness in various scientific fields. Clustering such complex objects is particularly advantageous for several reasons. Time series clustering deals with classifying the data points over time based on their behaviour. The analysis of the clusters' structure can help in easily detecting the main information in a dataset, such as regularities and anomalies. In particular, "the discovery of anomalies could be of primary importance to avoid the disruptive effect of the presence of outliers" [93, 13]. Time series databases are often very large and difficult to be handled. Hence, it is preferable to deal with structured datasets. Therefore, time series data are represented as a set of groups of similar time series by aggregation of data [1]. However, even when dealing with not very large datasets (typical of the case of multivariate time series of social phenomena) it may be useful to identify groups and synthesize information in a meaningful way.

Three main approaches can be adopted for the classification of time-series [54, 177]:

- *Feature-based clustering*

  This is a particularly useful approach for long and noisy time series, for which applying clustering time series based on the Euclidean distance in the space of points is not a good option. That is "because of the noisy present and the fact that the autocorrelation of the time-series is ignored" [177, 68]. Techniques belonging to this approach are based on features extracted in the time domain, frequency domain or wavelet decomposition of the time series [97, 179, 178, 94].

- *Model-based clustering*

  For methods belonging to this approach, time series are assumed as generated from "specific underlying models or by a combination of probability distributions, and the similarity between fitted models is evaluated" [177, 112]. Time series are clustered by means of parameters estimates or by means of the residuals of the fitted models [220, 253, 176, 96].

- *Observation-based clustering*

  This approach is based on the comparison of the observed time series (or a suitable transformation). It is useful if we want to cluster time series according to their geometric profiles. In the literature, several distance measures and clustering methods have been suggested [150, 151, 241, 91, 67, 68, 69, 70]. This approach is particularly suitable for short time series. For this reason, it is particularly appropriate for multi-indicator systems, in most cases characterised by few time occasions.

Dealing with a system of indicators means dealing with complexity. Thus, in classifying those systems we must take into account their complexity. The latter is not only conceptual, but also operational, in particular, related to the nature of multivariate time series. They often present a switching behaviour: they could have a dynamic pattern consistent with a given cluster for a certain time period and then a completely different one more similar to another cluster. This characteristic cannot be underlined with a traditional *crisp* approach, in which "each datum is exactly assigned to only one cluster obtaining exhaustive partitions characterised by nonempty and pairwise disjoint subsets" [92, 547]. Thus, the assignment of data to clusters is forced and, consequently, this can be inadequate in presence of data points that are almost equally distant from two or more clusters. The alternative is to adopt a *fuzzy* approach. Fuzzy clustering is an overlapping approach[12], based on the Fuzzy Set Theory [274], which allows units to belong to more clusters simultaneously depending on a certain membership degree [32]. There is no longer that a unit belongs or not to a generic cluster; it belongs to a cluster according to a certain membership degree bounded between 0 (complete non membership) and 1 (complete membership). Fuzzy approach is a natural way to address the uncertainty of systems of indicators. It is based on the evidence that the real world is so complex that it cannot be treated through clear and rigid propositions. The reality is never white or black, we must always take into account the *shades of gray*. Using stringent criteria (crisp approaches) to identify uniformity of behaviour among the statistical units in which a social construct is being measured through a system of indicators can generate misleading conclusions, since forcing attribution to one group rather than another can lead to loss of information.

Each cluster must be represented by a *prototype*, i.e. an object which presents its main characteristics. The identification of prototypes can be particularly complex in the case of time series [13]. According to Pierpaolo D'Urso et al. [93], a natural way to address this issue is to follow a *Partition Around Medoid* (PAM) approach; the cluster's prototype is an observed representative multivariate time series, the *medoid*. PAM approach is also convenient from a computational point of view[14]. In the context of multi-indicator systems, the choice of a PAM approach is particularly indicated, also because it facilitates the interpretation of the

12 Approaches of cluster analysis differ in how the different clusters relate to each other. Many cluster analysis approaches aim at finding partitions *mutually exclusive*, in which $c_i \cap c_j = 0$ for $i \neq j$. Overlapping clustering techniques are those that violate the condition of mutually exclusivity.

13 Dealing with a two-dimensional matrix (formula 3.1), prototype could be the (weighted) mean of the features of the objects belonging the cluster, a *centroid*. Dealing with three-way time data arrays (formula 3.2), this is not a good choice: a centroid multivariate time series is more difficult to comprehend and to achieve, even if in the literature there are some proposals [219].

14 "By adopting this approach it is possible to compute the distance matrix only once, since data do not change during the iterative clustering procedure" [93, 13].

results obtained.

In the clustering field, a main topic is the presence and the processing of *outliers*. In time series clustering, an outlier could be defined in several ways[15]. Different robust fuzzy clustering methods, classified in different approaches[16], have been proposed for neutralising the negative effects of outliers [93]. What should be emphasised here is that the presence of outliers must always be taken into account when dealing with multi-indicator systems and, where they are present, adopt robust synthesis procedures, both for units and indicators.

The synthesis of the units of a multi-indicator system, even if often little considered and applied by the literature in this field, constitutes an important tool both for operational and interpretative purposes. It is often a complementary tool to the synthesis of indicators. Synthesising indicators, we obtain a *reduced* system compared to the starting one. Even the analysis of this *reduced* system could be complex, especially if we want to have a general scenario that takes into account different domains and allows a comparison of units over time. Thus, classifying statistical units in groups can facilitate the understanding of phenomenon. This is the subject of the analysis in the fourth Chapter: starting from the time series of composite indicators created by Italian National Institute of Statistics (Istat) for the *Equitable and Sustainable Well-being* (BES) project, we classify the Italian regions over time considering the 12 BES domains. To do this, we apply a distance measure and a classification algorithm (described in the paragraph 4.3), the choice of which was made considering the characteristics described in this paragraph. Obviously, this is one of the possible methods that can be applied, all of which are suitable to the nature of phenomenon.

---

15 For instance, we can consider a multivariate time series as an outlier if its dynamics deviate markedly from the rest of data or even if one or more of its components have an anomalous behaviour.

16 Pierpaolo D'Urso et al. [93] enumerate the following: *noise* approach (outlier time series are assigned to the so-called noise cluster); *metric* approach (distance measures with robust properties are incorporated in the objective function of the clustering method); *trimmed* approach (the clustering method is applied to the time series remaining after a fixed fraction of outliers are eliminated) and *influence weighting* approach (a weighting system is incorporated in the clustering method for assigning objectively low weights to outlier time series).

3.4. Synthesising indicators over time using different approaches

The synthesis of indicators (columns) of a system is certainly the most widely discussed aspect in the scientific literature in this field. In its simplest form, the problem can be formalised as follows. Given the data matrix $\mathbf{X} \equiv \{x_{ij}\}$ (see formula 3.1) with N statistical units and M indicators, the objective is to synthesize it in a vector $\mathbf{v} \equiv \{v_i\}$, with N statistical units, in which the generic element v_i represents the synthetic value of the i-th unit with respect to all M indicators of the original matrix $\mathbf{X}$:

$$\mathbf{X} \equiv \left\{ \begin{matrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \ddots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{matrix} \right\} \Rightarrow \mathbf{v} \equiv \left\{ v_i \right\} = \begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_N \end{pmatrix} \tag{3.12}$$

Dealing with temporal data, as represented in the three-way time data array in the formula 3.2, the objective is to obtain for each generic unit i-th one synthetic measure for each temporal occasion t-th. The generic synthetic measure of the unit i-th at the generic time occasion t-th is obtained considering the values the unit assumes in the M basic indicators in the generic t-th temporal occasion. Considering all units and all time occasions, we obtain a two-way data matrix $\mathbf{V}$ as follows:

$$\mathbf{X} \equiv \{x_{ijt} : i = 1, \ldots, N;\ j = 1, \ldots, M;\ t = 1, \ldots, T\} \Rightarrow \mathbf{V} \equiv \left\{ v_{it} \right\}$$

$$\mathbf{V} \equiv \left\{ v_{it} : i = 1, \ldots, N; t = 1, \ldots, T \right\} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1T} \\ v_{21} & v_{22} & \cdots & v_{2T} \\ \vdots & \ddots & \ddots & \vdots \\ v_{N1} & v_{N2} & \cdots & v_{NT} \end{pmatrix} \tag{3.13}$$

where v_{it} is the synthetic value of the unit i-th at the time t-th. It is, therefore, necessary to focus on how to obtain the synthesis of indicators from a technical point of view. In other words, we should focus on the *arrow* $\Rightarrow$ in the formulas 3.12 and 3.13. From the technical point of view, the synthesis can be faced through two different approaches, *aggregative-compensative* and *non-aggregative*. We must remember what was specified in the introduction to this Chapter, that is, we deal with the synthesis of systems of cardinal indicators considered from a temporal point of view. Therefore, the synthesis' methods examined are those that allow a dynamic synthesis and that are suitable for cardinal variables.

3.4.1. The aggregative-compensative approach

As suggested by the term, the aggregative approach consists in the aggregation, by means of a mathematical function, of the basic indicators. These methodologies are defined *composite* indicators [240, 213]. Building a composite indicators is not an easy task and requires a step-by-step process [209].

The first step is the definition of phenomenon we want to measure. We have analysed the main aspects of measuring social complex phenomena in the second Chapter. It is always necessary that the concept refers to and is inserted within a theoretical framework that gives it meaning. Particular attention should be given to the *measurement model* as we have seen in paragraph 2.3.2. 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. As known, the choice of the measurement model depends by appropriateness to the phenomenon to be measured and on the nature and direction of relationships between constructs and measures. 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[17].

From the operational point of view, after the definition of the phe-

[17] The reflective measurement model is most widely used in psychological and management sciences. "The main approach allowing to deal with reflective models is undoubtedly *Factor Analysis*, which can be applied in order to test the hypothesised dimensional structure underlying the selected indicators. In particular, it allows indicators that fit better the latent dimensional structure to be synthesised. The approach is based upon the assumption that the total variance of each indicator represents a linear combination of three different components (additive assumption): *common variance* (due to the dimensional structure), *specific variance* (due to the specificity variance of each indicator), and *error*. Actually, this analysis allows, by estimating for each indicator the amount of common variance (communality), the reflective approach to be tested. Indicators turning out to be part of the same supposed underlying dimension can be meaningfully synthesised" [171, 122].

nomenon and the selection of basic indicators, the following phases are the *normalisation* of the basic indicators, and the *aggregation* of the normalised indicators [209, 213, 187]. Normalization is required to make the indicators comparable, because they often present different measurement units and ranges. The objective is to transform them into *pure numbers*. In the normalisation, it is necessary to define the *polarity* of the basic indicators, i.e. the sign of the relation between the indicator itself and the phenomenon. 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). For instance, if we want to construct a composite whose increase coincides with an improvement in health, the life expectancy would have positive polarity, while the smoking rate would be negative. If, on the contrary, we want to construct a composite whose increase indicates a worsening of health (for instance, a risk indicator), the life expectancy would have negative polarity, while the smoking rate would be positive. After the normalisation, all the indicators must have positive polarity, i.e. "an increase in the normalised indicators corresponds to an increase in the composite index" [187, 166]. We can identify two main methods useful to invert polarity:

- the *linear transformation*, in which we take the complement with respect to maximum value. Given the three-way time data array $\mathbf{X} \equiv \{x_{ijt}\}$ in the formula 3.2, the linear inversion of polarity is calculated as follows:

$$x'_{ijt} = Max_{x_j} - x_{ijt} \tag{3.14}$$

  where Max_{x_j} is the absolute maximum value of the indicator j-th, x_{ijt} is the value of the indicator j-th in the unit i-th at the time t-th and x'_{ijt} 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, standardisation and re-scaling.

- The *non-linear* transformation, in which we take the reciprocal of the value to be inverted. Given the three-way time data array

$\mathbf{X} \equiv \{x_{ijt}\}$ in the formula 3.2, the inversion is calculated as follows:

$$x'_{ijt} = \frac{1}{x_{ijt}} \tag{3.15}$$

This technique, typically used with indicization, is criticised 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. This ratios are particularly used for measuring *gender gap* [267]: 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:

$$x'_{ijt} = |\lambda_{x_j} - x_{ijt}| \tag{3.16}$$

where λ_{x_j} is the value of the indicator j-th in which the polarity inverts (the threshold).

There are various normalisation methods[18], each of them has strengths and weaknesses. The researcher must identify the most suitable normalisation method to apply to its research works considering its properties. Dealing with multi-indicator system over time, the most commonly used normalisation method is the $Min - Max$. Given the three-way time data array $\mathbf{X} \equiv \{x_{ijt}\}$ (formula 3.2), the normalisation is carried out as follows:

$$r_{ijt} = \frac{(x_{ijt} - \underset{it}{Min}_{x_{ijt}})}{(\underset{it}{Max}_{x_{ijt}} - \underset{it}{Min}_{x_{ijt}})} \tag{3.17}$$

[18] For a review of the most commonly used normalisation methods, please see: Matteo Mazziotta and Adriano Pareto [187].

where x_{ijt} is the value of the indicator j-th in the unit ith at the time t-th; $\underset{it}{Min}_{x_{ijt}}$ is the minimum value of the indicator j-th for all units i in all temporal occasions t; $\underset{it}{Max}_{x_{ijt}}$ is the maximum value of the indicator j-th for all units i in all temporal occasions t and r_{ijt} is the normalised value of the indicator j-th in the unit i-th at the time t-th. We use the formula 3.17 if the indicator j is positive, otherwise we apply the following:

$$r_{ijt} = \frac{(\underset{it}{Max}_{x_{ijt}} - x_{ijt})}{(\underset{it}{Max}_{x_{ijt}} - \underset{it}{Min}_{x_{ijt}})} \tag{3.18}$$

Another method is to calculate the *index numbers* by choosing a fixed base (the value assumed by an indicator in a specific year, usually the first or last in the series). In this case it is necessary that all the indicators have positive polarity, so, if necessary, some must be inverted before normalisation, using one of the procedures previously described. The normalisation is achieved as follows:

$$r_{ijt} = \frac{x_{ijt}}{\underset{it}{Ref}_{x_{ijt}}} \tag{3.19}$$

where x_{ijt} is the value of the indicator j-th in the unit i-th at the time t-th; $\underset{it}{Ref}_{x_{ijt}}$ is the fixed base for the indicator j and r_{ijt} is the normalised value.

The following step is the *aggregation* of the normalised indicators. In the literature, many methods have been proposed for constructing composites [240], in particular for cross-sectional multi-indicator system [213, 187], which can also be used for systems over time. 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 the literature, there are different approaches to the weighting issue. Within the composite indicator framework, methods for weighting indicators can be broadly categorised into three main groups [114, 492].

- the *equal weighting*, i.e. giving to all the indicators the same weight;
- the *statistic-based weighting*, in which weights derive from the statistical characteristics of the data and are attributed as the results of a statistical method (for instance, principal component analysis);
- the *public/expert opinion-based weighting*, which relies on inputs from the participating public or experts, whose judgements ultimately determine the weights to be assigned to individual indicators [213].

No agreed methodology exists to weight basic indicators. The simplest weighting strategy, i.e. attributing equal weight to all basic indicators, considering them equally important [209], is the most commonly used. This method is not without criticism; especially from those who consider a possible misconception to the underlying logic, according to which the "weight assigned to a variable can be directly interpreted as a measure of its importance to the value of the composite" [26, 12]. The statistical method is very questionable, because most of the time it is based on the correlations among basic indicators and, as we have seen, their interpretation changes according to the measurement model. Probably, the best method could be the one based on the stakeholders/experts' opinion. 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.

Different classifications for aggregation methods exist. As shown by Xiaoyu Gan et al. [114], they include those based on the semantics of aggregation [27, 122] and those based on the degree of compensation tolerated [213]. Based on the latter classification, widely used aggregation methods include:

- *Additive aggregation methods*

  These methods employ functions that sum up the normalised values of basic indicators to form a composite index. The most widespread additive method is the *weighted arithmetic mean*. Given the three-way data array $\mathbf{R} \equiv \{r_{ijt}\}$ of the normalised data[19], the synthesis is given by:

[19] In these cases we consider the hypothesis in which the data should be normalised. If the normalisation is not necessary, we can use the array $\mathbf{X} \equiv \{x_{ijt}\}$ of the original data (formula 3.2).

$$v_{it} = \sum_{j=1}^{M} r_{ijt} w_j \qquad (3.20)$$

where w_j is the weight of the indicator j-th. In the case of equal weighting, i.e. if $w_i = \frac{1}{M}$, we have the simple arithmetic mean. This technique implies full compensability, such that poor performance in some indicators can be compensated for by sufficiently high values in other indicators.

- *Multiplicative aggregation methods*

  Geometric aggregation methods use multiplicative instead of additive functions. The most widespread geometric aggregation function is the *weighted geometric mean*. Given the three-way data array $\mathbf{R} \equiv \{r_{ijt}\}$ of the normalised data, the synthetic measure is given by:

$$v_{it} = (\prod_{j=1}^{M} r_{ijt}^{w_j})^{1/M} \qquad (3.21)$$

  where w_j is the weight of the indicator j. In the case of equal weighting, we have the simple geometric mean. Geometric mean-based methods only allow compensability between indicators within certain limitations (partially compensative). This requirement exists because of the "geometric-arithmetic means inequality" [27, 114], which limits the ability of indicators with very low scores to be fully compensated for by indicators with high scores. Simultaneously, significant marginal effects maybe measured using geometric methods when increasing the values of indicators with relatively low absolute values [213].

- *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. The *substitutability* is a fundamental issue in composite construction. The components of a composite index are called non-substitutable if a compensation among them is not allowed (i.e. a deficit in one component may not be compensated by a surplus in another). A non-compensatory approach generally requires the use of non-linear functions, such as Multi-Criteria Analysis [208].

Within this approach, a particularly used method is the *Adjusted Mazziotta-Pareto Index* ($AMPI$). This partially non-compensatory composite indicator has 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. It is a variant of the Mazziotta-Pareto Index (MPI), based on a $Min - Max$ normalisation and a re-scaling 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 [186]. This normalisation procedure allows assessing absolute changes over time. "Using AMPI, we compute the score of each unit independently of the others, in contrast to the MPI where the mean and standard deviation of the individual indicators are required" [187, 179]. Moreover, various analyses have shown that this method is more robust than others are [183, 185]. Given the three-way data array $\mathbf{X} \equiv \{x_{ijt}\}$ (formula 3.2), first data is normalised by using a variant of the $Min - Max$ method as follows:

$$r_{ijt} = \frac{(x_{ijt} - MIN_{x_j})}{(MAX_{x_j} - MIN_{x_j})} * 60 + 70 \tag{3.22}$$

where x_{ijt} is the value of the indicator j-th in the unit i-th at the time t-th; MIN_{x_j} and MAX_{x_j} are the two goalposts of the indicator j-th and r_{ijt} is the normalised value. If the basic indicator has positive polarity, the formula 3.22 is used; otherwise, the formula 3.23 is calculated:

$$r_{ijt} = \frac{(MAX_{x_j} - x_{ijt})}{(MAX_{x_j} - MIN_{x_j})} * 60 + 70 \tag{3.23}$$

The two goalposts are defined as follows:

$$Ref_{x_j} \pm \Delta \quad \text{with} \quad \Delta = \frac{(\underset{it}{Max}_{x_{ijt}} - \underset{it}{Min}_{x_{ijt}})}{2} \tag{3.24}$$

where Ref_{x_j} is the reference value, i.e. the value of the indicator j-th in a specific unit i-th at a specific time t-th and $\underset{it}{Max}_{x_{ijt}}$ and $\underset{it}{Min}_{x_{ijt}}$ are, respectively, the maximum and the minimum value of the indicator j-th 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:

$$AMPI^{\pm} = \underset{it}{\mu}_{r_{ijt}} \pm \underset{it}{\sigma}_{r_{ijt}} * \underset{it}{cv}_{r_{ijt}} \tag{3.25}$$

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 unit i-th at the temporal occasion t-th. The sign $\pm$ 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 characterised by the combination of a *medium effect* $(\underset{it}{\mu}_{r_{ijt}})$ and a *penalty effect* $(\underset{it}{\sigma}_{r_{ijt}} * \underset{it}{cv}_{r_{ijt}})$, which allows penalising 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. What might seem a limit of $AMPI$, on the contrary, is one of its qualities, as it allows highlighting the presence of a strong variability in the time series of the basic indicators. The value 100 represents the reference value; therefore, $AMPI$ indicates how each unit is placed with respect to the goalposts.

Despite its success, the aggregative approach has been deeply criticized as inappropriate and often inconsistent, from both conceptual and methodological point of view [110, 172]. Critical conceptual issues are the definition of the theoretical framework and the related identification of relevant variables. However, these are critical aspects of the entire measurement process, as analysed in the second Chapter, and do not only concern the construction of composite indices. From the methodological point of view, critical aspects in composite indicator approach are the normalisation of basic indicators, the weighting and the selection of the mathematical function by which combining normalised indicators. We have already discussed the weighting issue in the previous pages. It should be added that the problem of the weighting system becomes more complex in the case of longitudinal analysis. In fact, it is not certain that a system of weights designed for a specific temporal occasion is also valid for the others. Just as the concept changes over time, so the system of indicators and, consequently, the system of weights provided for the latter can change. The definition of a weighting system for multi-indicator systems over time is an issue that has not been much addressed in the literature and remains an open question. Also for this reason, the selection of robust indicators validated by official statistics and the use of the same weight for each of them seems to be the choice not only conceptually, but also methodologically more acceptable and efficient.

The normalisation, as said, is necessary to allow the comparison between different indicators for unit of measure and variability. However, this is a very delicate operation, like all those carried out on the data. Moreover, from a conceptual point of view, standardisation does not solve the problem of putting together different measures, of *mixing apples and oranges*.

One of the main problems with aggregative methods is related to the way in which they are calculated, i.e. as a combination of basic indicators. As shown by Leonardo S. Alaimo and Filomena Maggino [13], composite indicators *flatten* the dynamism of phenomena. In particular, the authors identify two different errors related to the application of this approach. The *simplifying fallacy* refers to the risk of fallacious conclusions produced through an extreme simplification. An example could be represented by the classical way to conduct correlations analysis, which many studies are based on in constructing synthetic indicators.

Apart from taking into consideration only the presence of linear correlation, they do not consider the variability in time. This clearly limits the effectiveness of this analysis. This is an over-simplification. The *composition-through-compensation fallacy* refers to situations in which the composite indicator can produce same values for different situations. Even in the case of a robust and partially non-compensative synthesis procedure such as $AMPI$, the fact remains that the composites flatten the differences. We observe and analyse this situation in the application reported in the fifth Chapter. This questions the effective full discriminatory capacity of composites. A correct understanding of the phenomena requires the use of procedures that respect the values of each unit's profiles. Aggregating is useful in order to simplify the complexity; however, it does not allow a precise analysis, crushing and flattening the differences. Making synthesis through compensative aggregation is not able to render a full understanding of the complexity of social phenomena. In other words, reducing a set of indicators to a single number flattens the differences between statistical units, *making comparable their incomparabilities*.

Marco Fattore [105] underlines other critical aspects. The management of multidimensional systems of ordinal data excludes the possibility of directly applying the composite indicator approach. Ordinal attributes cannot be synthesised by using an aggregative method, designed for numerical variables. In fact, ordinal scores cannot be treated as numbers. Despite this, we often see their transformation into numerical scores, by more or less sophisticated scaling tools, in order to make possible their synthesis by aggregative procedures. These procedures may lead to controversial and incorrect results. Apart from the mathematical and methodological implications, such operations pose delicate conceptual questions. "One could legitimately ask why concepts naturally conceived in ordinal terms should be forced into numerical settings" [105, 193–194].

In conclusion, despite its success, the aggregative approach is controversial due to conceptual and methodological difficulties. 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 *way* 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. In the next paragraph, we analyse one of those alternative procedures, that based on the application of the Partial Order Theory. As anticipated in the introduction to this Chapter, this method is particularly suitable for the treatment of ordinal data, but it can also be applied to systems of cardinal indicators, allowing to overcome some of the limitations of the aggregative approaches. Obviously, this is not a *perfect* or the *best method*; it also has its limits, which will be discussed in the next paragraph.

3.4.2. The non-aggregative approach

Posets are among the most common mathematical objects, Let's suppose we want to buy a car. It is likely that we do prefer some cars to others, but that in some cases we may not express any preference, leaving the alternatives unordered, i.e. *incomparable*. Similar situations are typical of social sciences, when a set of statistical units are scored against a set of ordinal attributes and resulting profiles have conflicting scores. Let's suppose we want to measure the level of life satisfaction of 5 individuals and to do so we consider whether they have experienced three positive situations, namely x, y and z. We suppose those situations are equally important. What is the level of life satisfaction of the subjects considered? Can we say which subject's life is better and which is worse? In Table 3.1, we report the profiles of the 5 individuals according to the scores in the three considered situations.

Tab. 3.1. Example: life satisfaction according to three positive situations.

Subject	X	Y	Z
A	Yes	Yes	Yes
B	Yes	Yes	No
C	Yes	No	Yes
D	No	Yes	Yes
E	No	No	No

Undoubtedly, individual A has the best profile, as does individual E the worst. Individuals B, C and D are incomparable, because they have totally different profiles. The traditional way of analysing such a situation consists in encoding the three variables, assigning, for example, to the No value 0 and to the Yes value 1. Well, in this case the next step would be to obtain a synthesis for each subject. This is usually done by applying an aggregative procedure. Apart from the conceptual error to treat attributes as numbers, the results are also absolutely misleading. In Table 3.2, we have coded the three variables and calculated two synthesis, one with the arithmetic mean and one with the geometric mean. The results highlight other problems. Subjects A and E obtain in both synthesis results that represent their situation. On the contrary, subjects B, C and D have the same value with both procedures, despite a totally different profiles in the three variables. Moreover, the geometric mean assigns to the three subjects considered the value 0, as well as to the subject E. We can conclude that in the aggregative-compensative approach the problem is not so much in the compensation, as in the nature of those methods, which, as mentioned above, make comparable the incomparable.

Tab. 3.2. Example: life satisfaction according to three positive situations: coding variables; arithmetic mean; geometric mean.

Subject	X	Y	Z	Arithmetic mean	Geometric mean
A	1	1	1	1.00	1.00
B	1	1	0	0.67	0.00
C	1	0	1	0.67	0.00
D	0	1	1	0.67	0.00
E	1	1	0	0.00	0.00

These considerations are valid for both ordinal and cardinal variables. The application of partial order theory allows the overcoming of those problems. It is a mathematical discipline which combines elements of Graph theory and Combinatorics. In the broadest sense is Graph theory that branch of Discrete Mathematics which studies relations. Within the context of the analysis of indicators the relations are defined on the basis of profiles [8].

3.4.2.1. Poset: definitions and formalisation

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 $\trianglelefteq$ 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 [210, 76, 225, 242]:

- 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 x_i and x_j belonging to the set X, if x_j is better than x_i and, at the same time, x_i is better than x_j, then the two elements are identical; i.e. if $x_i \trianglelefteq x_j$ and $x_j \trianglelefteq x_i$ then $x_i = x_j$, with $x_i, x_j \in X$;

- *transitivity* is present if the units are, at least, ordinal scaled and states the possibility of defining an order among them. i.e. if $x_i \trianglelefteq x_j$ and $x_j \trianglelefteq x_c$, then $x_i \trianglelefteq x_c$, with $x_i, x_j, x_c \in X$.

For instance, consider the set $X = \{a, b, c, d\}$ and the following list of comparabilities: $a \trianglelefteq a, b \trianglelefteq a, c \trianglelefteq a, d \trianglelefteq a, b \trianglelefteq b, c \trianglelefteq c, d \trianglelefteq b, d \trianglelefteq c, d \trianglelefteq d$. The set X and the relation $\trianglelefteq$ define the poset $P = (X, \trianglelefteq)$. The *incidence matrix* is a tool to define the structure of comparabilities. Let $k = |X|$ be the cardinality of X, the *incidence matrix* is a $k \times k$ boolean matrix $\mathbf{Z}$ summarising the comparability relation $\trianglelefteq$. The elements of $\mathbf{Z}$ may assume the values:

$$z_{ij} = \begin{cases} 1 & \text{if } x_i \trianglelefteq x_j \\ 0 & \text{otherwise} \end{cases}$$

for all $x_i, x_j \in X$. For instance the incidence matrix of the previous example is:

$$\mathrm{Z}_P = \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} . \tag{3.26}$$

It is necessary to define the *cover relation* to provide a graphical representation of comparabilities. Consider two elements $x_i, x_j \in X$, the element x_j covers the element x_i, $x_i \prec x_j$, if x_j dominates x_i, $x_i \trianglelefteq x_j$, and there is no other element $z \in X$ that jointly dominates x_i and is dominated by x_j, $x_i \trianglelefteq z \trianglelefteq x_j$. A directed acyclic graph can describe the cover relation $\prec$. The *Hasse diagram* is the graphical representation of this graph where the orientation from top to bottom substitutes the arrows. Two edges connected by a path are comparable by transitivity;

otherwise, they are incomparable. Figure 3.1 shows the Hasse diagram of the example previously introduced.

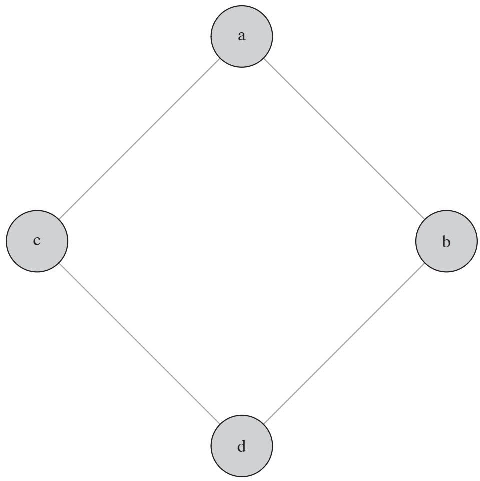


Fig. 3.1. Example of an Hasse diagram.

Finite posets always have *maximal* elements, i.e. those which are below no other elements (in Figure 3.1, the element a) and minimal ones, i.e. elements which are above no other elements (like element d in Figure 3.1). A subset of poset elements which are mutually comparable is called a *chain* (for instance, in the Figure 3.1 the elements d, c and a form a chain). On the contrary, a subset of poset elements which are mutually incomparable is called an *antichain* (for instance, in the Figure 3.1 the elements c and b are an antichain).

A poset gives us two kinds of information about its elements. First, there is an information related to the existence of comparabilities (we can define it *vertical* information) For instance, we can affirm that the element a is better than b. At the same time, there is also a *horizontal* information related to the presence of incomparabilities. This type of information shows the ambiguities within the set of poset elements, its *fuzziness*. If we take into account the two elements c and b, we cannot establish who is better and who is worse. There are ambiguities that reflect in the different *relational position* of elements, within the network of comparabilities and incomparabilities that defines the partial order [105]. An *extension* of $P = (X, \trianglelefteq)$ is a poset $P_e = (X, \trianglelefteq_e)$ on the same set X but equipped with a partial order relation $\trianglelefteq_e$ extending the relation $\trianglelefteq$. Therefore, all the pairs of elements comparable in $\trianglelefteq$ are comparable in $\trianglelefteq_e$ and some pairs comparable in $\trianglelefteq_e$ are not comparable in $\trianglelefteq$. A *linear extension* of P is an extension of P where all the elements of the set X are comparable. Therefore, it is a complete (or linear) order obtained extending the initial poset. In simple words, a poset which is a chain is also called a linear order. A poset usually has more than one linear extension. Let Ω_P be the set of all the linear extensions of P. We define *mutual ranking probabilities* (MRP) *matrix* of P the matrix $\mathrm{M}_P = (m_{ij}) \in \mathbb{R}^{k \times k}$, with $x_i, x_j \in X$ and m_{ij} is the relative frequency of linear extensions in Ω_P such that the element x_i is dominated by the element x_j. The MRP matrix of the example is:

$$\begin{array}{cc} & \begin{array}{cccc} a & b & c & d \end{array} \\ \mathrm{MRP}_P = & \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & 1.00 & 0.50 & 0.00 \\ 1.00 & 0.50 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 \end{bmatrix} \begin{array}{c} a \\ b \\ c \\ d \end{array} . \end{array} \tag{3.27}$$

The intersection of linear extensions (corresponding to the set of comparabilities they have in common) is equal to the original poset. In other words, a poset comprises all and only those comparabilities its linear extensions agree. Moreover, a set of linear extensions uniquely identifies only one specific poset. Thus, two finite posets have different sets of linear extensions and any finite poset is the intersection of its linear extensions [242]. In Figure 3.2 we report the linear extensions of the poset in the example.

This is an important property that can be used for the construction of synthetic measures starting from posets. We can use posets to define the structure of comparabilities underling multi-indicators systems. Once the structure is defined, we can analyse it through mathematical tools. Among the different alternatives, in this thesis we consider one score-vectors strictly order-preserving, the *average height*. By this measure, we can obtain a linear order by a partial order representing a system of indicators; in most cases, the objective of synthesis is exactly this one. "In fact, once evaluation scores are assigned to poset elements, these can be ordered, producing a linear order. In many cases, moreover, it is the final ranking, rather than the precise scores, what really matters (e.g.

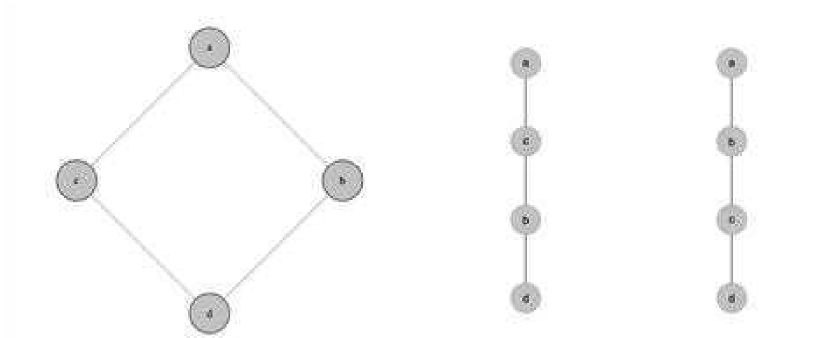


Fig. 3.2. Linear extensions of the Hasse diagram of the example.

think of a policy-maker that must allocate funds for a social program, to more deprived people)" [105, 201]. A very simple way to achieve this goal is by means of average height, i.e. by the association to each element in a finite poset P of a score representing its *position* on a *low-high* axis. The procedure can be summarised as follows:

1. List all the linear extensions Ω_P of P.

2. For each linear extension, compute the height of the elements of P. We define the height of an element x of the poset P as 1 plus the number of elements (the profiles) below x in the complete order.

3. For each element of P, compute the average height (avh) over the linear extensions of P.

Thus, the *average height* of an element $x \in X$ of P is the average over Ω_P of number of elements $y \in X : y \trianglelefteq_l x$ in the linear extension $P_l = (X, \trianglelefteq_l) \in \Omega_P$. It corresponds to the column sums of the MRP matrix and it is bounded between a minimum (the value 1) and a maximum (the total number of poset elements). In our example, the average height of each element is reported in Table 3.3

We must introduce an important property of average height. Let $P_A = (X, \trianglelefteq_A)$ and $P_B = (X, \trianglelefteq_B)$ be two posets sharing the same set of elements but with different structures of comparability $\trianglelefteq_A$ and $\trianglelefteq_B$. Let $\mathbf{h}_A$ be the average height vector of P_A and $\mathbf{h}_B$ the average height vector of P_B. The average height is an absolute scale representing the power of dominance of each element of X. Therefore, the two vectors $\mathbf{h}_A = [h_{A1}, \dots h_{Ai}, \dots h_{Ak}]^{\mathrm{T}}$ and $\mathbf{h}_B = [h_{B1}, \dots h_{Bi}, \dots h_{Bk}]^{\mathrm{T}}$ are comparable: if

Tab. 3.3. Example: average height of poset corresponding to Hasse diagram in Figure 3.1.

Subject	Average height
A	4.00
B	2.50
C	2.50
D	1.00

$h_{Ai} > h_{Bi}$, the element $x_i \in X$ is more dominant in the structure A than in the structure B and the difference $h_{Ai} - h_{Bi}$ is a measure of such increment.

3.4.3. Applying posets to multi-indicator systems over time: a new methodological proposal

In the previous pages, I have presented the problem of synthesising a multi-indicator system over time, formalised in the formula 3.13. In this paragraph, I propose a possible solution based on poset. I need to make an important preliminary clarification. The indicators used are quantitative variables. In this application, I do not evaluate any difference or ratio to their values to get distances or proportions. I only analyse the resulting structure of comparabilities: the poset. It can be considered a loss of information. In effect, posets are the natural representation of multidimensional ordinal data [104]. Their use with quantitative data is possible and reduces the set of operations and choices to be made in order to synthesize indicators. For instance, no normalisation and aggregation are used to get scores. Results of poset-based methodologies have to be intended in terms of analysis of the structure of comparabilities. However, they produce an overview useful to socio-economic decisions without assumptions that may distort the results.

The first step of this procedure is to give all indicators the same polarity. In particular, the polarity of all of them must be positive, i.e. the higher the value of the indicator, the better the situation of the measured concept. In this way, nodes in the highest positions of the Hasse diagrams will indicate better situations than those in the lowest positions. If some indicators present negative polarity, we must to invert them taking into account the temporal nature of system. Given the three-way data array $\mathbf{X} \equiv \{x_{ijt} : i = 1, \ldots, N;\ j = 1, \ldots, M;\ t = 1, \ldots, T\}$, in each T matrices $(N \times M)$ we must invert the polarity of the indicators that have it negative. To do this, we can use one of the procedures described in paragraph 3.4.1.

Let's consider a system of 5 units, 3 indicators and two temporal occasions t_1 and t_2. Table 3.4 reports that system. It is a three-way data array $\mathbf{X} \equiv \{x_{ijt} : i = 1, \ldots, 5;\ j = 1, \ldots, 3;\ t = 1, 2\}$ that can be seen as a set of 2 matrices of order (5×3), each of which represents a temporal slice of $\mathbf{Y}$.

Tab. 3.4. Example: a multi-indicator system of five units, three variables and two times.

Subject	A_{t_1}	Y_{t_1}	Z_{t_1}	A_{t_2}	Y_{t_2}	Z_{t_2}
a	0.3	0.5	0.7	0.4	0.5	0.8
b	0.7	0.6	0.8	0.6	0.7	0.9
c	0.2	0.4	0.4	0.7	0.2	0.3
d	0.1	0.5	0.2	0.6	0.7	0.3
e	0.9	0.8	0.6	0.3	0.1	0.1

For each of the two matrices independently, we can calculate the incidence matrix and construct the Hasse diagrams, reported in Figure 3.3. Just from the simple observation of the Hasse diagrams we can draw important information; it is evident, for example, that the relationship structure of the system is different in the two times considered. The aim is to synthesize the system of indicators. To do this, we can calculate the average height of the system considered in the two different times, as reported in Table 3.5.

Tab. 3.5. Example: average height distribution for the multi-indicator system in Table 3.4: times t_1 and t_2.

Subject	Average height t_1	Average height t_2
a	3.000	2.875
b	4.625	4.750
c	1.375	3.500
d	1.875	2.875
e	4.125	1.000

The results obtained allow an *intra-temporal* comparison of the units within the system. For example, we can say that element e is better than element c at time t_1 or that, at time t_2, this situation reverses. Anyway, we cannot make an *inter-temporal* comparison of units. We observe that

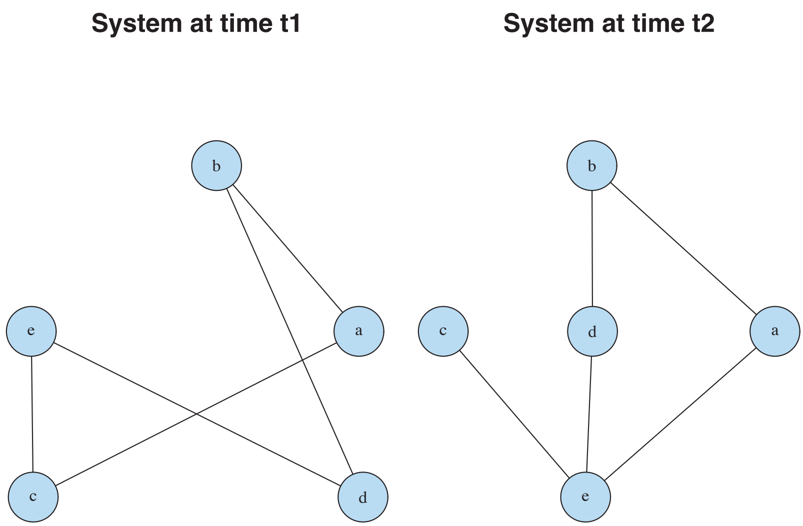


Fig. 3.3. Example of Hasse diagrams of a multi-indicator system in two different times.

the position in the ranking of unit e is worsening from time t_1 to time t_2, but we do not know why this happened. For instance, it may have happened that all the elements had a very marked improvement in the indicators considered, while the element e could have increased slightly and been overtaken by the other units. Another possibility could be that the element e has been drastically reduced from the time t_1 to the time t_2 compared to the other units. To obtain a measure allowing comparisons over time, we must *merge* the posets. Given two finite posets Λ and Π, we merge them by setting $x \leqslant_{\Lambda\Pi} y$ if and only if one of the following conditions is valid:

1. $x, y \subset \Lambda$ and $x \leqslant_{\Lambda} y$;
2. $x, y \subset \Pi$ and $x \leqslant_{\Pi} y$;
3. $x \subset \Lambda$; $y \subset \Pi$ and $x \leqslant_{\Lambda\Pi} y$;
4. $x \subset \Pi$; $y \subset \Lambda$ and $x \leqslant_{\Lambda\Pi} y$.

In other words, by merging the two posets we maintain their initial structures of comparability, adding other comparabilities that are an expression of the temporal comparison among the elements. In this way, it will be possible to make inter-temporal comparisons. By merging the two posets in the example, we obtain the Hasse diagram in Figure 3.4. We calculate the average height of the merged poset, obtaining the results in Table 3.6. In this case, it is possible to compare the same unit over time: for example, we can observe that the unit d is improved from time t_1 to time t_2. In addition, we can compare different units with each other over time: for example, we can say that unit d at time t_2 is worse than unit e at time t_1 and better than it at time t_2.

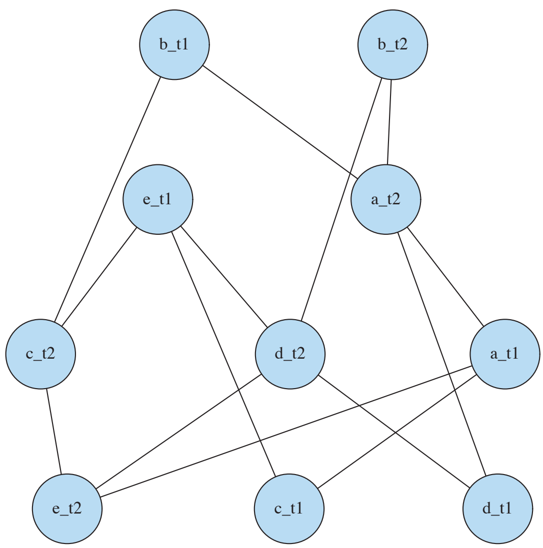


Fig. 3.4. Example of Hasse diagram of a multi-indicator system by merging two different times.

As Marco Fattore [106] states, the average height may solve the problem of getting a ranking out of a partial order, but loses any information on the attribute scores of nodes' profiles. As a consequence, for example, we cannot compare the average heights, across different groups of nodes. To solve this problem and to anchor the average height computation to a *common reference system*, we can introduce the concept of *embedded scale* [106, 16]. The procedure consists of identifying some benchmark profiles that form a scale of increasing levels embedded in the origi-

Tab. 3.6. Example: average height distribution for the multi-indicator system in Table 3.4: merged poset at times t_1 and t_2.

Subject	Average height t_1	Average height t_2
a	4.718	6.611
b	8.984	9.027
c	2.335	4.874
d	2.534	5.511
e	8.655	1.752

nal poset The benchmarks provide points that help anchoring both the comparisons between profiles in the Hasse diagram and the average heights to a reference scale. Figure 3.5 shows the Hasse diagram obtained by adding an embedded three-level scale to the poset described in the Figure 3.4.

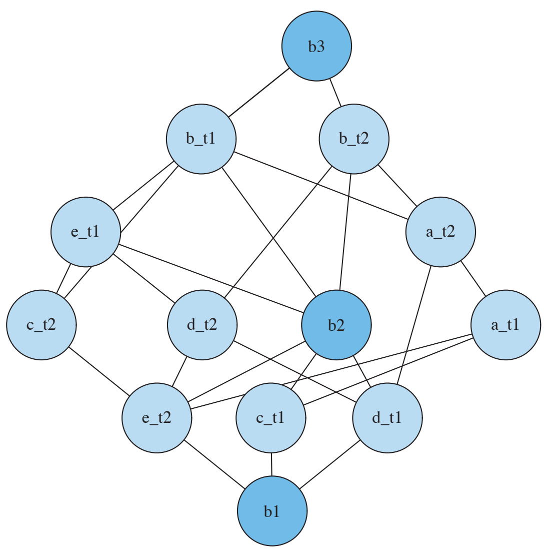


Fig. 3.5. Example of Hasse diagram of a multi-indicator system by merging two different times with embedded scale.

In this example, we add three nodes:

- $b1$ with profile $\{0, 0, 0\}$;

- $b2$ with profile $\{0.5, 0.5, 0.5\}$;
- $b3$ with profile $\{1, 1, 1\}$.

Benchmark profiles get their average heights as well, so that reference points are introduced into the final ranking, partly quantifying it. The embedded scale allows the quantification of reference profiles; it introduces into the evaluation procedure a minimum amount of exogenous information, which is then spread across the poset, consistently with the structure of the partial order relation [106].

Tab. 3.7. Example: average height distribution for the multi-indicator system in Table 3.4: merged poset at times t_1 and t_2 and embedded scale.

Subject	Average height t_1	Average height t_2
a	6.002	8.227
b	10.982	11.008
c	3.253	6.444
d	3.429	7.016
e	10.757	2.800
b1	1.000	1.000
b2	7.084	7.084
b3	13.000	13.000

Table 3.7 shows the average height distribution at the two times obtained with the introduction of embedded scale. The latter allows a more precise measurement, giving some benchmark points.

4. Synthesis of statistical units over time: an application to well-being in the Italian regions

Even if we act to erase material poverty, there is another greater task, it is to confront the poverty of satisfaction - purpose and dignity - that afflicts us all. Too much and for too long, we seemed to have surrendered personal excellence and community values in the mere accumulation of material things. ... Yet the gross national product does not allow for the health of our children, the quality of their education or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials. It measures neither our wit nor our courage, neither our wisdom nor our learning, neither our compassion nor our devotion to our country, it measures everything in short, except that which makes life worthwhile. And it can tell us everything about America except why we are proud that we are Americans ...

Robert F. Kennedy, 18th March 1968

Well-being is a clear example of a complex concept, the analysis and measurement of which has often been marked by profound limitations or even errors. From Robert F. Kennedy's quotation [145], it can be seen that well-being has been and still is considered as an uni-dimensional concept, exclusively associated with the GDP of a nation. This is a clear mistake. "The non-complex vision of the human and social sciences holds that there is a separate economic reality, a psychological reality, a demographic reality, and so on. ... The economic dimension contains the other dimensions and there is no reality that we can comprehend with a single dimension" [205, 45:46].

In this Chapter[1], I deal with the analysis of the concept of well-being, starting from its definition (paragraph 4.1). Among the various alternative measures to GDP, I take into account the one proposed by the Italian National Institute of Statistics (Istat) with the Equitable and Sustainable Well-being (BES) project. From the methodological point of view, a very important issue, the *synthesis of the statistical units*, is addressed (paragraph 3.3). Dealing with multi-indicator systems, we can often note that the synthesis of indicators, although necessary, is often not sufficient to allow us a clear analysis of the phenomenon, especially over time. The presence of dimensions and domains makes the analysis even more complex. The reading of the time series obtained may not be easy and immediate. It is, therefore, often useful to group the units according to several synthetic indicators, expression of different domains or sizes, so as to have an overview of the phenomenon. Clustering techniques are used to do this. In the specific case of well-being (and in general in the study of social complex phenomena), these techniques must have some characteristics to be more suitable (as shown in paragraph 3.3). Considering the time series of the BES synthetic indicators, I use the *Dynamic Time Warping-based Fuzzy C-Medoids model with exponential transformation* (described in paragraph 4.3.2), a model proposed by Pierpaolo D'Urso et al. [93], particularly suitable for the classification of data I deal with. The results (reported and discussed in paragraph 4.4) show that a non-rigid and forced units classification has been obtained, which exactly reflects their trends.

4.1. Well-being: towards a possible definition and measurement

All measurement processes in the social sciences must start with a robust conceptual definition. That is the reason why, in order to measure well-being, we must define it first. This is a very difficult task, due to the *complexity* of this concept. Throughout history, various notions of well-being have been discussed depending on cultural influences and prevailing political regimes [120]. Traditionally, well-being is equated with and *limited* to economic welfare. Historically, this association be-

[1] A previous version of this research work was published [95].

tween well-being and a flourishing economy was born in the early years of the industrial revolution, when the satisfaction of people's basic needs for food and shelter was of paramount importance and, therefore, economic growth was considered the essential precondition to increase quality of life [87]. The Gross Domestic Product (GDP) represents the fundamental measure of the production of each economic system. It is an useful indicator of a nation's economic performance and the most commonly used measure of well-being. However, the latter use of GDP is improper and questionable. Its own creator, Nobel Prize in Economics Simon Kuznets, warned against using this index to measure well-being. First of all, because the well-being of a nation cannot be deduced from a measure of national income. In addition, the use of GDP would lead to an *oversimplification* of the complexity of well-being. "The valuable capacity of the human mind to simplify a complex situation in a compact characterisation becomes dangerous when not controlled in terms of definitely stated criteria. With quantitative measurements especially, the definiteness of the result suggests, often misleadingly, a precision and simplicity in the outlines of the object measured. Measurements of national income are subject to this type of illusion and resulting abuse, especially since they deal with matters that are the center of conflict of opposing social groups where the effectiveness of an argument is often contingent upon oversimplification" [156, 5–6]. As measure of well-being, GDP has some important limitations, including:

- the exclusion of non-market transactions. GDP only considers the value of goods and services traded on the market, while many factors that contribute to people's well-being are not bought and sold;
- the failure to account for the degree of income inequality in society;
- the failure to indicate whether the nation's rate of growth is sustainable or not;
- the failure to consider the costs imposed on human health and the environment of negative externalities arising from the production or consumption of the nation's output;
- treating the replacement of depreciated capital the same as the creation of new capital.

In general, no single economic indicator can provide a complete assessment of the well-being of nations.

What is well-being? Answering this question is a very difficult task. There is not universally accepted definition: different conceptualisations are the combination of heterogeneous components, which assume different meanings in different contexts and cultures. These conceptualisations can be used to quantitatively capture the concept of well-being[2]. According to Brent Bleys [38], the most important conceptualisations are *utilitarianism* (including both the 'revealed preferences' approach and the happiness approach), the *fulfilment of human needs* (including sustainable development) and *capabilities and functionings*. In identifying a possible definition of well-being, the starting point is the consideration that it is involved with the progress of a nation, which is not reduced only to the economic component. From a general point of view, the progress of a nation is defined by "well-being of individuals and society, its fair distribution (*equity*) and sustainable promotion (*sustainability*)" [168, 803]. At individual level, well-being refers to *quality of life*, which, according to Wolfgang Zapf [275], can be structured in two macro dimensions:

- *living conditions*, i.e. outcomes, resources, capabilities, external circumstances, subjective evaluations;
- *subjective well-being*, which can be conceived as a composite construct of two components: the *cognitive* component, related to the process by which individuals evaluate their life; the *affective* component, referring to the emotions experimented by individuals during their lives [85].

Societal well-being involves dimensions such as economic and social cohesion, integration of individuals and groups, social connection, and social ties (social capital), observed at both micro-level and macro-level [52]. The concept of equity concerns the distribution of well-being within population and is related to concepts like inclusion and exclusion. Sustainability refers to the "possible erosion/durability of those conditions with reference to the present and future generations' needs" [168, 808]. No approach is able to fully describe well-being. Different perspectives focus only on certain aspects, not considering the complexity of the phenomenon. Identifying a comprehensive definition of this

2 For a review of different conceptualisations of well-being, see: Des Gasper [116].

concept is difficult. We have to consider the individual and societal level, both defined by objective and subjective aspects and measured through objective and subjective indicators. "A good and healthy society is that in which each individual has the possibility to participate in the community life, develop skills, abilities, capabilities and independency, adequately choose and control his/her own life, be treated with respect in a healthy and safe environment and by respecting the opportunities of future generations" [170, 214]. This can be considered a possible multidimensional definition of well-being.

Well-being is a complex and multidimensional phenomenon, difficult to monitor. A great variety of dimensions of different nature contribute to its definition. Moreover, like others, this phenomenon evolves over time, adapting to the changing needs of individuals. It presents the main characteristics of a *Complex Adaptive System* [197, 255]. It is made up of different dimensions (elements) of different types, which are linked together in linear and non-linear way. Well-being evolves over time, modifying both its dimensions and the links between them. Its measurement requires the definition of systems of indicators capable of capturing its different aspects.

It is clear that GDP is not an ideal yardstick for the well-being of citizens in the various countries, because the latter depends on many other aspects than the economic ones linked to production and consumption [41]. This is clearly stated in the report by the Commission on the Measurement of Economic Performance and Social Progress, the so-called *Stiglitz Commission*[3], probably the most influencing work in this field [251]. "The Stiglitz Commission's recommendations for the measurement of progress reduce the emphasis toward economic indicators in favor of a multidimensional approach that considers social and environmental well-being as important as the economic well-being" [120, 12]. The conclusions of the Stiglitz Commission gave rise to a debate in the international community on identifying measures that could go *beyond GDP*. As Ed Diener and Martin E.P. Seligman [86] point out, countries should establish regular assessments of well-being, complementary to economic indicators as GDP. These measures are important for policy-makers, because economic indicators can omit much of what

[3] The Commission on the Measurement of Economic Performance and Social Progress was set up by French President Nicolas Sarkozy in January 2008. It produced a final report in September 2009 calling for a shift of emphasis from measuring economic production to measuring people's well-being.

is important or give misleading information, as Robert F. Kennedy [145] argues.

Alternative indicators have been developed to provide a more well-rounded measure of a nation's well-being by different national and international organizations. Among the most important, we can remember:

- *Human Development Index* (HDI), based on Amartya Sen's capabilities functionings theory [245]. Elaborated by the United Nations Development Programme (UNDP), it take into account three dimensions (considered the basic capabilities central to human development): a long and healthy life, knowledge and a decent standard of living. The HDI was first calculated in 1990 and the HDI rankings of most of the UN countries are published in the Human Development Report.

- *Index of Social Health* (FISH), published by the Fordham University Institute for Innovation in Social Policy since 1987, as a reliable measure of the social prosperity, especially of the American society [198]. It combines sixteen social indicators, closely linked to both the stages of life and to social institutions such as the labour market, the social welfare programs, the school, and the family. These indicators represent an integral part of society and their monitoring could give information about the quality of life.

- *Better Life Index* (BLI), elaborated by the Organisation for Economic Co-operation and Development (OECD). It is a first attempt to bring together internationally comparable measures of well-being in line with the recommendations of the *Stiglitz Commission*. First published on 24 May 2011, it includes 11 dimensions of well-being (for detailed information, see the BLI website).

- *Genuine Progress Indicator* (GPI). Its aim is to redefine progress developing an economic indicator that attempts to get much closer to the economic reality that people experience. GPI includes more than twenty positive and negative aspects of our economic lives [63]. It uses the same personal consumption data as the GDP but takes into account a number of other factors, such as income distribution or the value of volunteer and housework. The result is a substantively different picture than that presented by the GDP.

- *Social Progress Index* (SPI) measures the extent to which countries provide for the social and environmental needs of their citizens. It includes 54 indicators in the areas of basic human needs, foundations of well-being, and opportunity to progress show the relative performance of nations. The index is published by the non-profit *Social Progress Imperative*. The SPI measures the well-being of a society by observing social and environmental outcomes directly rather than the economic factors [222].

- *Happy Planet Index* (HPI) developed by the New Economics Foundation, combines environmental impact with well-being to measure the environmental efficiency with which people live long and happy lives [180].

All new measures are a combination of both income and non-income variables such as life expectancy, literacy rates, environmental indicators, measures of inequality and so on. By including these variables, they provide a measure of well-being that goes beyond the nation's GDP value. To cover the conceptual dimension's variability, they assume a *multi-indicator approach*, i.e. the adoption of several indicators for each conceptual dimension. This approach allows the overcoming (or, at least, the reduction) of problems produced by the single indicator approach. "In fact, multiple measures allow the conceptual dimensions to be measured with more precision (multiple measures allow random errors to be compensated), accuracy and discriminant capacity" [172, 94].

The *Equitable and Sustainable Well-being* (BES) project, conducted by the Italian National Institute of Statistics (Istat), is probably the most advanced experience in the field of measuring well-being. It is a joint initiative, launched in 2010 by Istat and the Italian National Council for Economics and Labour (CNEL). The project is very ambitious and aims at identifying new indicators for measuring the progress of the country through a particular process able to involve different actors (unions and management, civil society, academic experts). The cornerstone was to consider concepts related not just to macro-economics but also to equity and sustainability with reference to social and environmental dimensions[4]. The BES framework includes 12 well-being domains and

[4] For a detailed summary of the path of the Bes and its methodological developments, see: Italian National Institute of Statistics [136].

around 130 individual indicators drawn from official statistics. This complex multi-indicator system needs the adoption of approaches that would allow for more concise views. As known, *synthesis* is the keyword (as shown in the third Chapter). The adoption of a synthetic approach is the only way for understanding a complex phenomenon like well-being. It is impossible to understand it in its individual parts; it is necessary to renounce to explain analytically the complexity and try to understand the whole system as an indivisible entity. Therefore, in order to understand and measure well-being by means of the multi-indicator system provided by BES, it is necessary to use synthetic measures that allow the vision of the phenomenon as a whole.

The synthesis of the basic indicators of BES was one of the topics of discussion of the Scientific Committee[5]. The study and testing of the synthesis took place from the end of 2010, taking as main reference the volume "Handbook on Constructing Composite Indicators. Methodology and user guide" [213] of the Organisation for Economic Co-operation and Development (OECD) and the Joint Research Center (JRC). To construct the synthetic measures, Istat adopted the *composite indicators approach* (we discuss in detail this topic in paragraph 3.4). Among alternative aggregation methods, Istat chose the *Adjusted Mazziotta-Pareto Index* (AMPI), described in paragraph 3.4. Starting from the 2015 BES Report, Istat calculated composite indicators for each domain on a regional basis, so as to allow a territorial comparison of the levels of well-being of the Italian territory. This is a very important aspect if we consider the differences that have always characterised the country and that find their expression in the so-called North-South gap. The composites are updated annually in the BES Report and the time series are available from 2010, allowing a temporal and spatial comparison.

By constructing composites, Istat synthesises the BES framework in a new multi-indicator system consisting of 1 composite for each BES domain in time series from 2010. It is clear that even the analysis of this *reduced* system is complex and difficult, especially if we want to have a general scenario that takes into account all domains (or groups of them) and allows a comparison of regions over time. One possible way to make this system easier to understand is to use a time series clustering

5 In the BES project there were two different committees: a *Steering Committee*, made up of Istat, CNEL experts and stakeholders, which identified the 12 domains; a *Scientific Committee*, made up of Istat researchers and academic experts, which identified the indicators to measure well-being in each dimension.

approach. We deal with *multivariate time series*, or multivariate time trajectories, which presents a three-way structure "units x variables x times" (paragraph 3.1.1). Starting from Istat composites, I applied a robust fuzzy clustering method based on the Dynamic Time Warping distance and a classification algorithm described in paragraph 4.3. The aim is to analyse the evolution over time of well-being in the Italian regions, finding hidden patterns or similar groups and highlighting their characteristics. In this way, we can examine the changes in well-being and take into account its territorial characteristics, through the identification of clusters of similar regions.

4.2. Description of data

The data source is the freely downloadable BES dataset. In particular, as previously written, we use the time series of the composites. Data are available from 2010 to 2017, but not for all domains. For the domains *Quality of services* and *Income and Inequality*, data are available until 2016. Thus, we have decided to take into account data until 2016 for all composites, to make the analysis homogeneous.

Istat produces a composite for each domain, with the exception of:

- **Work and life balance**, divided into *Employment* and *Quality of work*;
- **Economic well-being**, divided into *Income and inequality* and *Minimum economic conditions*;
- **Safety**, divided into *Homicides* and *Predatory Crime*.

Therefore, we deal with 15 composites observed in time series. Given its complex nature, it may be misleading to analyse well-being by bringing all those indicators together in a single analysis. For instance, Istat decided not to create a single synthetic indicator of well-being, but to leave the various domains separate. However, grouping domains in a meaningful way can certainly be interesting and facilitate an overall reading of the main aspects of the phenomenon. Thus, we decide to group the BES domains using the three dimensions of the main framework of sustainable development: environmental, economic and social. Sustainable development is fundamental to the well-being of society, which must be achieved in each of the dimensions considered. The idea that a good society is not just an economically prosperous society is now universally shared. Other important factors allow the realisation of individual and societal well-being, primarily the social cohesion and inclusion and the quality of the environment. Therefore, it seems reasonable to divide the different BES domains into three macro-domains (corresponding to the three dimensions of sustainable development), as shown in Figure 4.1.

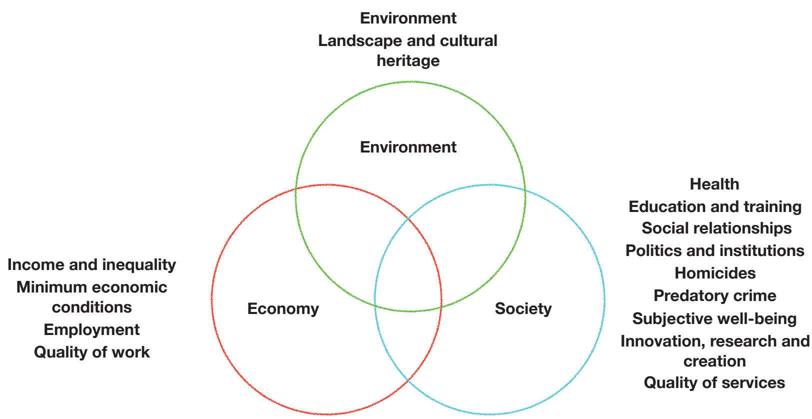


Fig. 4.1. BES Domains: environmental, economic and social dimensions; personal elaboration.

4.3. Methods

As stated in the paragraph 3.3, the statistical methods to classify multivariate time series must present specific characteristics. In particular, the most important elements to consider in clustering dynamic data are, the (dis)similarity or distance measure, the prototype extraction function (if applicable), the clustering algorithm itself, and cluster evaluation [1]. In order to deal with the complexity and uncertainty of this concept, the adoption of a fuzzy approach is certainly the better solution. Furthermore, this is an useful tool for policy makers, because it allows them to identify groups of regions with similar behaviour that can be the target of similar policies and, at the same time, anomalous regions, not comparable with the others, that must be the recipients of targeted and specific policies. The BES time series considered are short; they present only 7 temporal occasions. For this reason, we decide to adopt an observation-based approach. At the same time, we use a PAM and robust approach. The model used that present all this characteristics is the *Dynamic Time Warping-based Fuzzy C-Medoids with exponential transformation* proposed by Pierpaolo D'Urso et al. [93]. In the following paragraphs, we will describe the distance measure (paragraph 4.3.1), the clustering algorithm (paragraph 4.3.2) and the cluster validity index (paragraph 4.3.3) used.

4.3.1. Distance measure: the Dynamic Time Warping distance

The Dynamic Time Warping (DTW) distance is a dynamic algorithm that compares two series and tries to find the optimum warping path between them under certain constraints [28, 232, 140]. It allows the overcoming of some of the limitations associated with the Euclidean distance [140]. DTW for multivariate time series stretches or compresses the patterns of two objects locally, in order to make their shape as similar as possible. It is based on the dilatation or contraction of two (multivariate) time series locally, in order to make their shape as similar as possible.

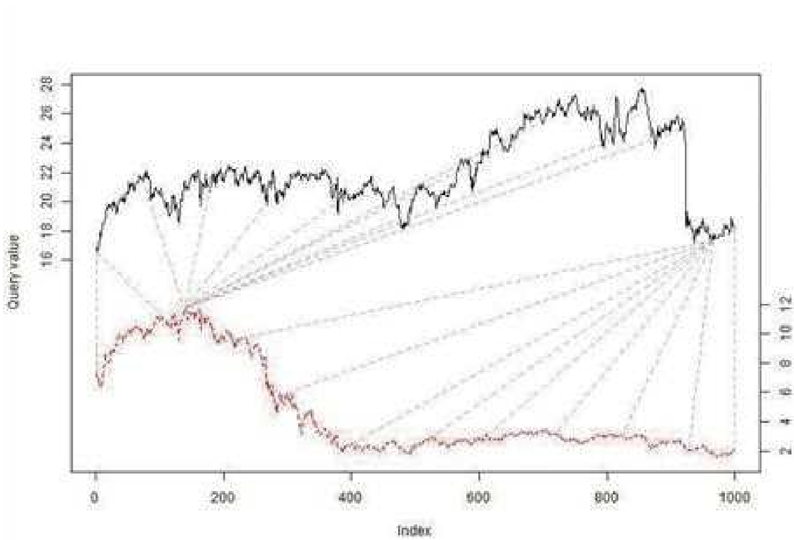


Fig. 4.2. Dynamic time warping distance.

Figure 4.2 shows an example of how DTW works by aligning two time-series. The starting and ending points of the series must match, but other points may be *warped* in time in order to find better matches. The total distance between two time series is computed through the so-called *warping curve* or *warping path*, which ensures that each data point in one time series is compared to the closest data point in the other one. Given two time series $\mathbf{A} = (a_1, \ldots, a_i, \ldots, a_n)$ and $\mathbf{B} = (b_1, \ldots, b_j, \ldots, b_m)$, the *warping curve* $\Phi(k), k = 1, \ldots, T$ is given by:

$$\begin{aligned} \Phi(k) &= (\phi_a(k), \phi_b(k)) \\ \phi_a(k) &\in 1, \ldots, N \\ \phi_b(k) &\in 1, \ldots, M \end{aligned} \tag{4.1}$$

under the following constraints:

1. boundary condition:

$$\Phi_1 = (1,1), \ \Phi_T = (N, M) \tag{4.2}$$

2. monotonicity condition:

$$\begin{aligned} \phi_a(k) &\leq \phi_a(k+1) \\ \phi_b(k) &\leq \phi_b(k+1) \end{aligned} \tag{4.3}$$

The warping curve realigns the time indices of $\mathbf{A}_i$ and $\mathbf{B}_j$ by using the warping functions ϕ_a and ϕ_b. The total dissimilarity between the two *warped* multivariate time series is:

$$\sum_{k=1}^{T} d(\mathbf{x}_{i,\phi_a(k)}, \mathbf{x}_{j,\phi_b(k)}) m_{k,\Phi} \tag{4.4}$$

where $m_{k,\Phi}$ is a local weighting coefficient; $d(\ldots)$ is, usually, the Euclidean distance for multivariate time series [119]:

$$d(i, j) = (\|\mathbf{a}_i - \mathbf{b}_j\|)^{\frac{1}{2}}. \tag{4.5}$$

The DTW distance is the one which corresponds to the optimal warping curve among the several warping curves, $\hat{\Phi}(k) = (\hat{\phi}_a(k), \hat{\phi}_b(k),\ k = 1, \ldots, T$ which minimizes the total dissimilarity between $\mathbf{A}_i$ and $\mathbf{B}_j$:

$$d_{DTW}(\mathbf{A}_i, \mathbf{B}_j) = \sum_{k=1}^{T} d(\mathbf{a}_{i,\hat{\phi}_a(k)}, \mathbf{b}_{j,\hat{\phi}_b(k)}) m_{k,\hat{\Phi}}. \tag{4.6}$$

The robust version is obtained by the exponential transformation of the DTW distance:

$$_{exp}d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_j) = 1 - \exp\left\{-\beta\, d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_j)\right\} \tag{4.7}$$

where β is a suitable parameter (positive constant) determined according to the variability of the data [93]. This transformation assigns *small weights* to distant units in the clustering process guaranteeing robustness of the procedure to outlying data.

4.3.2. Dynamic Time Warping-based Fuzzy C-Medoids model with exponential transformation

For clustering the Italian regions based on the BES time series, we use the Dynamic Time Warping-based Fuzzy C-Medoids clustering model with Exponential transformation (DTW-Exp-FCMd) [93]:

$$\begin{cases} \min: & \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m\ {}_{exp}d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_j) = \\ & \sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}^m \left[1 - \exp\left\{-\beta\, d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_j)\right\}\right] \\ s.t. & \sum_{c=1}^{C} u_{ic} = 1,\ u_{ic} \geq 0 \end{cases} \tag{4.8}$$

where m is a parameter controlling the fuzziness of the partition (we used $m = 2$) and u_{ic} is the membership degree of the unit i-th to the cluster c-th, obtained by solving the formula 4.8 with the Lagrangian multipliers method:

$$u_{ic} = \frac{1}{\sum_{c'=1}^{C}\left[\frac{\left[1-\exp\left\{-\beta\, d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_j)\right\}\right]}{\left[1-\exp\left\{-\beta\, d^2_{DTW}(\mathbf{A}_i, \mathbf{B}_{j'})\right\}\right]}\right]^{\frac{1}{m-1}}}. \tag{4.9}$$

4.3.3. Cluster validity: the Xie-Beni index

It is necessary to define the optimal partitioning of the data into clusters. The procedure of evaluating the results of a clustering algorithm, named *cluster validity*, consists of a set of techniques for finding a set of clusters that best fits natural partitions of given data sets [262]. Even if it deals often with the identification of the "correct" number of clusters C, the cluster validity problem concerns the quality or the degree to which the final partition of a cluster algorithm approximates the real structure of a data set [109, 113]. There are different cluster validity indices for fuzzy methods [272, 75, 55, 265]. We select the optimal number of clusters C by using the Xie-Beni criterion [272]:

$$\begin{aligned}\min_{C\in\Omega_C} : I_{XB} &= \frac{\sum_{i=1}^{n}\sum_{c=1}^{C} u_{ik}^{m} d_{DTW}^{2}(\mathbf{A}_i, \mathbf{B}_j)}{I \min_{j,j'} d_{DTW}^{2}(\mathbf{A}_i, \mathbf{B}_{j'})} \\ &= \frac{S}{I \min_{c,c'} d_{DTW}^{2}(\widetilde{\mathbf{A}}_i, \mathbf{B}_{j'})}\end{aligned} \tag{4.10}$$

where Ω_C represents the set of possible values of C $(C < I)$ and d_{DTW}^2 is a dissimilarity measure (formula 4.7). The numerator of I_{XB} represents the total within-cluster distance, i.e., the objective function of the clustering model considered. The ratio S/I is a measure of internal cohesion of the partition. The smaller this ratio, the more the cohesion of the partition with a given number of clusters. The minimum distance between centroids at the denominator of I_{XB} is called separation. The greater this distance, the more the separation of the partition. The optimal number of clusters C is identified in correspondence with the lower value of I_{XB}.

4.4. Analysis of results and discussion

We analyse the results of each dimension in individual sub-sections. First, we present the multivariate time series belonging to the dimension. Subsequently, we analyse the characteristics of the clusters medoids of the optimal partition, obtained through the application of the cluster validity criterion, and compare them with the national data. Finally, the trends of any fuzzy regions are analysed and compared with those of medoids.

In Table 4.1, we report the results of the application of the Xie-Beni index for $2 \leq C \leq 4$. According to this criterion, the optimal partition is

Tab. 4.1. Cluster validation: Xie-Beni index for different value of C; economic dimension; social dimension; environmental dimension.

Dimensions of well-being	C=2	C=3	C=4
Economic	0.06	0.12	0.63
Social	0.91	0.94	1.56
Environmental	0.27	0.81	0.74

the one that minimizes the index value. For this reason, for all dimensions we choose the solution with two clusters, $C^{*}=2$. For the evaluation of the fuzziness of the clusters, we need to specify a cut-off point for the membership degree. According to [178], if we have a two-cluster situation and the membership degrees in both clusters are between 0.3 and 0.7, it would be considered that there is a reasonable level of fuzziness in the cluster membership of the time series. Consequently, the value 0.7 has been chosen as cut-off. Therefore, those regions that do not have at least that value as membership degree to a cluster are considered fuzzy. For more information on the choice of cut-off, see [178].

4.4.1. Economic dimension

This dimension includes 4 composites:

- *Employment*,
- *Quality of work*,
- *Income and inequality*,
- *Minimum economic conditions*.

Figure 4.3[6] shows their trends for the Italian regions and the country. The so-called North-South gap is quite evident. The central-northern regions, in fact, have trends better than the national ones. On the contrary, in almost all the southern regions (except Abruzzo, whose values are in line with the national ones in some composites) we observe distances from the national average values in all composites.

[6] I realise all figures using R and STATA statistical software. In particular, we use the R packages *ggplot2* [270] for the representation of time series and *corrplot* [268] for the membership degrees matrices; the STATA module *spmap* [221] for the cartograms.

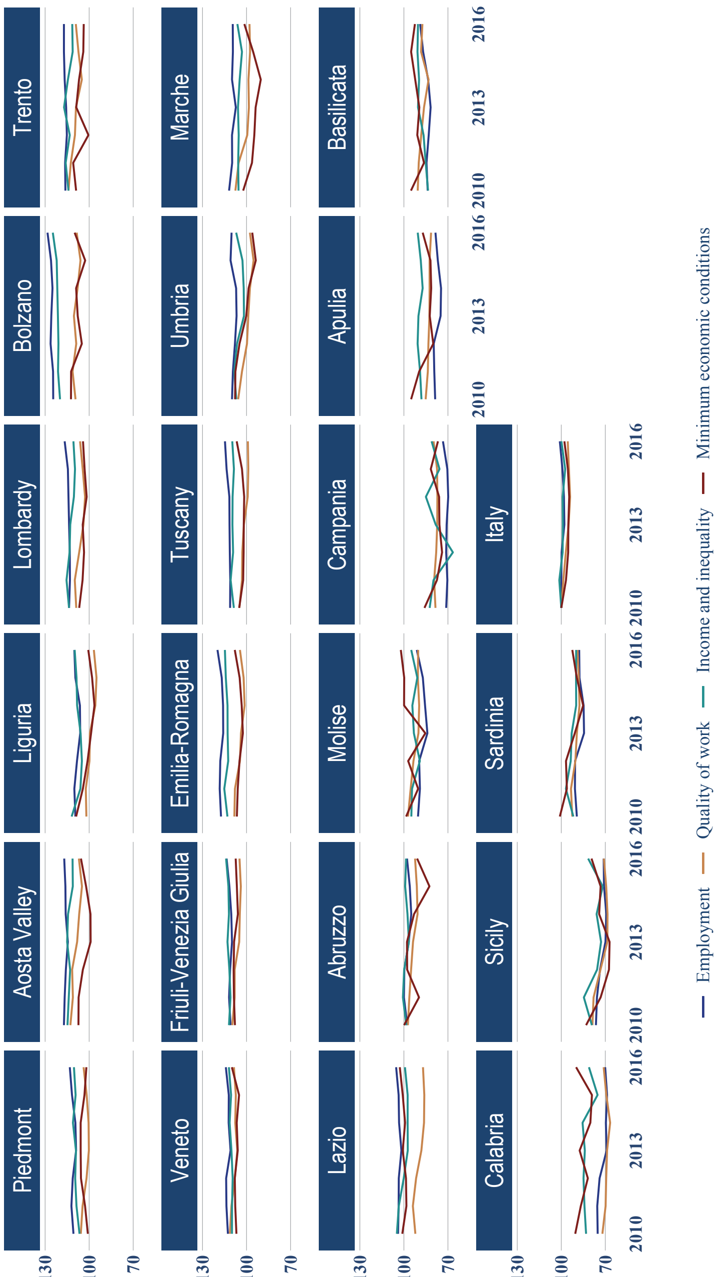


Fig. 4.3. Composites of economic dimension: regional and national data; time series 2010-2016.

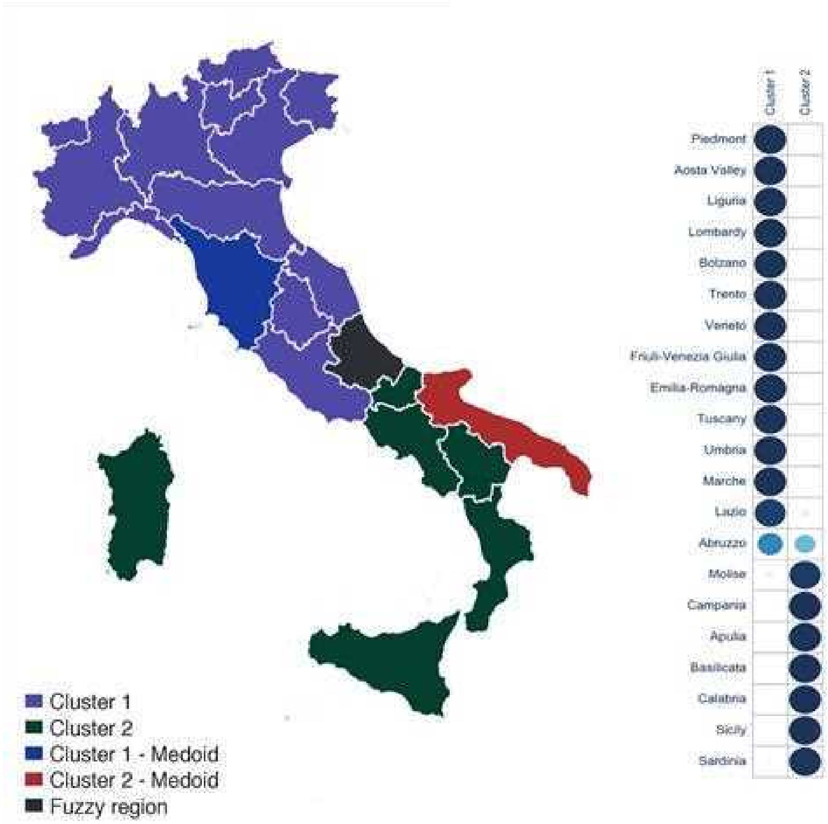


Fig. 4.4. Economic dimension of well-being: clusters' composition and membership degrees.

As shown in Table 4.1, based on the Xie-Beni criterion the optimal number of clusters is $C^* = 2$. The solution identifies 2 medoids, Tuscany for cluster 1 and Apulia for cluster 2, and one region with a fuzzy behaviour, Abruzzo. Figure 4.4 shows the subdivision of the Italian regions according to the cluster to which they belong and the matrix with the membership degrees. The regions' membership to their respective clusters, apart form Abruzzo, is clear and unambiguous. The split of the country is evident, with the northern-central regions all belonging to cluster 1, while the southern regions to cluster 2.

The clusters are clearly characterised; the first has better trends than Italy, while the second worse ones. This is clearly shown in Figure 4.5, which presents the comparison between the two medoids, respectively

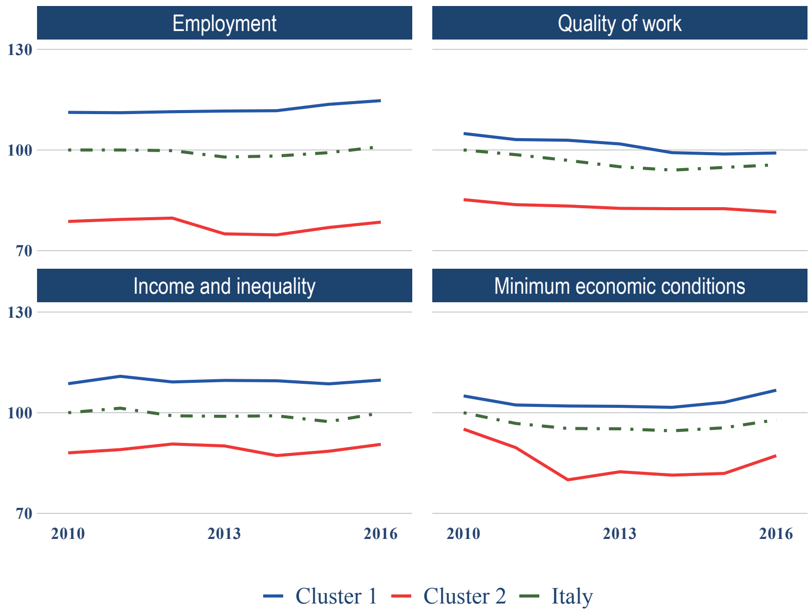


Fig. 4.5. Economic dimension of well-being: comparison among cluster 1, cluster 2 and Italy.

Tuscany and Apulia, and Italy. The economic differences between the areas of the country are confirmed in this analysis as a structural characteristic of the system, rooted in time. In particular, in the *Employment* domain we observe the greatest difference between the two clusters, with an average distance of 23 points[7]; moreover, this distance does not tend to decrease over time (as, for instance, we can observe in the domain *Quality of work*, even if due to a worsening of the regions of cluster 1).

Abruzzo is classified as a fuzzy region; as shown in Figure 4.6, it is perfectly located in the middle of the two clusters (its membership degree is 0.6 to cluster 1 and 0.4 to cluster 2). The region presents values very similar to the national ones in all composites. The trend of the *Minimum economic conditions* domain is particular. It starts from values in line with cluster 2, but from 2012 it deviates significantly from it, due to its improvement and a corresponding worsening of the trend of cluster 2. From 2014, we observe a worsening that leads to a re-alignment to

[7] It should be remembered that the values of the composites are within a range (70, 130). The value only shows the distance between the two time series.

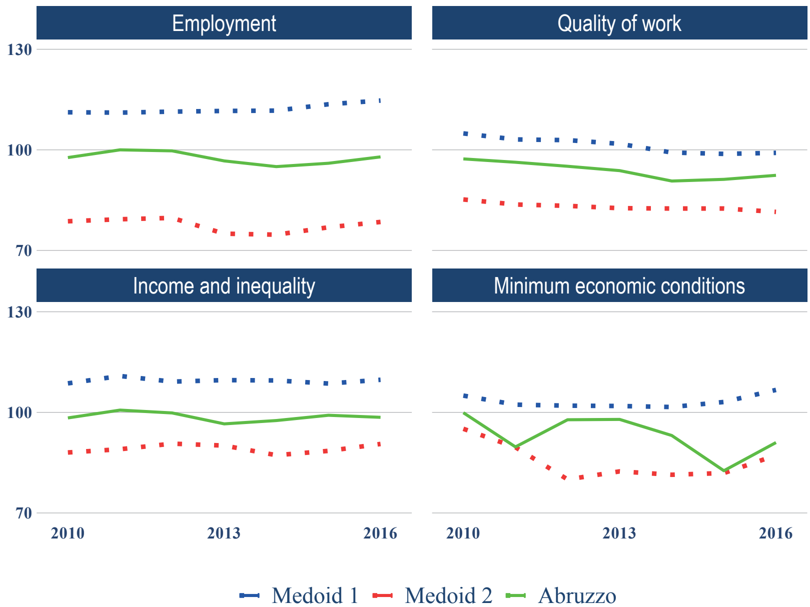


Fig. 4.6. Economic dimension of well-being: comparison between clusters' medoids and Abruzzo.

the values of cluster 2. This trend can be considered as a clear example of the fuzzy nature of many time series.

4.4.2. Social dimension

This is probably the dimension of well-being with the most interesting results. We include 9 composites:

- *Health*,
- *Education and training*,
- *Social relationships*,
- *Politics and institutions*,
- *Homicides*,
- *Predatory crime*,
- *Subjective well-being*,

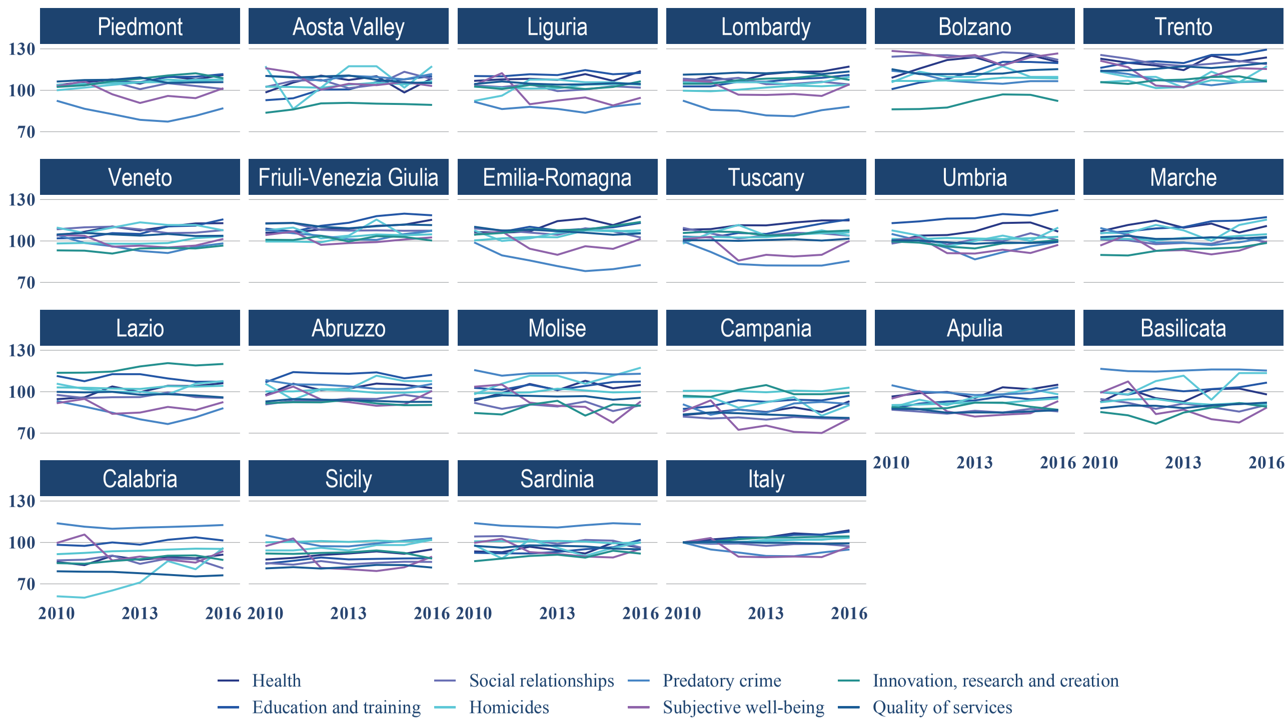


Fig. 4.7. Composites of social dimension: regional and national data; time series 2010-2016.

- *Innovation, research and creation*,
- *Quality of services*.

Looking at Figure 4.7, reporting the trends of the composites from 2010 to 2016, we can observe a situation profoundly different from that of the economic dimension (Figure 4.3). The North-South gap does not appear as clear as in the previous analysis for all domains. For instance, a particularly interesting case to analyse is that of *Predatory Crime* domain. Many northern regions (Piedmont, Liguria, Lombardy, Emilia Romagna, Tuscany) present low values in this domain. These areas have high levels of economic well-being (Figure 4.3); therefore, it is quite obvious to expect that here we can register higher levels of crime such as theft, robbery and pick-pocketing. At the opposite, in some southern regions (Sardinia, Calabria, Basilicata and Molise) there are better values in this composite. There seems to be a relationship between this composite and the *Subjective well-being* one. The perception of security is an important component in determining how an individual is satisfied with his or her life. Therefore, it is reasonable to think that the low levels of subjective well-being observed in some northern regions are partly linked to a higher perception of insecurity, in turn caused by a higher percentage of predatory crime[8].

Another interesting case is that of the domain *Innovation, research and creation* for the Aosta Valley, which presents a very low trend compared to that of Italy (with an average distance of 10 points) and of other composites in the region. The data, which could be considered wrong, can be easily explained if we consider the basic indicators selected for the creation of the composite: the percentage of R&D expenditure on GDP, the percentage of knowledge workers and the percentage of employees in cultural and creative enterprises[9]. The Aosta Valley has values halved compared to the national figure in the R&D expenditure and the percentage of knowledge workers is also much lower.

According to the the Xie-Beni criterion (Table 4.1), the optimal partition is obtained for $C^* = 2$. The two medoids are Veneto for cluster 1

[8] It should be pointed out that the composite on predatory crime is constructed taking into account three basic indicators (the percentages of robberies, burglaries and pick-pockets), which refer to the number of complaints. The latter tends to be underestimated. This underestimation tends to be higher in the south of the country.

[9] For a detailed description of the indicators, see: https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being/indicators.

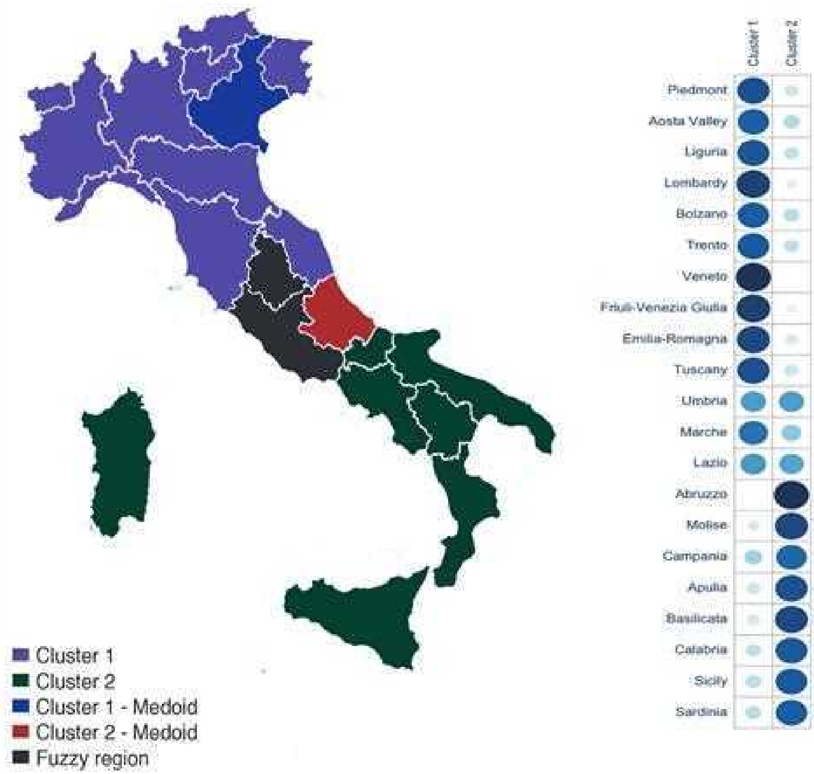


Fig. 4.8. Social dimension of well-being: clusters' composition and membership degrees.

and Abruzzo for cluster 2; there are 2 fuzzy regions, Lazio and Umbria. Figure 4.8 shows regions are divided between the two clusters and the membership degrees matrix. As observed in the previous case (Figure 4.4), there is an evident split in the country, with the north-central regions all belonging to cluster 1 and the southern ones to cluster 2. The membership of the regions is clear, but we can see a less clear-cut situation than that of economic dimension. Indeed, some regions, although not fuzzy, present a considerable membership degree in the cluster they do not belong to (for instance, Marche, Aosta Valley, Campania. See Table 4.2 in Appendix for the values).

Looking at the characteristics of the two medoids (Figure 4.9), cluster 1 has better values than cluster 2 in almost all time series, except for

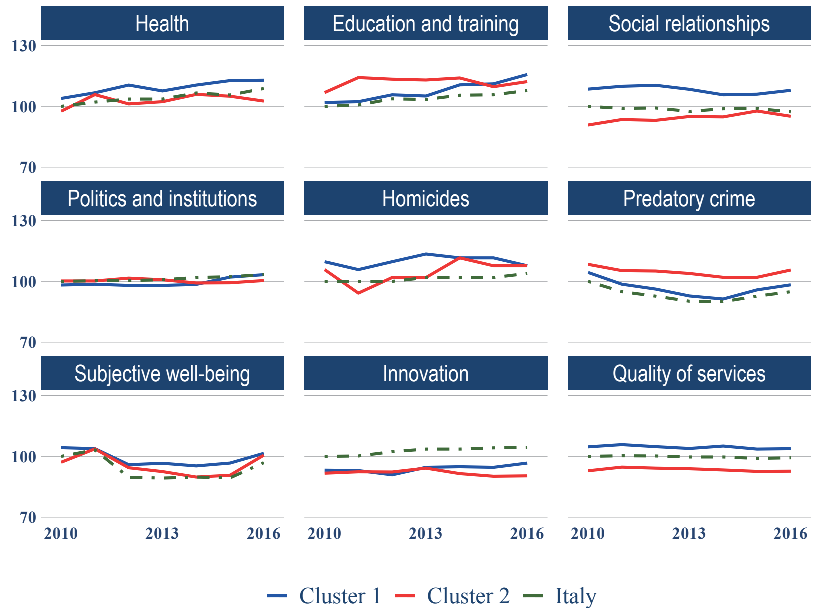


Fig. 4.9. Social dimension of well-being: comparison among cluster 1, cluster 2 and Italy.

Predatory Crime (this confirms what we wrote about the time series). Contrary to what emerges from Figure 4.5, Italy is not always exactly in the middle of the two clusters[10]. Moreover, some domains show interesting trends. In the *Education and training* domain, there is a reversal of the trends. Indeed, cluster 1 registers values lower than those of cluster 2 from 2010 to 2014; starting from 2015, the situation changes, mainly due to a drastic worsening of cluster 2 (the composite passes from 114 to 109). The values return to growth in 2016, but do not reach those of cluster 1. In the *Innovation, research and creation*[11] domain, we observe a situation in which the two medoids start from very close values, but then differentiate over time (from 2013, cluster 1 has a trend better than cluster 2). In addition, both medoids have trends lower than the national average.

The chosen solution identifies 2 fuzzy regions, Umbria and Lazio, whose trends are shown in Figure 4.10, compared with those of the 2

---

10 In particular, we can observe this situation in the domains: *Education and training*, *Predatory crime*, *Subjective well-being* and *Homicides*.

11 In figures 4.9 and 4.10, we have renamed this domain, using "*Innovation*", for reasons of space.

medoids.

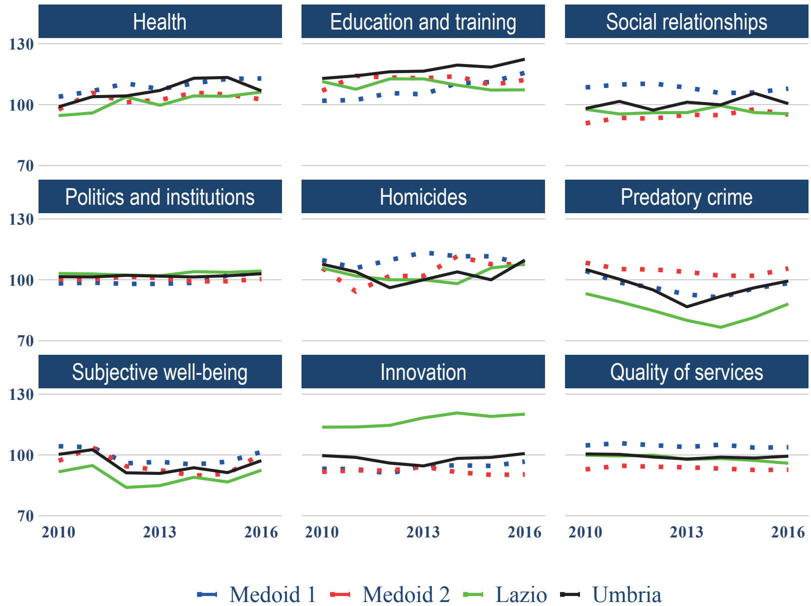


Fig. 4.10. Social dimension of well-being: comparison between clusters' medoids, Lazio and Umbria.

Umbria can be considered perfectly fuzzy; its membership degree is 0.5 to cluster 1 and 0.5 to cluster 2. This region is similar to cluster 2 in *Homicides*, *Subjective well-being* and *Education and training*; it is between the two clusters in *Quality of services*; in the remaining domains, it shows similar trends to those of cluster 1.

Lazio has a membership degree of 0.52 to cluster 1 and 0.48 to cluster 2. It is between the two clusters in *Quality of services* (like Umbria) and in line with cluster 2 in *Homicides*, *Health* and *Social relationships*. In *Predatory crime*, the region is considerably below the values of cluster 1 (on average, 10 points below during the period considered). To fully understand the value of the composite, it is necessary to examine the basic indicators used for its construction: in particular, Lazio has the highest pick-pocketing rate among the Italian regions (on average, in Italy this rate is 6.8‰, while in Lazio it is 12.7‰). In *Innovation, research and creation*, the region has the highest value among all regions, mainly due to the basic indicator "Impact of knowledge workers on employment".

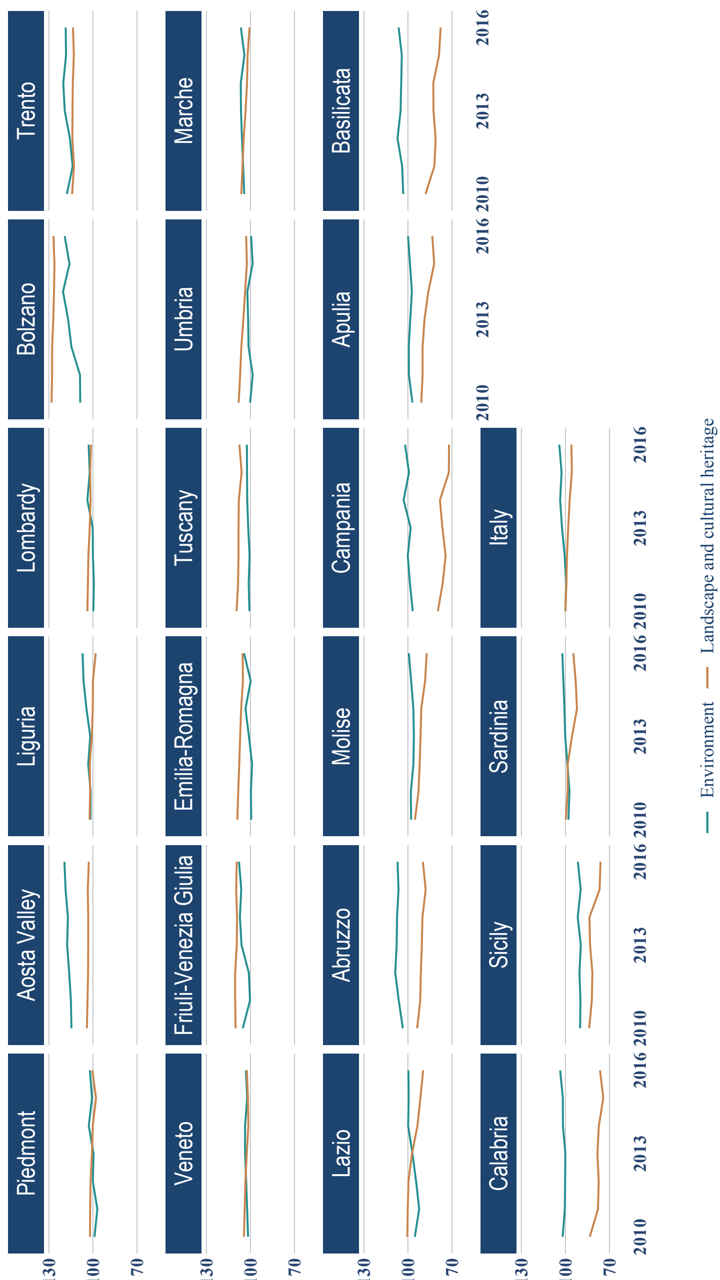


Fig. 4.11. Composites of environmental dimension: regional and national data; time series 2010-2016.

4.4.3. Environmental dimension

In this dimension we include 2 composites:

- *Environment*,
- *Landscape and cultural heritage*.

As we can see from Figure 4.11, the northern areas have better trends than the southern ones, especially in the *Landscape and cultural heritage* domain.

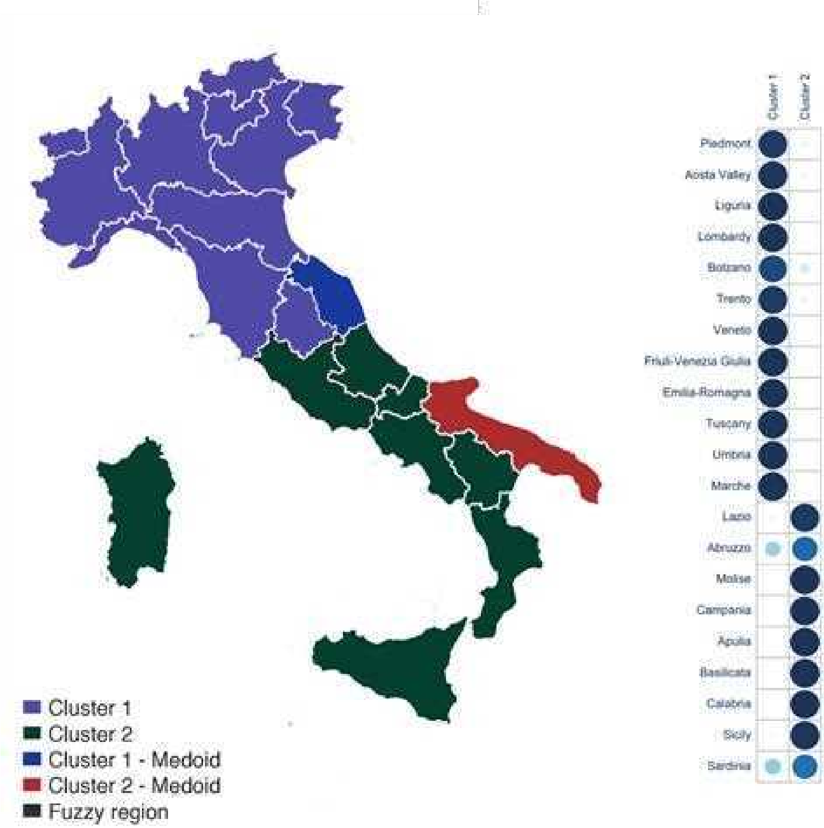


Fig. 4.12. Environmental dimension of well-being: clusters' composition and membership degrees.

As in the previous cases, also in this one we have two clusters, represented by Marche (cluster 1) and Apulia (cluster 2); there is no fuzzy region. All north-central regions belong to cluster 1 (except Lazio, included in cluster 2) and all southern ones to cluster 2; the membership of the regions is clear and unambiguous (see Figure 4.12).

In Figure 4.13, we report the medoids' trends compared to that of Italy. It is evident that the medoid 1 represents those regions with values better than the Italian ones; at the opposite, cluster 2 includes all those regions with trends worse than the national one, especially in *Landscape and cultural heritage*.

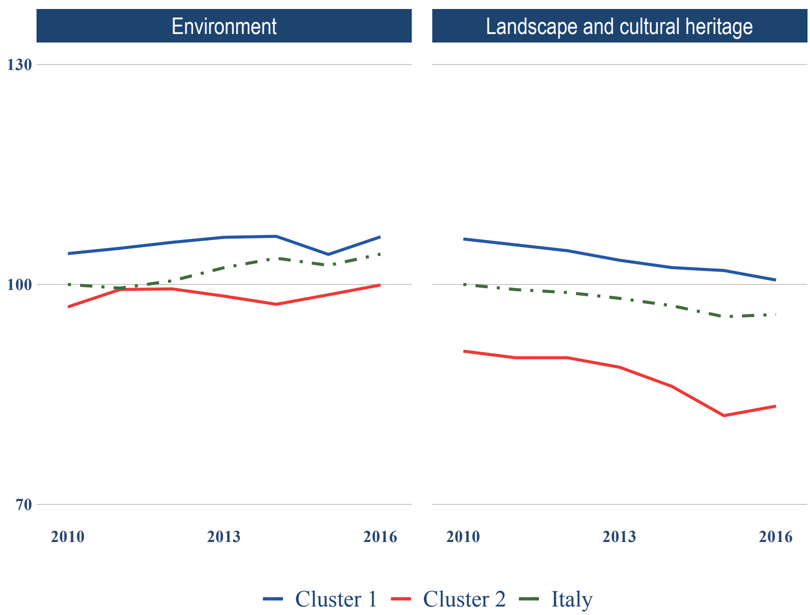


Fig. 4.13. Environmental dimension of well-being: comparison among cluster 1, cluster 2 and Italy.

4.5. Final and conclusive remarks

Well-being is a multidimensional and complex phenomenon, whose measurement requires a multi-indicator approach and the creation of a system of basic indicators, capable of taking into account its different aspects. The complexity of this concept is difficult to be managed and its analysis could be not easier. In other words, the picture depicted by the indicators can be difficult to be interpreted. That is why the adoption of some synthetic views could help in obtaining an overall view. In most cases, the synthesis focused on the indicators that are part of the system; however, it may not be sufficient to offer a stylised view of the phenomenon. The system of synthetic measures, although reduced, may not be easy to read. Indeed, the synthesis concerns not only the indicators but also units composing the observed reality.

We have examined the system of composite indicators constructed by Istat for the measurement of BES. By the application of a time series clustering method, we have tried to synthesize the information contained in the initial system, avoiding excessive simplification.

The clustering model used allows the study of well-being taking into account its complexity. In fact, the model has a series of characteristics that make it particularly suitable for this analysis. The identification of the medoids allows a clear characterisation of the different regions simplifying, without trivialising, the understanding of the phenomenon. The adoption of a fuzzy approach makes it possible not to *crush and flatten* the differences, highlighting the presence of regions with behaviour that is not clearly classifiable, which are often the most interesting.

The analyses have shown that the so-called North-South gap is clear. For all three dimensions, two clusters are identified, composed more or less by the same regions; one cluster gathers the regions with trends that tend to be better than the national ones and one cluster with the worse trends. However, in the social dimension we have seen how some composites have a behaviour that is not always so clear. Moreover, the model has allowed us the identification of the regions with behaviours that differ from the clusters and to highlight their specificities.

Tab. 4.2. Membership degrees: economic, social and environmental dimensions.

Regions	Economic dimension		Social dimension		Environmental dimension	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Piedmont	1.00	0.00	0.85	0.15	0.95	0.05
Aosta Valley	1.00	0.00	0.79	0.21	0.97	0.03
Liguria	1.00	0.00	0.83	0.17	1.00	0.00
Lombardy	1.00	0.00	0.92	0.08	1.00	0.00
Bolzano	0.99	0.01	0.8	0.2	0.89	0.11
Trento	1.00	0.00	0.82	0.18	0.95	0.05
Veneto	1.00	0.00	1.00	0.00	1.00	0.00
Friuli-Venezia Giulia	1.00	0.00	0.93	0.07	1.00	0.00
Emilia-Romagna	1.00	0.00	0.89	0.11	1.00	0.00
Tuscany	1.00	0.00	0.85	0.15	1.00	0.00
Umbria	1.00	0.00	0.50	0.50	0.99	0.01
Marche	1.00	0.00	0.69	0.31	1.00	0.00
Lazio	0.93	0.07	0.52	0.48	0.04	0.96
Abruzzo	0.60	0.40	0.00	1.00	0.26	0.74
Molise	0.05	0.95	0.11	0.89	0.00	1.00
Campania	0.00	1.00	0.25	0.75	0.01	0.99
Apulia	0.00	1.00	0.15	0.85	0.00	1.00
Basilicata	0.00	1.00	0.12	0.88	0.01	0.99
Calabria	0.01	0.99	0.18	0.82	0.01	0.99
Sicily	0.00	1.00	0.19	0.81	0.02	0.98
Sardinia	0.02	0.98	0.20	0.80	0.29	0.71

5. Synthesis of statistical indicators over time: an application to sustainable development in the Italian regions

We fundamentally depend on natural systems and resources for our existence and development. Our efforts to defeat poverty and pursue sustainable development will be in vain if environmental degradation and natural resource depletion continue unabated.

United Nations General Assembly [258, 19]

Sustainable development is a crucial issue in the institutional and academic debate. Like well-being, this is a complex concept whose meaning has changed over time. This complexity is clear from the quotation of the United Nations Secretary-General at the beginning of this Chapter. Pursuing sustainable development involves the consideration of different important aspects, first of all the environment.

In this Chapter[1], I deal with the analysis of the concept of sustainable development, starting from its definition (paragraph 5.1). The concept is reconstructed by taking into consideration the main stages of its institutional path. The main problems linked to this concept and its correct definition are also presented. The paragraph 5.1.1 deals with the territorial declination of sustainable development goals. This is a crucial topic in the international debate, which has been addressed in this research work to provide a framework for the analysis that is carried out, which takes into account the Italian regions.

[1] A previous version of this research work was published in Leonardo S. Alaimo et al. [7].

From the methodological point of view, a very important issue is addressed, the *synthesis of the statistical indicators*. I present the main aspects of this topic in paragraph 3.4. In particular, I apply two methods of synthesis. An aggregative-compensative method that has a well-established literature, the Adjusted Mazziotta-Pareto Index ($AMPI$), presented in paragraph 3.4.1. The other is a *new* method, of non-aggregative type, based on the Partial Order Theory, illustrated in the paragraph 3.4.3. The element of innovation is the proposal of a poset application for the synthesis of longitudinal multi-indicator systems. To test the validity, the proposed method are compared with $AMPI$, the most robust of aggregative methods [183, 185]. I compare the results obtained by different methods by using the correlation coefficient for repeated measures presented in paragraph 3.2. The syntheses are preceded by an exploratory analysis, carried out by calculating the coefficients of correlation *between* and *within* observations and the principal component analysis on average (the methods are described in paragraph 3.2).

Starting from the assumption that sustainable development cannot be considered as a unique multi-indicator system, I have taken into consideration 2 of the 17 Sustainable Development Goals (SDGs), goal 1 (*End poverty*) and goal 3 (*Health and well-being*). The statistical units are the Italian regions. For each goal, 2 dimensions have been identified, *present condition* and *risk*, which allow a full understanding of the concept of sustainable development (paragraph 5.2). The analyses were conducted separately for the dimensions of each goal and the results presented and discussed in specific sections (paragraph 5.4).

5.1. Sustainable development: a *contested* concept

Sustainable development is a main issue of international debate about human society and its future. It is one of the most challenging concepts ever developed with the aim of ensuring a dignified life in society for everyone. The term does not have a univocal definition in the literature, although it is widespread and there are several research trends in this field [269]. As shown by Robert B. Gibson, Selma Hassan, and James Tansey [118], different disciplines have contributed to the sustainability debate. Thus, there are many definitions currently in circulation, often divergent from one another [118, 100, 123]. Moreover, the concept has changed over time, taking on new meanings related to the different phases of the international debate. We can consider sustainable development as a *contested concept* [53]. It has an ambiguous and vague definition leading to different interpretations and meanings. For this reason, we reconstruct the sense of this concept through a brief analysis of the main phases of the international debate on this issue.

The concept of sustainable development is closely linked to that of sustainability, i.e. the capacity of any system to maintain itself indefinitely. The importance of sustainability for economic growth was underlined in 1972 by the so-called Club of Rome[2]. One of the conclusions of its book, *Limits to Growth*, is "if the present growth trends in world population, industrialisation, pollution, food production and resource depletion continue unchanged, the limits to growth on this planet will be reached sometime within the next one hundred years. The most probable result will be a rather sudden and uncontrollable decline in both population and industrial capacity" [190, 23]. The Club of Rome bases its analysis on the awareness of social and environmental negative consequences linked to an idea of development focused only on economic growth and technological progress. The authors underline that growth is a limit to development, and focus for the first time on the need for development which takes into account the scarcity of resources.

The term sustainable development powerfully entered the international debate in 1987 via the so-called Brundtland Commission and its report, *Our Common Future*, which was the response to a long debate, begun after the conclusions of the analysis of the Club of Rome, about the negative impact of human activities on the natural environment. "Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [266, 41]. This may be considered the first definition of sustainable development, emphasising its inter-generational aspect. It is synthetic, but undoubtedly exhaustive: it is delineated not only in the field of economic quantities, but it is multidimensional, rich and varied. Its objective is a satisfaction of needs that is *constrained*: the constraint is determined by the need to ensure a living space for development also for the future generations. "In essence, sustainable development is a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development;

[2] The Club of Rome commissioned a group of researchers from the Massachusetts Institute of Technology (MIT) to carry out a study to investigate the causes and the long-term consequences of the growth of five variables: population, industrial capital, food production, consumption of natural resources and pollution [157, 1-2].

and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspirations" [266, 43]. There has been a transformation of the meaning of sustainability, no longer focused only on scarcity of resources and the importance of their conservation (such as in *Limits to Growth*), but also on the satisfaction of the actual and future generations' needs. At the same time, it is clear that sustainable development is not an aim in itself, but an instrument that must ensure the achievement of actual and future generations' needs. The final objective is the creation of well-being in a twofold perspective. On the one hand (*objective well-being perspective*), it must create and ensure the appropriate material, social and political conditions available to the entire population to live a good life; on the other hand (*subjective well-being perspective*), it must guarantee opportunities and experiences for individuals to meet the needs of their lives. The *Rio Declaration on Environment and Development*, adopted by the United Nations Conference on Environment and Development held in Rio de Janeiro in 1992, emphasises this concept, stating that human beings are at the centre of concerns for sustainable development and the right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations [257]. The basis of this idea of sustainability is no longer the view economic growth as a limit: the prerequisite for sustainable development is the relationship between human activities, including the economic ones, and the environment, which does not diminish the chances for future generations to meet their needs. Furthermore, for the first time the Rio Summit underlines the importance of defining policies aimed at achieving sustainable development, by identifying 27 principles and an action program, the so-called *Agenda 21*, to help governments in obtaining this goal.

According to Jeffrey D. Sachs, the definition of sustainable development has evolved over time. The inter-generational equity, the need for preserving resources for future generations, is one of the main characteristics of this concept. However, the only way to achieve this goal is to conceive development as a multidimensional concept taking into account economic, social and environmental aspects. Nowadays, the definition focuses on this "*holistic approach* linking economic development, social inclusion and environmental sustainability" [238, 6]. The long-term stability of society is only achievable through the integration of these three pillars. This is clearly stated in *The Future We Want* [259], stressing the need to pursue development by achieving economic, social and environmental sustainability. "Sustainable development demands action on its three dimensions and as long as these are activated through policies fostering economic growth, greater social equality and the reduction of negative environmental impacts, the needs of current and future generations are expected to be enhanced" [124, 13]. So, we can consider sustainable development as a *three-way holistic framework*, involving three complex systems, economic, social and environmental, interacting with one another. We can better understand the importance of these components by reporting the definition of development proposed by Amartya Sen: "development can be seen as a process of expanding the real freedoms that people enjoy" [246, 3]. This definition focuses on human freedoms. Undoubtedly, economic growth plays a central role in the satisfaction of human freedoms. However, the idea that a good society is not just an economically prosperous society is now universally shared. "From the general point of view, societal well-being involves dimensions like economic and social cohesion, integration of individuals and groups, social connection and social ties (social capital), referring to dimensions observed at both macro and micro level" [169, 100]. Thus, there are also other important factors allowing individuals the realisation of their needs and freedoms, primarily the social cohesion and inclusion and the quality of the environment in which they live. This is the central point that allows us to move from the concept of development to that of sustainable development. It implies a new vision of society; it is a new way of understanding our lives.

Since the publication of the Brundtland Commission's report, the concept of sustainable development has been criticized, mainly because it seems too confusing and contradictory to be useful in practice [118]. Some scholars reject the very idea that development could be sustainable. Serge Latouche, for instance, criticises this concept, defining it a *mystification*. "The term is so broad that it can be applied to anything and everything" [161, 10]. He defines it an *oxymoron*, highlighting that the only development we know is that arising from the industrial revolution: an economic war among men and against nature. It is senseless to define the development as sustainable, because it is against its own nature. Furthermore, in the literature there is no consensus about the three-way framework. "One of the main obstacles to developing a common conceptual framework incorporating social, economic and ecological problems is the lack of genuine consensus among experts in each discipline as to how ecological, economic and social systems relate to one another" [100, 40]. In other words, the debate focuses on the role, the *weight*, to give to each component and their mutual relations. The three pillars (economic, social, and environmental) are one such framework, but many others are possible. Robert K. Turner, David W. Pearce, and Ian Bateman [256] suggest that the various approaches and definitions differ from each other because they are linked to two opposite perspectives, respectively labelled as *strong* and *weak* sustainability. The first one, the *ecocentric* perspective, seems to come "close to rejecting even a policy of modified development based on the sustainable use of nature's assets" [256, 54]. It emphasises environmental protection. Indeed, one of the main criticisms levelled at the three-way framework is its anthropocentric vision, which considers environment instrumental and subjected to human needs, putting them at the center of sustainable development. This principle is strongly rejected by the different ecocentric perspectives[3] developed over the years. According to Wolfang Sachs, "sustainable development calls for the conservation of development, not for the conservation of nature" [239, 34]. Human society is only a part of nature, and environment cannot be considered as a dimension of sustainable development. Nevertheless, it is the necessary condition for any kind of human activity, including the development. The objective of sustainability should be to limit, or even halt, economic growth and the use of particular resources. On the contrary, the *technocentric* perspective focuses on the free markets and argues that the sustainability notion is too vague to be helpful. Thus, nature is only instrumental in satisfying human needs. This perspective is based on a different understanding of the role of economic dimension, in terms of both development and growth. Creating well-being is only achievable through increasing the value of total capital. As underlined by Robert Solow, we do not have to worry about the scarcity of resources: the only thing we have to leave to future generations is the capacity to create well-being and not some particular natural resource [101]. Starting from Robert K. Turner, David W. Pearce, and Ian Bateman's reflection, the notions of weak and strong sustainable development have been debated in the recent literature, and a number of indicators or frameworks have been proposed to capture them. For weak sustainability, efforts have focused on the possible transformation of the well-known macroeconomic indicators, gross national product and

---

[3] For a complete analysis of the ecological paradigm on sustainable development, please see: Jennifer Elliott [100].

gross domestic product, into an indicator of sustainable development. For strong sustainability, the concept of critical natural capital has been introduced for capital stocks that cannot be replaced by other stocks of environmental or other capital to perform the same functions [199]. Some authors have also criticised the three-way framework, estimating that the pillars that underpin sustainable development are more than three. Jeffrey D. Sachs identifies four major dimensions, adding to the three traditional ones the *good governance*, "playing a central role in the eventual success or failure of SDGs" [238, 502]. The importance of good governance is also underlined in *The Future We Want*: "we acknowledge that democracy, good governance and the rule of law, at the national and international levels, as well as an enabling environment, are essential for sustainable development, including sustained and inclusive economic growth, social development, environmental protection and the eradication of poverty and hunger" [259, 2]. Other authors [131] incorporate the *institutional* dimension. Keith Nurse [212] argues that sustainable development manages not only social, economic and environmental capital, but also the *cultural* one. Some researchers have discussed more than four dimensions of sustainable development [216, 244]; others consider it as the interaction of the environmental and human systems in a *two-part coupled* framework [224]. Figure 5.1 reports the main alternatives to the three-way holistic framework. Despite the criticism, the tripartite model, as elaborated in Agenda 21, remains dominant and hegemonic in the literature and it is the basis of the indicators' system proposed by the United Nations.

5.1.1. Sustainable Development Goals and Indicators: the issue of their regional declination

Even if sustainable development can be considered a global issue, the question related to how to achieve it remains. The Brundtland Commission definition is certainly evocative, but it is equally difficult to make it operational. The three-way framework is based on an anthropocentric vision, according to which in order to be sustainable, development should ensure the satisfaction of the needs and the well-being of present and future generations, by combining economic growth, social inclusion and environment and setting measurable goals focused on priority areas. Thus, achieving sustainable development implies also a *normative approach*. Governments must define appropriate policies for increasing

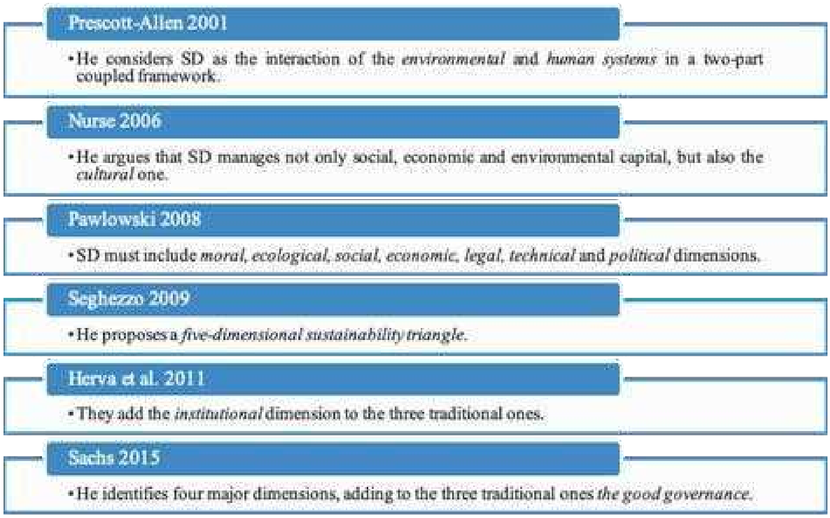


Fig. 5.1. Dimensions of sustainable development: different theories.

current well-being and, at the same time, not reducing that of future generations. They have to make choices on how best to use its total capital stock today to increase current economic activities and welfare and to save or even accumulate for future generations. In order to help and guide governments in achieving these objectives, sustainability appraisal – i.e. "an integrated assessment of the environmental, social and economic effects of proposed actions at all levels of decision making" [74, 368] – has become a fundamental decision support tool [47].

The individuation of a set of goals and the definition of an indicator framework are undoubtedly useful for defining and assessing policies and actions. The United Nation Conference on Sustainable Development held in Rio on June 2012, also known as *Rio+20 Summit*, identified a number of principles that should inspire the definition and the choice of the Sustainable Development Goals (SDGs). "SDGs should be action-oriented, concise and easy to communicate, limited in number, aspirational, global in nature and universally applicable to all countries" [259, 47]. Over the next three years, it developed an intensive debate involving governments, civil society and other stakeholders around the world, which led to the adoption of the so-called *Agenda 2030* at the United Nations Sustainable Development Summit in September 2015. The SDGs form a part of the Agenda 2030: they are a framework of 17 goals and 169 targets across social, economic and environmental areas of SD, defined according to the principles of *Rio+20 Summit*. Goals were selected to cover the three traditional dimensions of sustainable development. These goals are the successors and the evolution of the *Millennium Development Goals* (MDGs). The latter were established in 2000 by the Millennium Summit of the United Nations; they were eight international development goals to be achieved by 2015. While the MDGs focused only on social issues and were targeted at developing countries, the SDGs provide goals and targets in three dimensions and are applicable to all countries, also to the developed ones. However, it should be noted that not all the targets are applicable to all countries and in the same way, as clearly stated in the Agenda 2030. "Targets are defined as aspirational and global, with each Government setting its own national targets guided by the global level of ambition but taking into account national circumstances. Each Government will also decide how these aspirational and global targets should be incorporated into national planning processes, policies and strategies" [260, 13]. For this reason, it becomes essential to identify a global indicator framework to know and monitor the situation of each country with respect to each goal and target, to be able to plan and implement actions that take into account the strengths and weaknesses of the different national realities. "Indicators are being developed to assist this work. Quality, accessible, timely and reliable dis-aggregated data will be needed to help with the measurement of progress and to ensure that no one is left behind. Such data is key to decision making" [260, 12]. The need to assess the sustainable development of societies has grown together with the importance of this issue in the international debate, in public opinion and among stakeholders; at the same time, it is no easy task [114].

The starting point is the definition of an indicators' framework, whose main aim is to provide an information-driven architecture that will help in the definition of policies. Indicators should measure characteristics of the human society that ensure its continuity and functionality far into the future. They serve as monitoring and signalling mechanisms. "The optimal sustainability indicators are those that capture the essential characteristics of the system and show a scientifically verifiable trajectory of maintenance or improvement in system functions" [199, 3]. The global indicator framework was developed by the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) and adopted by the United

Nations General Assembly on 6 July 2017. It includes 232 indicators[4], divided among the 17 goals and 169 targets.

At this point, it is necessary to understand whether SDGs could really be a system of indicators. In their general statement, and in their further specification into targets, the 17 goals are a mixture of the causal chain from inputs to outputs to outcomes [142]. Therefore, the 17 goals cannot be considered a system of indicators. They represent different things:

- domains (Life below Water, Life on Land, Industry and Infrastructure);
- conceptual issues (Gender Equality, Good Health and well-being), crossing different domains;
- goals (no Poverty, no Hunger).

The most correct way of conceiving SDGs appears to be to consider them as *alarm bells*, which refer to (real or ideal) systems of indicators. In other words, they seem to be extrapolated from a system, since they stand out particularly serious situations, urging *ad-hoc* policies. Since SDGs are not a system, it does not seem possible to define a unique synthetic measure for sustainable development. Making a synthesis in this context can only be meaningful to do a summary comparison in terms of time and space, but nothing more. In this way, they perform a function of analysis and monitoring, useful for defining direct and specific actions.

According to the *Agenda 2030*, data should also be collected and reported sub-nationally: each government must develop indicators, at national and regional level, which complement the global framework, giving attention to the territory. Not surprisingly, the territory can be considered as the result of the interaction of the same subsystems (environmental, economic and social) of sustainable development according to the three-way framework. Each human group lives in a specific territory, defined as a geophysical space, in which certain economic and social relations are developed. Thus, every human group has a specific territorial localisation and corresponds to a specific social and cultural identity. Therefore, it is fundamental in the process of defining policies and actions taking into account not only the national specificities, but

[4] The total number of indicators is 244. Nevertheless, some indicators repeat under two or three different targets.

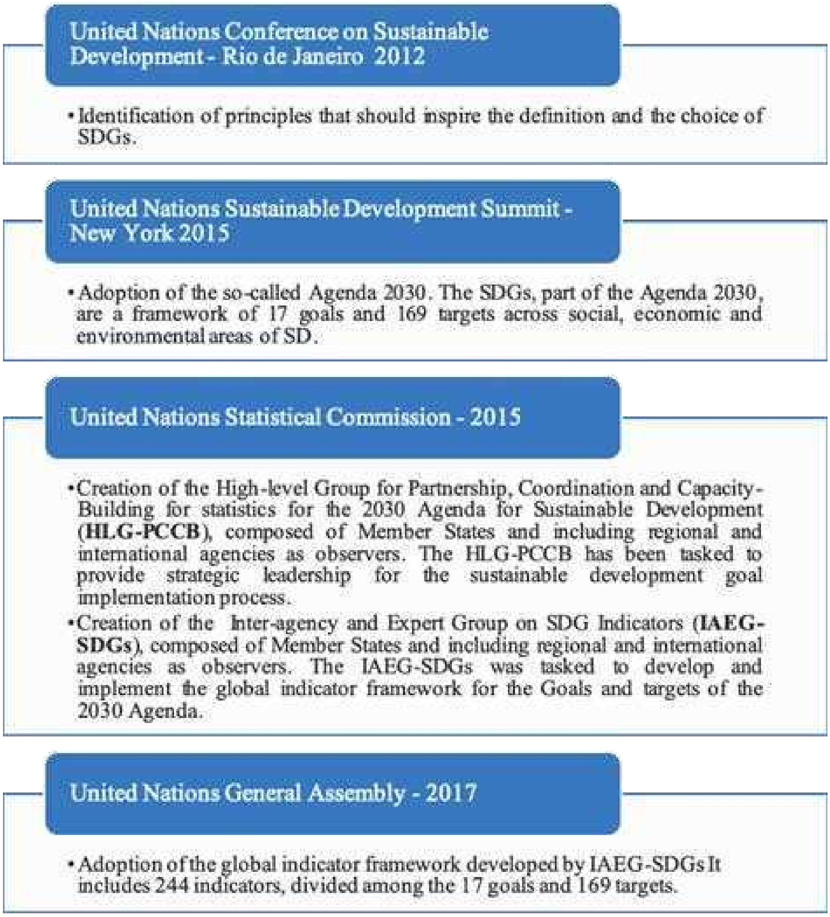


Fig. 5.2. Main phases of sustainable development goals and indicators institutional definition process.

also and especially, the sub-national ones. The realisation of *Agenda 21* at local level remains not legally binding, although by the end of 2000 many countries had policies and frameworks for sustainable development at local and regional levels [37]. It is necessary to *act locally and think globally*. "Sustainability is a policy strategy at the global, national and local levels" [215, 169]. The importance of sub-national and local realities in achieving sustainable development is certainly a central theme of the research in this field [53]. The regionalisation of the indicator framework and the introduction of sustainable development assessment systems at local level can be considered as a cutting-edge research in this field, as demonstrated by different analyses carried out with reference to territories of different countries [79, 252, 127, 149, 200, 128, 4]. The common conclusion of all these studies is that sustainable development cannot be considered independently from the local context. Goals and targets, conceived for all nations, must be adapted to sub-national realities and we must select specific indicators to monitor them.

The necessity of taking into account the sub-national specificities is even more important for Italy, a country historically characterised by strong regional specificities and differences, whose radicalisation is embedded in the so-called North-South gap. Although in the literature there are different positions about the origin of the gap[5], it is generally recognised that the disparity in development, not only the economic one, between the north and the south of the country has been accentuated since unification and throughout Italian history. Title Five of the Constitution of the Italian Republic recognises and regulates sub-national entities. In particular, in this research work we focus on regions, local authorities with legislative power and, therefore, with the authority and the tools to define policies. Following the constitutional reform of 2001, the general legislative power in Italy belongs to the State and the regions, placed on the same level. Based on Art. 117 of the Italian Constitution, legislative competence is allocated by subject. In some matters, the competence is *exclusive* to the State (for example, foreign policy, armed forces, immigration); in others it is *concurrent*, i.e. regions have legislative power, except for the determination of fundamental principles, reserved to the State (for example, health, education, protection and safety at work); finally, the Regions have legislative power with reference to any matter not expressly reserved for the legislation of the State (*residual* or *exclusive* competence). In addition, we must remember that five Italian regions (Sicily, Sardinia, Aosta Valley, Trentino-South Tyrol and Friuli-Venezia Giulia) enjoy particular forms and conditions of autonomy (the so-called *Autonomous regions with special statute*). Therefore, it is clear that we cannot ignore the analysis of the regions' situation and their direct involvement in the definition of strategies and actions.

We examine and monitor the Italian situation with regard to the

---

5 Some authors point out that even at the time of national unification Italy was characterised by evident inequalities and backwardness of the South compared to the rest of the country and how this situation has influenced the future development. Other authors, on the contrary, reduce the importance and role of the initial gap between the different areas in future development. For a review, please see: Amedeo Lepore [163].

achievement of some SDGs, based on the territorial regional analysis, to highlight potential differences or territorial homogeneity. To do this, we construct synthetic measures useful for making the analysis of the phenomena easier, referring to the basic indicators for further details. The choice of considering only some SDGs is linked to the main purpose of this analysis. The crucial point is not so much the analysis of the results, as the highlighting of certain methodological aspects relating to the synthesis of indicators. Moreover, the selection of only some SDGs is consistent with the idea of conceiving them not as a system, but as *alarm bells*. In particular, we focus on two goals, both belonging to the social dimension of sustainable development, according to the three-way holistic framework:

- **Goal 1 –** End Poverty
- **Goal 3 –** Good Health and Well-being

The objective is to construct synthetic indices for each goal considered, one using the aggregative approach and one using the non-aggregative one. In particular, we want to underline how the use of the proposed non-aggregative methodology based on poset not only is particularly appropriate for the longitudinal analysis of cardinal variables, but also allows the overcoming of the *flattening* of phenomena typical of aggregative approaches and discussed in paragraph 3.4.1. To test the validity of the proposed procedure, we compare the results obtained with the two different procedures.

5.2. Description of data

We selected 15 basic indicators, divided among the goals considered, all in time series from 2010 to 2017. The source of the data is the Italian National Institute of Statistics (Istat) data-warehouse (`http://dati.istat.it`). In particular, we have used the last available data from three datasets: *Sustainable Development Indicators* [138], *Equitable and Sustainable Well-being* [137] and *Territorial Indicators for Development Policies* [139]. The selection of the indicators has been influenced by the need to have data available at regional territorial disaggregation level. This means that we could not take into account variables of potential interest (e.g. individuals in absolute poverty) because they did not have data available at regional level. The Brundtland Commission's definition of SD has been the guide for the selection of the basic indicators. As mentioned above, in order to be sustainable, development must ensure the well-being of current generations and, at the same time, not compromise the ability of future generations to achieve it. Therefore, we selected all the indicators useful for the analysis of regional realities and appropriate for either monitoring the *present condition*, or for providing information on a future one (*risk*). We believe that by distinguishing these two aspects it is possible to analyse the phenomena considered. This distinction is partly inspired by the composition of the Human Development Report 2014 [129] and of the Equitable and Sustainable Well-being Report 2015 [136], which described the framework of the sustainability of well-being by using the concept of *risk* and *resilience* factors. We choose the regions as units of analysis to highlight potential territorial differences or historical, geographic and cultural tradition homogeneity in the Italian case.

5.3. Methods

The analysis of each goal starts with the exploratory analysis. The procedure adopted (paragraph 3.2) consists in the calculation of the correlations between (CB) and within (CW) observations and the analysis of the average PCA, for the dimensions (*present condition* and *risk*) of each goal. To construct synthetic measures, we use two different approaches:

- the *composite indicator* approach, the traditional way to synthesize cardinal variables. In particular, we use the *Adjusted Mazziotta-Pareto Index* ($AMPI$), described in paragraph 3.4.1.
- the *non-aggregative* approach. We propose the application of poset, according to the procedure described in paragraph 3.4.3.

Regarding $AMPI$, we use as reference value that of Italy in 2010. The value 100 of $AMPI$ represents the reference value; therefore, $AMPI$ indicates how each unit is placed with respect to the goalposts. In this application, all the composites are positive, i.e. increasing values of each index correspond to positive variations of the phenomenon considered in each goal; then we used $AMPI$ with negative penalty.

The posets have already been used for the treatment of systems of cardinal indicators [106, 9, 8, 12]. The new element in this research work is the application of this methodology to build synthetic measures for the analysis of phenomena over time, according to the procedure described in paragraph 3.4.3. Thus, we construct the *temporal* poset, obtained by merging the posets of each year. We add an embedded scale, to improve the quality of measurement. We have chosen to use a 5-level scale, defined as follows:

- min, with a profile given by the minimum value in all indicators;
- $Q1$, with a profile given by the first quartile of all indicators;
- $Q2$, with a profile given by the second quartile of all indicators;
- $Q3$, with a profile given by the third quartile of all indicators;
- max, with a profile given by the maximum value in all indicators.

We also add the nodes representing Italy. In doing so, we can compare the regional trends respect the national one. Moreover, we have a reference value (the Italian one at year 2010) that we can use to obtain a solution similar to $AMPI$. In particular, we compute the average height of temporal poset and, after, the index numbers (paragraph 3.4.1) by using as reference value that of Italy in 2010.

We want to highlight how the application of poset methodology could allow the overcoming of some problems of the aggregative methods. Synthesis often tends to be representative of situations profoundly different from one another, as the result of different values in the basic indicators, or similar situations between them. We question the effective discrimination power of composite indexes: in fact, data results reveal that the same values in the composite often refer to deeply different situations. This problem can be overcome by applying posets. To test the validity of the proposed procedure, we compare the results obtained with the two different methods, by using the rank correlation on average. The procedure is similar to that of correlation *between* observations, but, instead of the Pearson's coefficient ρ (paragraph 3.2), we calculate the Kendall correlation coefficient τ:

$$\tau = \frac{c-d}{c+d} = \frac{2S}{n(n-1)} \tag{5.1}$$

where c is the number of concordant pairs; d is the number of discordant pairs and $S = (c-d)$ [144]. Kendall's τ measures the degree of a monotone relationship between variables, and like Spearman's ρ, it calculates the dependence between ranked variables; thus, it can be calculated for continuous as well as ordinal data[6]. Finally, we compare the values obtained with the two methods with those of the basic indicators used by the calculation of the correlation between (CB). Before proceeding, we must make a clarification. The construction of composite indicators involves several choices (indicator selection, data normalisation, weights and aggregation methods). The robustness of composite indicators must, therefore, be evaluated. For this purpose, an uncertainty and sensitivity analysis is fundamental [213]. Obviously, these analyses are not sufficient to guarantee the reasonableness of a composite, which must be guided by a robust theoretical framework; however, they could be useful tools for evaluation. These procedures have been designed and applied to allow the comparison between composite indicators; for this reason, they are not considered adequate to evaluate methods belonging to different approaches (aggregative and non-aggregative), as in our case. Hence, our choice to evaluate only the results obtained and the corresponding rankings, pointing to the differences in terms of *interpretation* of results obtained from the two different methods.

All the statistical elaborations and graphs have been carried out using the statistical software **R**. For the calculation of AMPI, we used the *Compind* package [112]. For posets, the *parsec* package [20] was used. We realised the graphic elaborations by using the packages *ggplot2* [270], *igraph* [72] and *factoextra* [143].

5.4. Presentation and analysis of results

We analyse and present the results for each goal in individual subsections. First, we carry out an exploratory analysis; the next step is the analysis and the comparison of the composite values obtained using $AMPI$ and poset. Before the application of poset, we must give all indicators positive polarity; thus, we must to invert those with negative polarity. We used a *linear transformation* to invert polarity (see paragraph 3.4). Given the three-way data array $\mathbf{X} \equiv \{x_{ijt} : i = 1, \dots, N;\ j = 1, \dots, M;\ t = 1, \dots, T\}$, representing a multi-indicator system of a specific goal, we consider all the T matrices $\mathbf{D}_t$ of order $(N \times M)$ (each of them is a *temporal* slice of $\mathbf{X}$):

[6] Kendall's correlation coefficient distinguishes itself from Spearman's one by stronger penalisation of non-sequential (in context of the ranked variables) dislocations.

$$\mathbf{D}_t = \left\{d_{ij} : i = 1...N; j = 1...M\right\} = \begin{Bmatrix} d_{11} & d_{12} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NM} \end{Bmatrix} \tag{5.2}$$

In each matrix $\mathbf{D}$, if an indicator has negative polarity we invert it as follows: $d'_{ij} = (max_{x_j} - d_{ij})$, where max_{x_j} is the maximum of the indicator j.

5.4.1. Goal 1 - End poverty. Present condition

For monitoring the present condition in the Italian regions regarding poverty (Goal 1), we select 4 basic indicators:

- the severe-material deprivation rate ($X1$);
- the share of total population living in a dwelling with a structural problem ($X2$);
- the housing-cost overburden rate ($X3$);
- the regional poverty index ($X4$).

Table 5.1 reports the definitions of the indicators and their polarity. Figure 5.3 reports the results of exploratory analysis. It shows a weak CW, which indicates that an increase over time in one indicator within one observation is not associated with an increase in the other indicators. This means that regions have different trends in each variable considered. For instance, from 2009 to 2017 the severe-material deprivation rate increases in Piedmont (from 5.7% to 9.0%), while the housing-cost overburden rate decreases (from 8.7% to 7.2%). Even the CB presents values quite low, with the exception of the relation between the severe-material deprivation rate and the regional poverty index (0.87). Looking at these results, we can deduce that the indicators selected represent different aspects of poverty conditions. The first principal component explains 56% of total variance; therefore, we can consider only one latent variable and construct only one composite. The choice of the 4 indicators allows the consideration of different aspects of that latent variable.

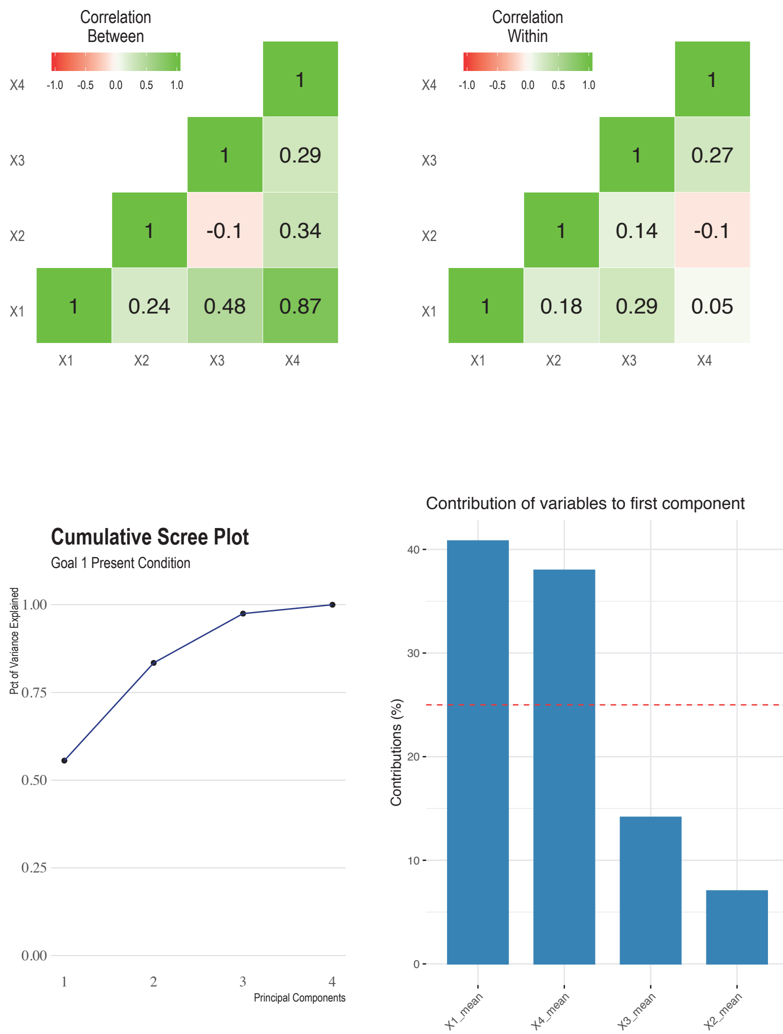
From 2004 to 2010, Italy shows a positive trend in all indicators: for instance, the regional poverty index decreases from 11% to 9.6%; the

Tab. 5.1. Indicators of Goal 1 - End poverty. Present condition: code; description; polarity.

Code	Basic Indicator	Description	Polarity
X1	Severe-material deprivation rate	Share of population living in households lacking at least 4 items out of the following 9 items: i) to pay rent or utility bills, ii) keep home adequately warm, iii) face unexpected expenses (of 800 euros in 2014), iv) eat meat, fish or a protein equivalent every second day, v) a week holiday away from home, or could not afford) vi) a car, vii) a washing machine, viii) a colour TV, or ix) a telephone.	NEG
X2	Share of total population living in a dwelling with a leaking roof, damp walls, floors or foundation, or rot in window frames of floor	Percentage of persons in the total population living in a dwelling with at least one of the following housing problem: a) Leaking roof or rot in window frames or floor b) damp walls/ floors/ foundation.	NEG
X3	Housing-cost overburden rate	Percentage of persons in the population living in households where the total housing costs represent more than 40% of disposable income.	NEG
X4	Regional poverty index (households)	The estimate of the incidence of relative poverty (the percentage of poor households and persons) is calculated on the basis of a conventional threshold (known as the International Standard of Poverty Line), which identifies the value of consumption expenditure below which a household is defined as poor in relative terms. The relative poverty threshold for a two-member household is equal to the average monthly expenditure per person in the country, which in 2015 was 1050.95 euro. Households consisting of two persons with a monthly expenditure equal to or less than this value are classified as poor. For households of different sizes, the line value is obtained by applying an appropriate equivalence scale.	NEG

housing-cost overburden rate passes from 12.3% to 7.7%. Thus, even if the Italian situation in 2010 is not the best performing, it is still the result of progressive improvements. The year 2010 marks the end of these improvements and the beginning of a negative trend, which we can consider a structural feature of Italy, characterizing the country up

Fig. 5.3. Exploratory analysis of basic indicators regarding Goal 1 present condition: correlation between observations; correlation within observations; cumulative scree plot of PCA on average data.



to today. All selected indicators suffer a marked worsening from 2010 to 2016: for example, the severe deprivation index goes from 7.4% to 12.1% (reaching the minimum value, 14.5%, in 2012). In 2017, the trend is reversed, with all indicators improving, with the exception of the regional poverty index, which reaches 13.5% (increasing steadily since 2010).

5.4.1.1. Synthesis by using AMPI

The trend described above is well reproduced by the national composite (Figure 5.4), which loses 7 points from 2010 to 2016, while in 2017 it shows a clear improvement, passing from 93 to 97. Regional composites confirm the national trend, with a general worsening from 2010 to 2016 and an improvement in 2017, except for Campania and Sardinia. In 2017, Campania confirms the same value as 2016 (67). But this situation is not the result of a stability in the basic indicators: in fact, the severe-material deprivation index loses 6 percentage points; while the regional poverty index gains 5 (the other 2 indicators remain constant). The composition and, consequently, the equilibrium among indicators change.

Sardinia has a lower value in 2017 (89) than in 2016 (93) due to a worsening in the regional poverty index and the housing-cost overburden rate. We observe a split in the country. The Northern Regions have values higher than the national ones throughout the period considered (except Liguria; from 2012 to 2015, which presents lower values than national data, while in 2016 and 2017, it remains in line with the Italian trend). Among Central Regions, Tuscany has a better trend than Italy; Marche, Umbria and Lazio show different trends over time, but in 2017, they all stand at higher values than the Italian ones. All the Southern Regions are below the national figure. The North-South gap is evident in Figure 5.4.

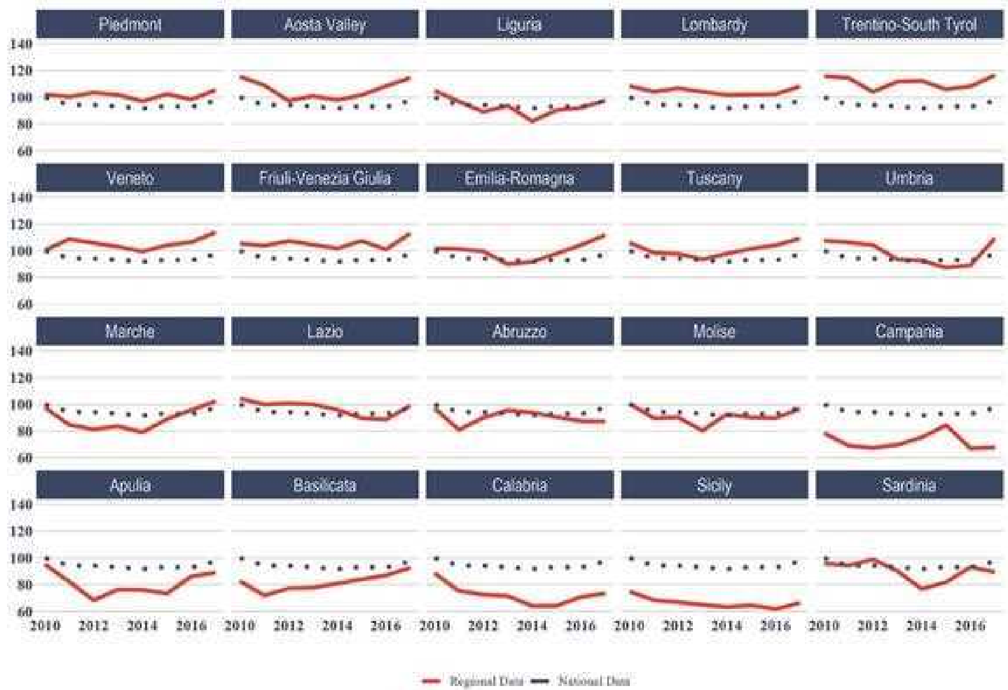


Fig. 5.4. Synthetic index of Goal 1 present condition: regional and national data; time series 2010-2017; AMPI: Italy 2010 = 100.

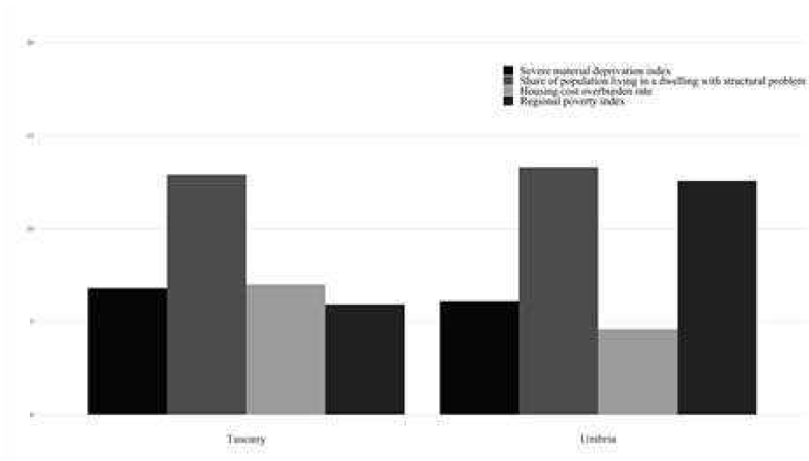


Fig. 5.5. Comparison between basic indicators of Goal 1 present condition: Tuscany and Umbria in 2017.

Different regions often have the same value in the composite, even though they have different compositions in the basic indicators. For instance, in 2017, the values for Tuscany and Umbria, although the same (109), are the result of different compositions in the basic indicators. As shown in Figure 5.5, Tuscany has worse values than Umbria in the severe-material deprivation rate (7% compared to 6% of Umbria) and the housing-cost overburden rate (7% compared to 4.5% of Umbria); it has much better values than Umbria in the regional poverty rate (6% compared to 13% of Umbria). On the contrary, Veneto and Friuli Venezia Giulia present the same composite (109), as result of similar compositions in basic indicators.

5.4.1.2. Synthesis by using poset

Figure 5.6 shows the individual Hasse diagrams for each year considered. The nodes represent the profiles of the regions in the indicators considered. Figure shows how the relationship structure has changed over time. However, as we have explained in the paragraph 3.4.3, in order to make inter-temporal comparisons between regions it is necessary to merge the posets and form a single *temporal* poset. At the same time, to improve the quality of the measurement obtained, we must introduce an embedded scale. We use a scale defined as illustrated in the paragraph 5.3. In Figure 5.7, we report the Hasse diagram of the temporal poset, obtained by merging the single-year posets (Figure 5.6).

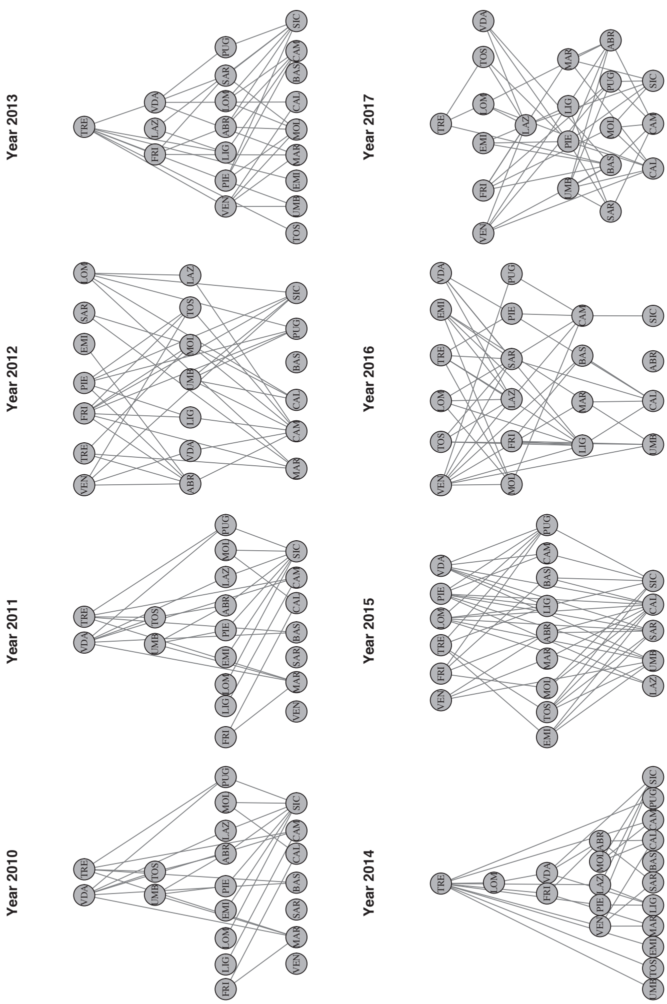


Fig. 5.6. Goal 1 present condition: Hasse diagrams from 2010 to 2017.

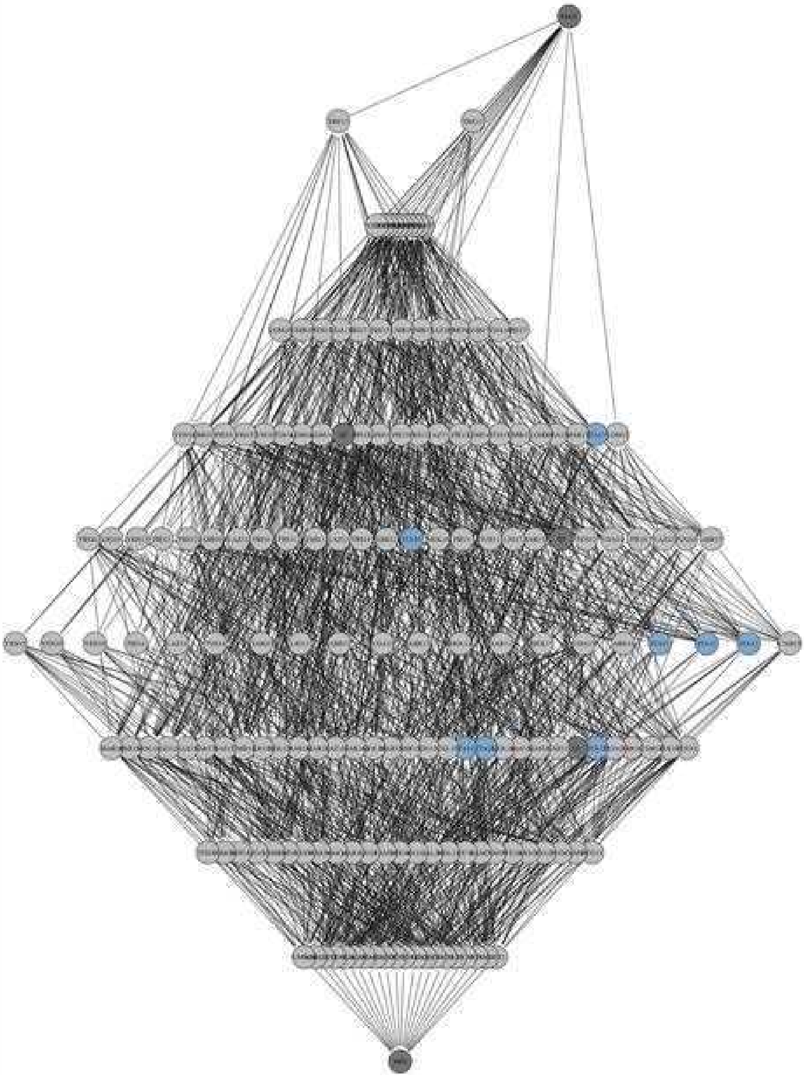


Fig. 5.7. Goal 1 present condition: temporal poset years 2010-2017.

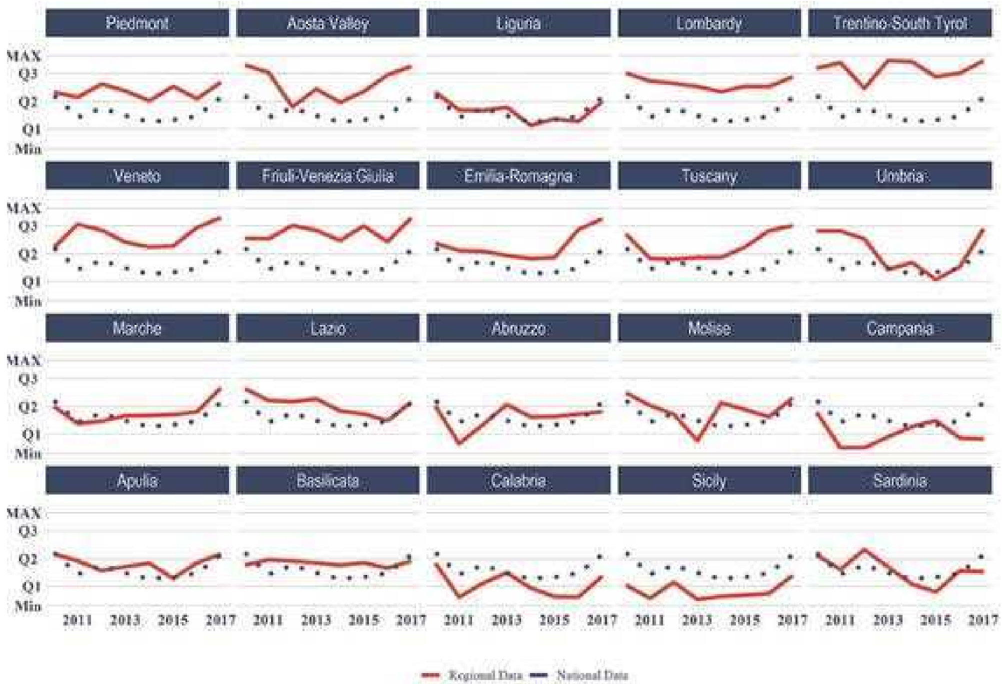


Fig. 5.8. Synthetic index of Goal 1 present condition: regional and national data; time series 2010-2017; index numbers from average height distribution; fixed base = Italy 2010.

We add the embedded scale (nodes in dark grey) and the nodes representing Italy (blue nodes in the Hasse diagram). In doing this, we can compare the regional trends respect the national one. We calculate the average height of temporal poset. Moreover, we use the Italian value at year 2010 as reference to compute the index numbers at fixed base. In Figure 5.8, we present the charts with the time series of the index numbers; the value 1 corresponds to the value of Italy in 2010. We have used the embedded scale as reference system of charts.

The Italian trend is very similar to that showed by $AMPI$. The national value decreases from 2010 to 2016, while in 2017 it shows a clear improvement. Regional composites confirm the national trend, with a general worsening from 2010 to 2016 and an improvement in 2017. As in the case of AMPI (Figure 5.4), the exception are Campania and Sardinia (we have already analysed the possible causes of this situation). The North-South gap is clear and evident. The two syntheses has a very high rank correlation ($\tau = 0.83$) rank on average; thus we can conclude that they report more or less the same results.

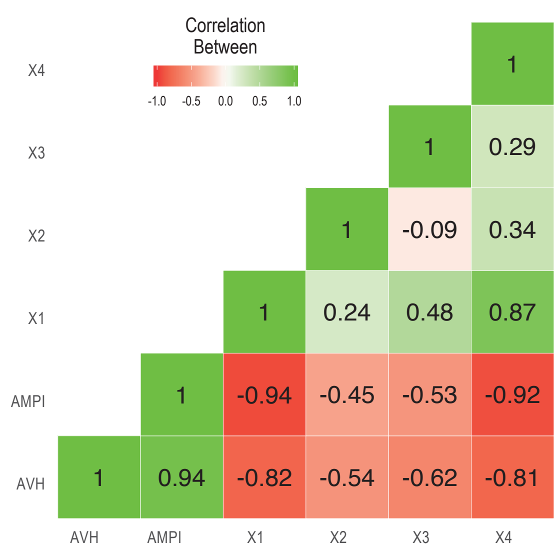


Fig. 5.9. Goal 1 present condition: correlation between (CB) average height (AVH), AMPI and basic indicators.

Both synthetic measures are quite correlated with the basic indicators

(Figure 5.9); we can observe lower correlation values with reference to $X2$ (the share of total population living in a dwelling with a structural problem) and $X3$ (the housing-cost overburden rate). As underlined in the exploratory analysis, these two indicators seem to represent a different aspect of poverty condition.

Differently from what was seen with the application of the aggregative method, in this case the different combinations in the basic indicators are differentiated by the attribution of different values. In the example analysed before, we have seen that the combinations in basic indicators of Umbria and Tuscany are different in 2017 (Figure 5.5). By using $AMPI$, the two regions have the same value. With the application of the poset procedure, Tuscany has an average height of 140 (the index number at fixed base Italy 2010 is 1.42), while Umbria presents a value of 134 (the index number at fixed base Italy 2010 is 1.35).

5.4.2. Goal 1 - End poverty. Risk

For the analysis of the risk of poverty, we consider 3 basic indicators:

- the low-work intensity rate ($X5$);
- the at-risk-of-poverty rate ($X6$);
- the economic distress index ($X7$).

Table 5.2 reports the definitions of the indicators and their polarity.

Figure 5.10 reports the results of the exploratory analysis. We observe a strong CB for all indicators, meaning that they present a strong association on average and, consequently, high values in one indicator correspond to high values in the others. However, the CW shows different results. In fact, only the low-work intensity rate and the at-risk-of-poverty rate have an appreciable coefficient (0.41). The CW between the low-work intensity rate and the economic distress index is even negative (-0.21). This result, which could be considered an error if read superficially, provides important information. It indicates that some regions tend to have divergent trends in these indicators. For instance, from 2010 to 2017, Calabria shows an enhancement in the economic distress index (it passes from 24.4% to 12.5%) and an aggravation in the low intensity rate (it increases from 17.5% to 22.4%). We observe similar situations in other regions (e.g. Friuli-Venezia Giulia, Molise, etc.). The conclusion is that the indicators have very similar trends on average, while very different ones from one region to another.

Tab. 5.2. Indicators of Goal 1 - End poverty. Risk: code; description; polarity.

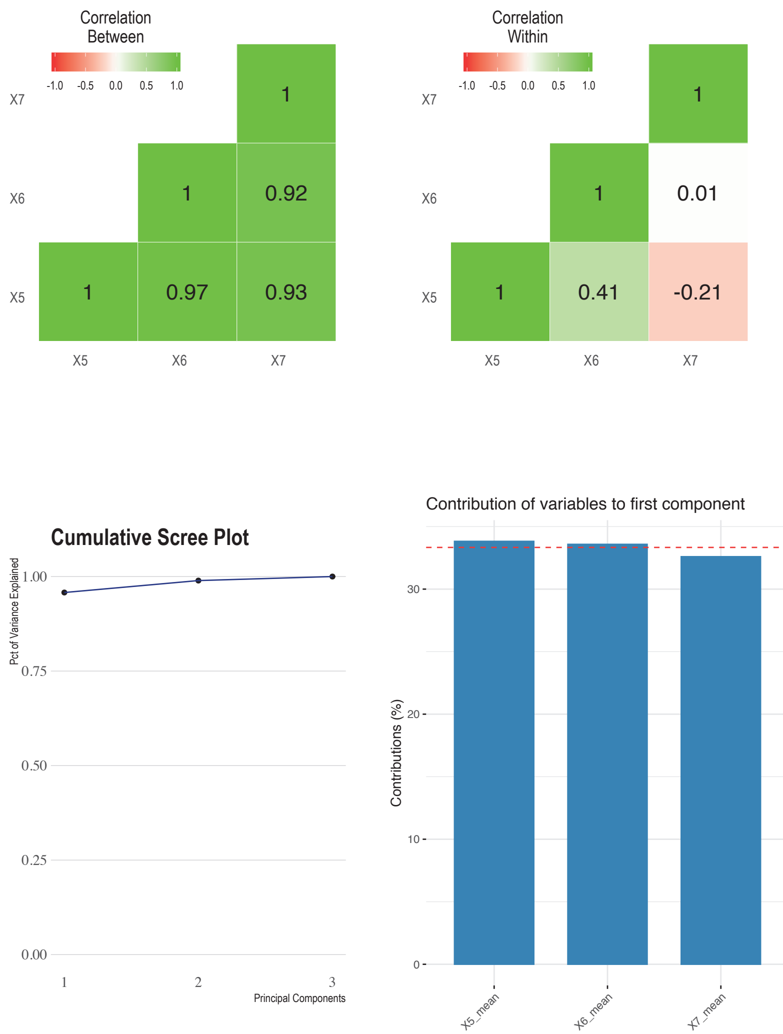
Code	Basic Indicator	Description	Polarity
X5	Low-work intensity rate	Proportion of people living in households with very low work intensity namely household members of working age (person aged 18-59 years, with the exclusion of dependent children aged 18-24) that have worked during the income reference year less than 20% of the number of months that could theoretically have been worked by the same household members.	NEG
X6	At-risk-of-poverty rate	Percentage of persons at risk of poverty, with an equivalised income less than or equal to 60% of the median equivalised income.	NEG
X7	Index of economic distress	Share of individuals in households that, considering all the available income, declare to get to the end of the month with great difficulty.	NEG

The correlation analysis seems to confirm that the indicators considered represent two different aspects of the risk of poverty, the objective and the subjective one. From the cumulative scree plot, it is quite evident that we can consider only one latent variable (the first component explains more than 90% of total variance). The contribution of the indicators to the first component is high.

5.4.2.1. Synthesis by using AMPI

The risk of poverty indicators in Italy improves between 2004 and 2010. However, this improvement only concerns the objective dimension. In fact, the economic distress index (i.e. the share of individuals in households that, considering all the available income, declare to get to the end of the month with great difficulty) increases from 15% to 17%. From 2010 on-wards, the Italian trend has been reversed. In fact, both the two indicators expressing the objective dimension worsen (from 2010 to 2017, the at-risk-of-poverty rate increases by 1.6% and the low-work intensity rate by 1.2%). On the contrary, the economic distress index improves from 17.4% in 2010 to 8.6% in 2017. The Italian composite (Figure 5.11) constantly decreases until 2014. From 2015, it begins to rise until reaching the value 105 in 2017.

Fig. 5.10. Exploratory analysis of basic indicators regarding Goal 1 risk: correlation between observations; correlation within observations; cumulative scree plot of PCA on average data.



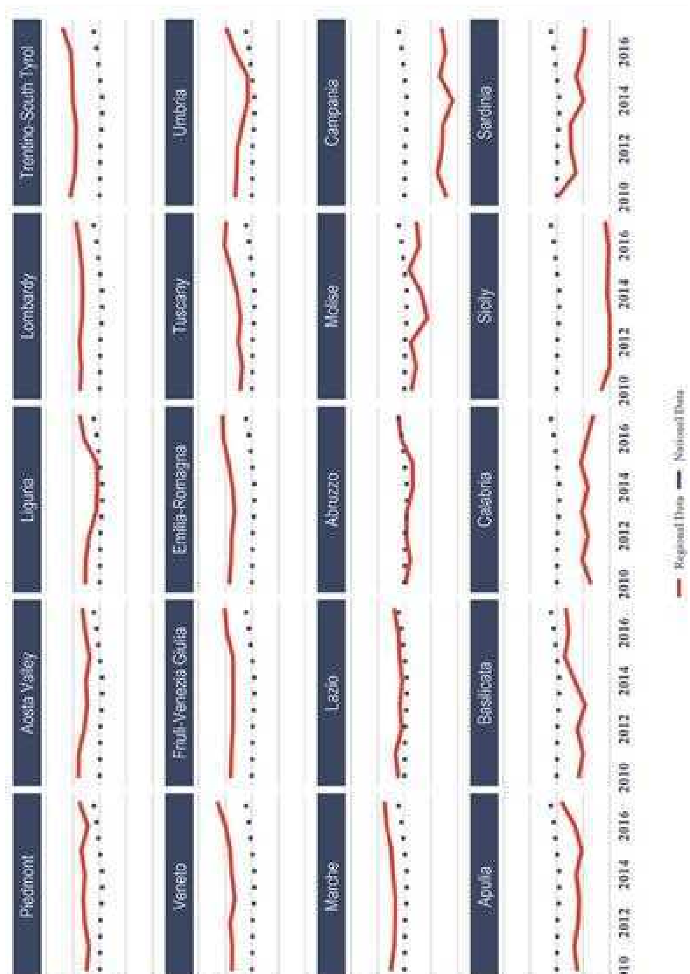


Fig. 5.11. Synthetic index of Goal 1 risk: regional and national data; time series 2010-2017; AMPI: Italy 2010 = 100.

This trend is mainly influenced by that of the economic distress index. Looking at regional data, the North-South gap is evident. All Northern and Central Regions have better trends than the national one, with values always higher than the Italian ones. Southern Regions (except Abruzzo, in line with Italy) present values and trends well below national data, in some cases highlighting distances that seem difficult to cover (for instance, Sicily, Calabria and Campania).

The results, considered together with those of the present condition (paragraph 5.4.1), show an alarming picture for the South of the country; not only there is a situation of very strong manifest poverty, but there is also an increased risk that the situation will worsen over time.

The same composite' value does not always represent the same situations. For example, despite having the same value in 2017 (115), Piedmont and Liguria present different combination in basic indicators (Figure 5.12). Piedmont has a lower low-work intensity rate (7.5% and Liguria 9.7%), while Liguria present a better value in the economic distress index (5.3% and Piedmont 8.9%).

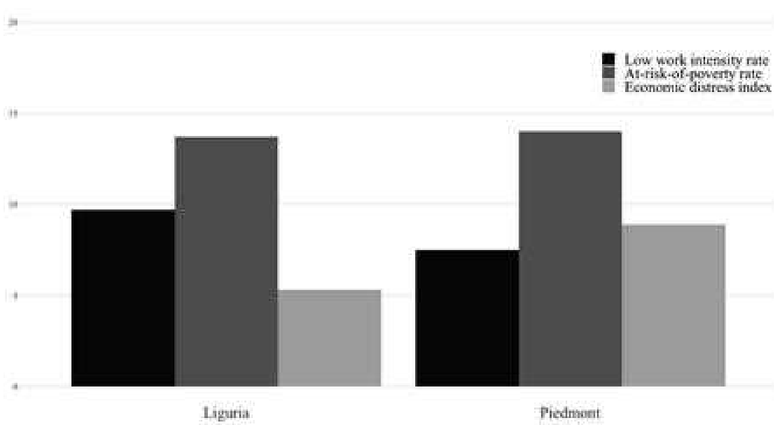


Fig. 5.12. Comparison between basic indicators of Goal 1 risk: Liguria and Piedmont in 2017.

5.4.2.2. Synthesis by using poset

In Figure 5.13, we report the Hasse diagrams for each year considered. The nodes represent the profiles of the regions in the three indicators considered for the risk of poverty.

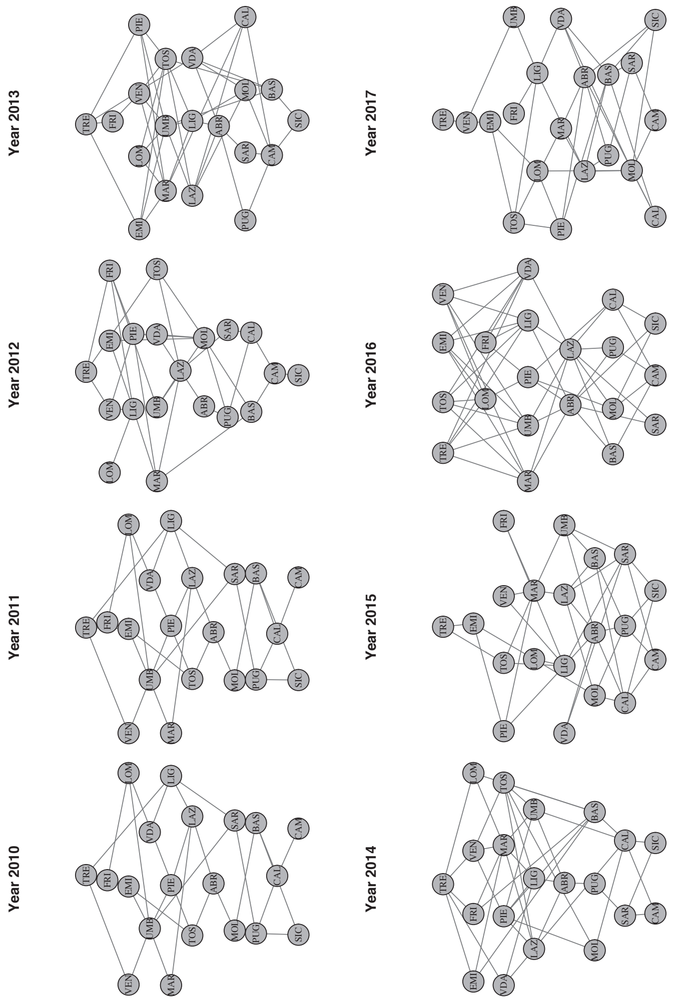


Fig. 5.13. Goal 1 risk: Hasse diagrams from 2010 to 2017.

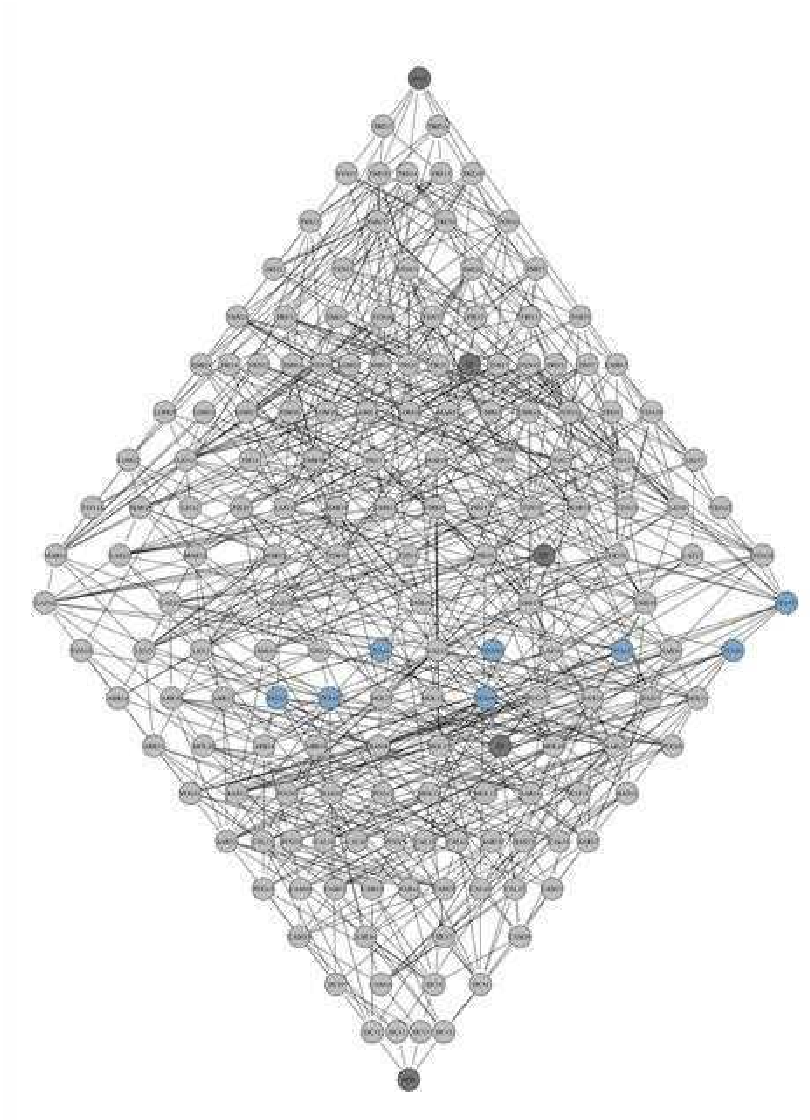


Fig. 5.14. Goal 1 risk: temporal poset years 2010-2017.

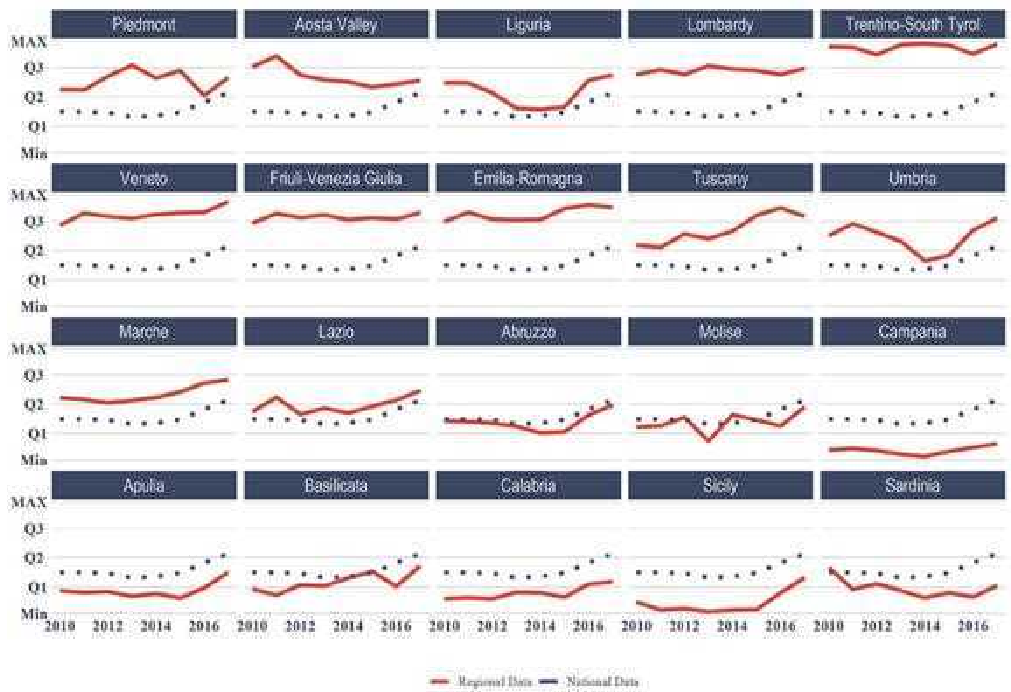


Fig. 5.15. Synthetic index of Goal 1 risk: regional and national data; time series 2010-2017; index numbers from average height distribution; fixed base = Italy 2010.

We can observe that the relationship position of regions has changed over time. For instance, Trentino-South Tyrol is on the top of different Hasse diagrams in all years; the position of other regions (for instance, Tuscany) changes over time. However, we cannot know anything about the temporal changes; as we have explained in the paragraph 3.4.3, in order to make inter-temporal comparisons, we must analyse the *temporal* poset. We merge the single-year posets and add the embedded scale, defined according to the rules presented in paragraph 5.3. In Figure 5.14, we report the Hasse diagram of the temporal poset, obtained by merging the single-year posets (Figure 5.13). We add the embedded scale (nodes in dark grey) and the nodes representing Italy (blue nodes in the Hasse diagram).

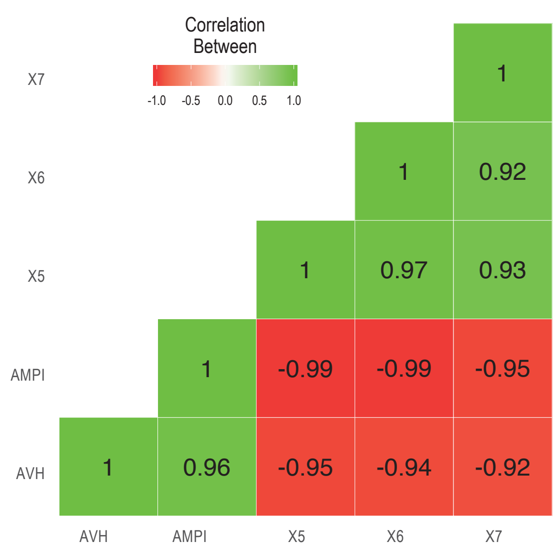


Fig. 5.16. Goal 1 risk: correlation between (CB) average height (AVH), AMPI and basic indicators.

We calculate the average height of temporal poset and compute the index numbers at fixed base (Italy 2010). Figure 5.15 shows the time series of the index numbers obtained from the average height; the value 1 corresponds to the value of Italy in 2010. As in Figure 5.8, the embedded scale is used as reference system of charts. The Italian trend is similar to that showed by $AMPI$. The national value decreases from 2010 to

2014, starting from 2015, it begin to rise. This increase is due to the net improvement in the economic distress index. Regional composites confirm the gap between the North and the South of the country.

The two syntheses has a very high average rank correlation ($\tau = 0.99$) and are both very highly correlated with the basic indicators (Figure 5.16). Thus, we can conclude that they report more or less the same results. As in the case of Goal 1 present condition, the application of poset allow the differentiation of units, by attributing different synthetic values to different combinations. Piedmont and Liguria, having the same $AMPI$, with the application of poset present different values: Piedmont has an average height of 116 (the index number at fixed base Italy 2010 is 1.82), Liguria has an average height of 120 (the index number at fixed base Italy 2010 is 1.88).

5.4.3. Goal 3 - Health and well-being. Present condition

We use 4 basic indicators to monitor the present condition regarding Goal 3:

- the life expectancy at birth ($X8$) - LE;
- the healthy life expectancy at birth ($X9$) HLE;
- the life expectancy without activity limitations at 65 years of age ($X10$) - LEL;
- the good health index ($X11$).

We want to consider both an objective dimension and a self-perceived subjective dimension. Table 5.3 reports the definitions of the indicators and their polarity.

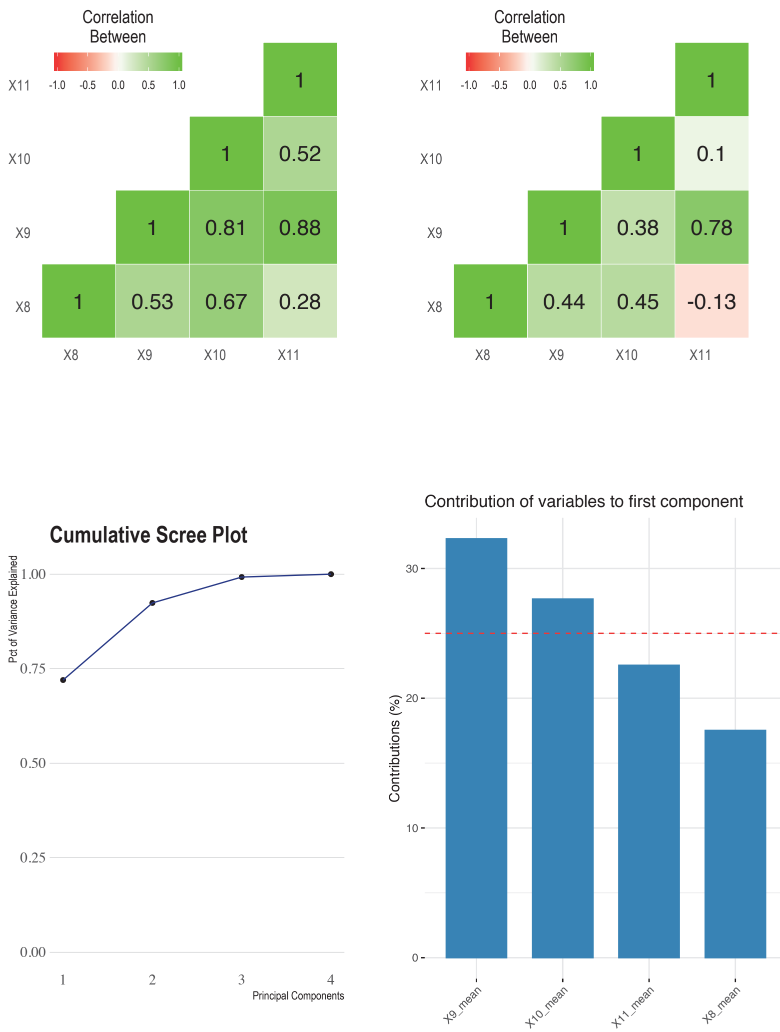
The results of the exploratory analysis (Figure 5.17) are interesting. We would be inclined to expect a high correlation between the three life expectancies and that the latter have an influence on the self-perceived health. Starting with the results of CB, LEL is strongly correlated to the other two. The correlation between LE and HLE is positive, but the coefficient is not as high as expected (0.53). This indicates a difference between the two indicators, expressed in terms of distance between their values. The CW among these three indicators confirms the direction of the links (all the correlations are positive), but the intensity is lower than in the CB. In brief, we can conclude that high values on average in an indicator correspond to high values in the other two and vice versa; at

Tab. 5.3. Goal 3 - Health and well-being. Present condition: code; description; polarity.

Code	Basic Indicator	Description	Polarity
X8	Life expectancy at birth	Average number of years that a child born in a certain calendar year can expect to live.	POS
X9	Healthy life expectancy at birth	Average number of years that a child born in a given calendar year can expect to live in good health, using the prevalence of individuals who respond positively ("well" or "very well") to the question on perceived health.	POS
X10	Life expectancy without activity limitations at 65 years of age	Average number of years that a person aged 65 can expect to live without restrictions on activities due to health problems, using the proportion of people who have responded that they have had restrictions, for at least 6 months, due to health problems in performing the activities that people habitually perform.	POS
X11	Good health index	Proportion of people claiming to be in good health.	POS

the same time, increases in an indicator within one observation tend to be associated with increases in the other indicators, even if some regions show different trends. The good health index has a strong CB with LEL (0.52) and, in particular, with HLE (0.88). We note that this relationship continues to be very strong (0.78) even in the case of CW. On the contrary, the CW between the good health index and the LEL is even negative (-0.13). As already highlighted (paragraph 5.4.2), this result is not a mistake, but indicates that Regions have diverging trends in these indicators. In particular, from 2010 to 2017 some regions (Piedmont, Veneto, Tuscany, Umbria, and Apulia) register an increase in LEL and a decrease in the good health index. For instance, in Veneto the first indicator passes from 8.6 to 10.3 and the second one from 72% to 70%. From these results, we conclude that, despite of the correlations between indicators are strong on average, the trends are often different from one region to another. This outcome allows us to consider these indicators as different aspects of the phenomenon considered. This seems to be confirmed by the analysis of contributions of basic indicators to the first component. The cumulative scree plot shows that the first principal component explains almost 75% of total variance and, consequently, we can construct only one composite.

Fig. 5.17. Exploratory analysis of basic indicators regarding Goal 3 present condition: correlation between observations; correlation within observations; cumulative scree plot of PCA on average data.



5.4.3.1. Synthesis by using AMPI

Due to the lack of data prior to 2009, we can only provide the *picture* of Italy for 2010 and analyse the following trend. In 2010, LE in Italy is rather high (approximately 82 years); the HLE is about 58 years of age, while LEL is 9 years; almost 71% of population claim to be in good health. In 2017, all indicators are growing compared to 2010, except for the good health index, which loses 1 percentage point. The distance between the years of age between LE (83) and HLE (59) remains unchanged. We can consider the distance between LE and HLE as a structural feature of the country, which can be simplified in the formula *long life, short health*: Italians live long, but spend 30% of their lives in a poor condition of health. The trend of the Italian composite (Figure 5.18) is constantly growing from 2009 to 2016 (109). In 2017, the composite loses two points, due to the decrease in all life expectancies. The decrease concerns all three indicators: the life expectancy at birth passes from 82.8 to 82.7; the healthy life expectancy at birth from 58.8 to 58.7 and the life expectancy without activity limitations at 65 years goes from 9.8 to 9.7. These decreases are very slight. However, we must consider that life expectancies vary very slowly over time and, therefore, could have a very strong weight, especially when compared to a positive trend in previous years.

Looking at regional trends, we can see that Trentino-South Tyrol presents the best trend, while Calabria the worst. In fact, Trentino-South Tyrol has the best values in all indicators throughout the period considered. For example, the index of good health is 81%, while for the country it is 71% (we consider the average data for simplicity).

At the opposite, Calabria has the worst values in almost all indicators, with the exception of LE, where Campania is the worst. There are significant differences among the Italian regions, with the Northern and Central ones having better conditions. These differences are mainly due to differences in HLE: for instance, the distance between Calabria and Trentino-South Tyrol in this indicator is on average 15 points.

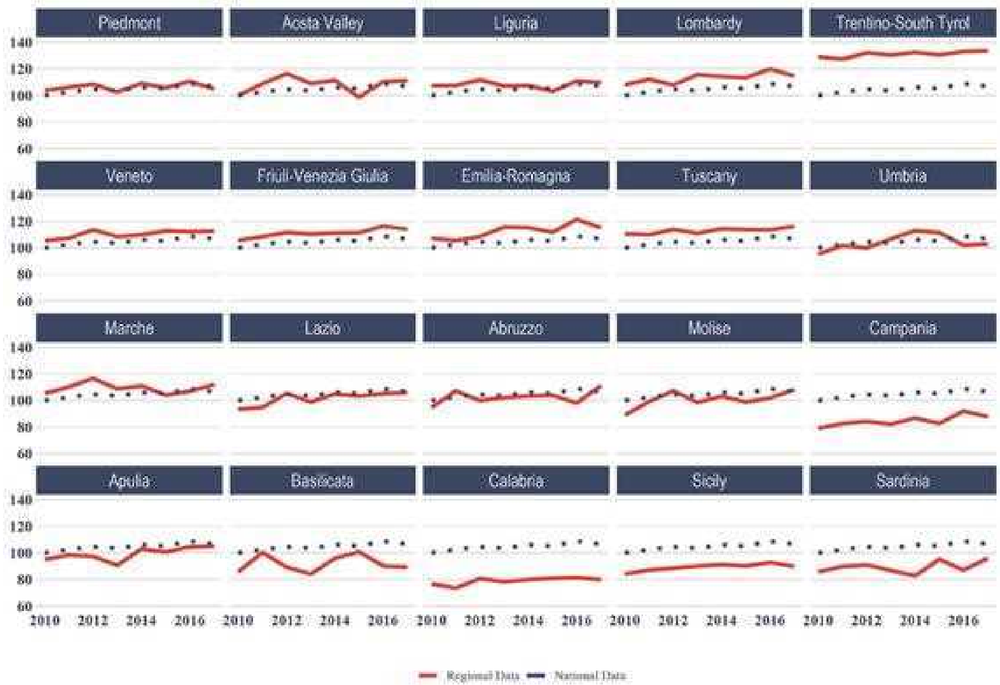


Fig. 5.18. Synthetic index of Goal 3 present condition: regional and national data; time series 2010-2017; AMPI: Italy 2010 = 100.

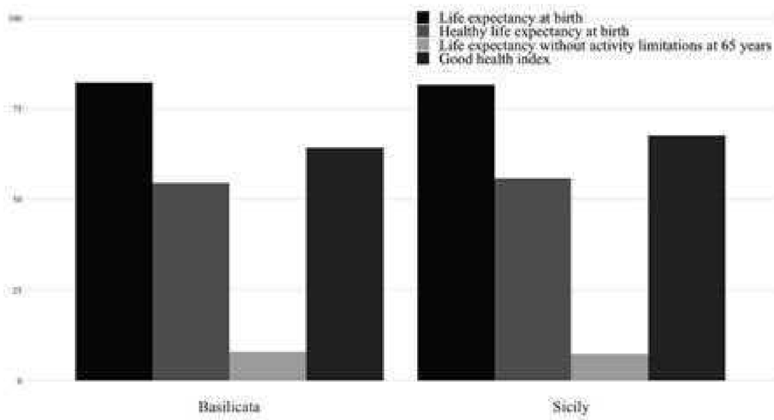


Fig. 5.19. Comparison between basic indicators of Goal 3 present condition: Basilicata and Sicily in 2017.

As seen in previous analyses, the same value of composite often represents very different situations. Basilicata and Sicily, despite having the same value in 2017 (89), have very different combinations in the basic indicators, as shown in Figure 5.19 (Sicily has better values in HLE and good health index, whilst worse in the others).

5.4.3.2. Synthesis by using poset

In Figure 5.20, the Hasse diagrams of different years are showed. It should be noted that, although they change over time, the relationships structures have some elements common to all years. For example, Trentino-South Tyrol is always the highest node in the various diagrams, except in 2016, where there are more regions at the top of the Hasse diagram. As said, we cannot know if this phenomenon is the result of a worsening of Trentino-South Tyrol or of an improvement of the other regions or of both these circumstances. Sicily, Campania and Calabria are always minimal elements of the different posets[7]. Figure 5.21 shows the temporal poset obtained by merging those in Figure 5.20 and adding the nodes representing Italy (blue) and the embedded scale (dark grey).

[7] In the 2012 Hasse diagram, the three nodes appear superimposed. However, the three regions have different combinations in the elementary indicators. The graph superimposes them for space economy, since they have the same comparabilities.

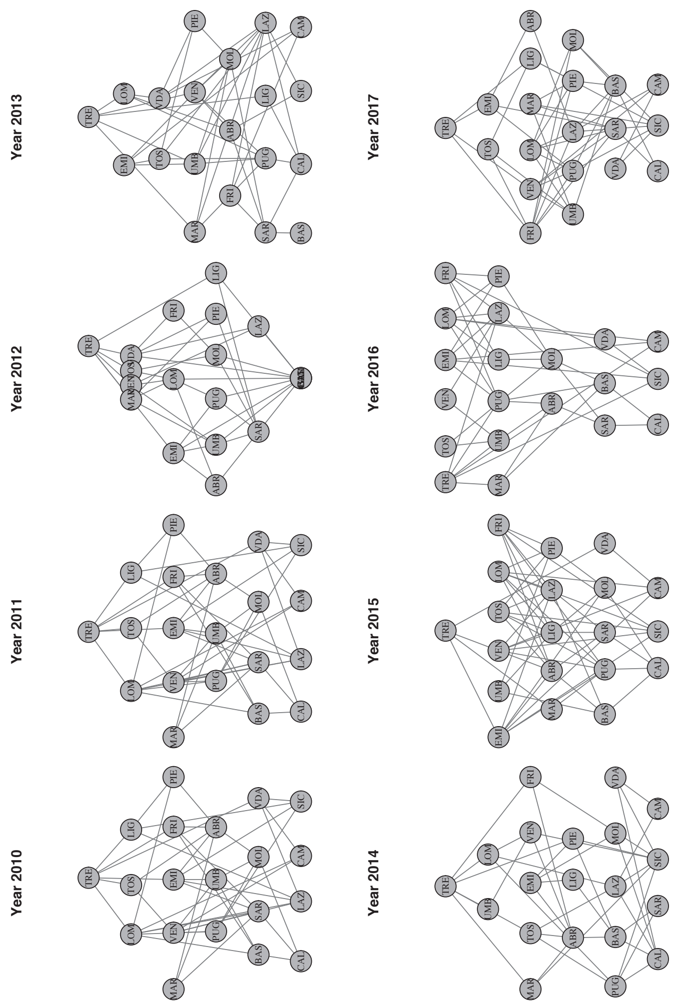


Fig. 5.20. Goal 3 present condition: Hasse diagrams from 2010 to 2017.

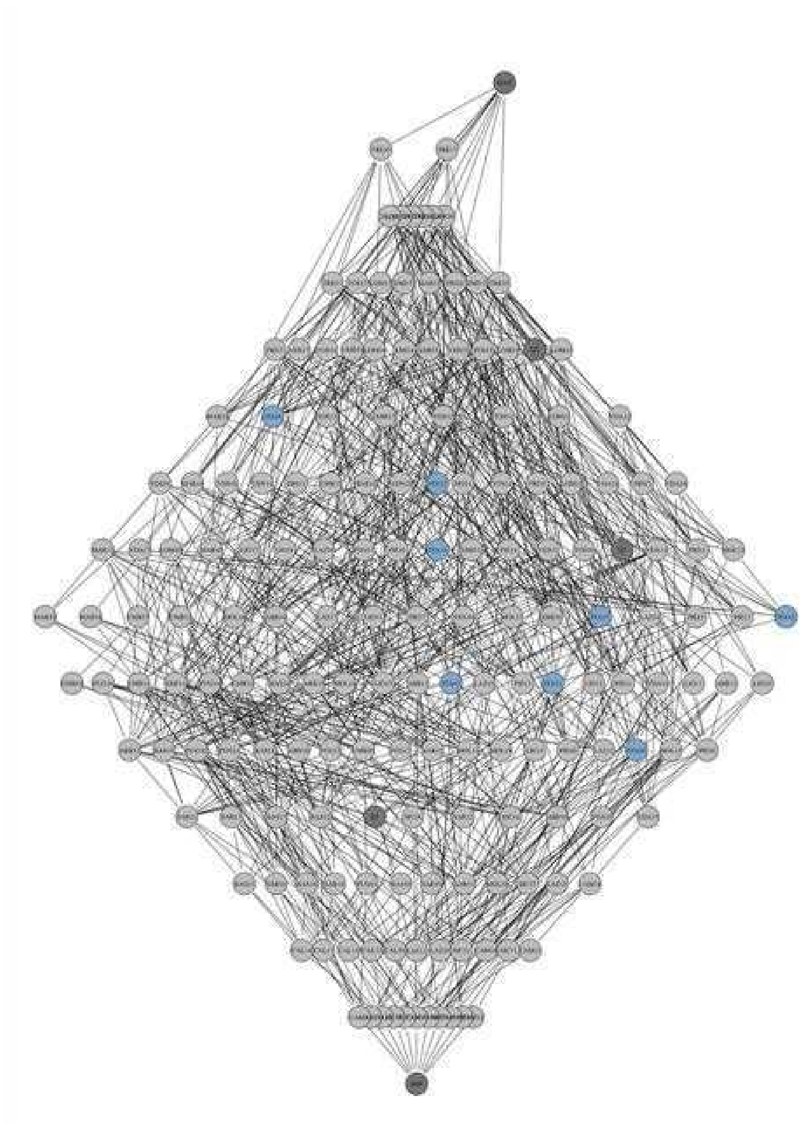


Fig. 5.21. Goal 3 present condition: temporal poset years 2010-2017.

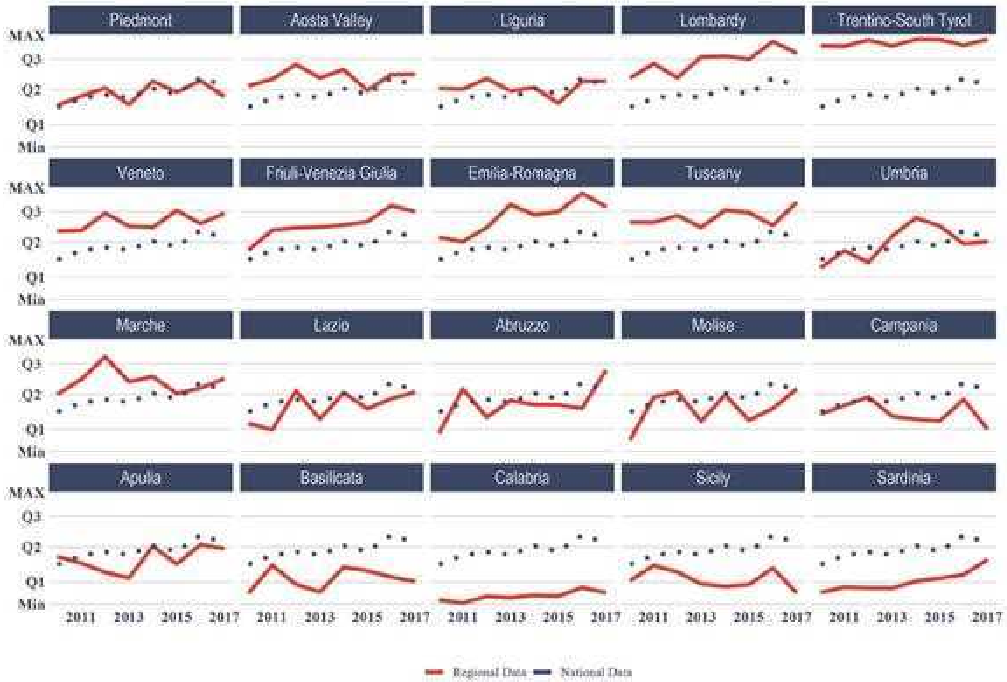


Fig. 5.22. Synthetic index of Goal 3 present condition: regional and national data; time series 2010-2017; index numbers from average height distribution; fixed base = Italy 2010.

We compute the average height and the index numbers using the Italian value in 2010 as fixed base. Figure 5.22 reports the charts with regional and national trends. As in the previous analyses, also in this case the national trend is almost the same as that obtained by applying $AMPI$ (obviously the scale is different); in fact, it is constantly growing from 2009 to 2016, while in 2017 it decreases (for comparing this results with $AMPI$ trend, see Figure 5.18). We have already explained the reasons of this trend in paragraph 5.4.3.1. Trentino-South Tyrol and Calabria are, respectively, the *best* and the *worst performer*; the result is the same obtained by the aggregative-compensative method. The North-South gap is marked.

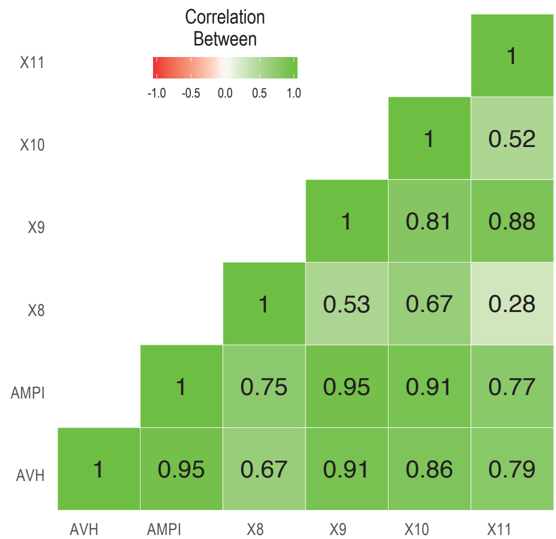


Fig. 5.23. Goal 3 present condition: correlation between (CB) average height (AVH), AMPI and basic indicators.

The two procedures lead to almost the same results; the rank correlation on average between the two measures is very high ($\tau = 0.92$). Looking at the correlation between the two syntheses and the basic indicators (Figure 5.23), we can observe high coefficients. As we have seen in previous analyses, the use of posets seems to be able to discriminate better among statistical units respect to their profiles. This also happens in this case. According to different combinations in basic indicators, Basilicata has an average height of 35.9 (the index number at fixed base Italy 2010 is 0.58), while Sicily has a value of 18.34 (the index number at fixed base Italy 2010 is 0.30).

5.4.4. Goal 3 - Health and well-being. Risk

In order to monitor risk, we select 4 basic indicators, expressing behaviours that may influence the level of health:

- smoking ($X12$);
- alcohol consumption ($X13$);
- adequate nutrition ($X14$);
- sedentariness ($X15$).

Table 5.4 reports the definitions of the indicators and their polarity.

Tab. 5.4. Indicators of Goal 3 - Health and well-being. Risk: code; description; polarity.

Code	Basic Indicator	Description	Polarity
X12	Smoking	Proportion of people aged 14 and over who report current smoking. The indicator is standardized using the Italian 2001 Census population as standard population.	NEG
X13	Alcohol consumption	Proportion of people aged 14 and over who have at least one behaviour at risk in the consumption of alcohol. The indicator is standardized using the Italian 2001 Census population as standard population.	NEG
X14	Adequate nutrition index	Percentage of people aged 3 years and over who say they take every day at least 4 portions of fruit and vegetables. The indicator is standardized using the Italian 2001 Census population as standard population.	POS
X15	Sedentariness index	Proportion of people aged 14 and over referring not to perform any physical activity.	NEG

Obviously, other behaviours may affect health; however, we cannot include them in our analysis because of lack of data. Similar considerations to those made in the previous pages emerge from the results of the exploratory analysis (Figure 5.24).

Fig. 5.24. Exploratory analysis of basic indicators regarding Goal 3 risk: correlation between observations; correlation within observations; cumulative scree plot of PCA on average data.

We observe that high CB values (both positive and negative) correspond to low CW values. These indicate that, as previously written, the indicators have some correlation on average and, at the same time, Regions often present different, and even divergent, trends among basic indicators. The first principal component explains more than 50% of variance, thus, we can construct only one composite. Looking at the Italian situation in 2010, the 22.8% of people aged 14 and-over report smoking and the 20% of people aged 14 or-over have at least one behaviour at risk in the consumption of alcohol[8].

The percentage of people aged 3 years and-over who say they consume at least 4 portions of fruit and vegetables every day is only 17.3% and the 39.5% of people aged 14 and-over say they do not perform any physical activity. In 2017, all the indicators increase, with the exception of the adequate nutrition index[9] (16.4%).

5.4.4.1. Synthesis by using AMPI

The situation described before is well summarised by the composite (Figure 5.25), which in 2017 reaches the value of 108. The Italian trend is positive throughout the period considered (with a slight decrease in 2015). Figure 5.25 shows that the North-South gap, although present, is less marked than in other analyses. For instance, the Aosta Valley has a trend worse than the national one for the whole period analysed (mainly due to high values in alcohol consumption[10]). Among the Southern Regions, Sardinia has the best trend and from 2016 has values higher than the Italian ones; in particular, in 2017 it has the highest value in the adequate nutrition index among all the regional ones (23.6%). In addition, different combinations in basic indicators could be represented by the same composite. Apulia and Campania, despite having the same value in 2017 (89), have very different combinations in the basic indicators (Figure 5.26).

---

8 All individuals who practice at least one of the behaviours at risk are identified as "consumers at risk", exceeding the daily consumption of alcohol (according to specific thresholds for sex and age) or concentrating in a single occasion of consumption the assumption of more than 6 alcoholic units of any drink (binge drinking).

9 The indicator has a decreasing trend until 2015. In 2016, it pick-ups, followed in 2017 by a slight decrease.

10 Looking at the average values for simplicity, 25.5% of regional population present at least one behaviour at risk in the consumption of alcohol compared to 17.7% of the national population

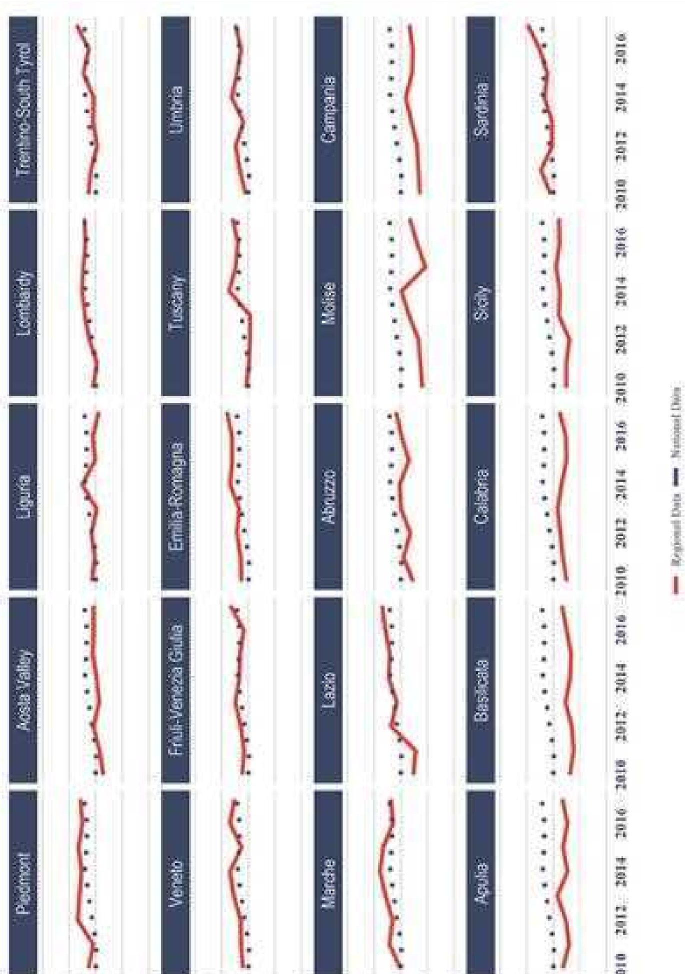


Fig. 5.25. Synthetic index of Goal 3 risk: regional and national data; time series 2010-2017; AMPI: Italy 2010 = 100.

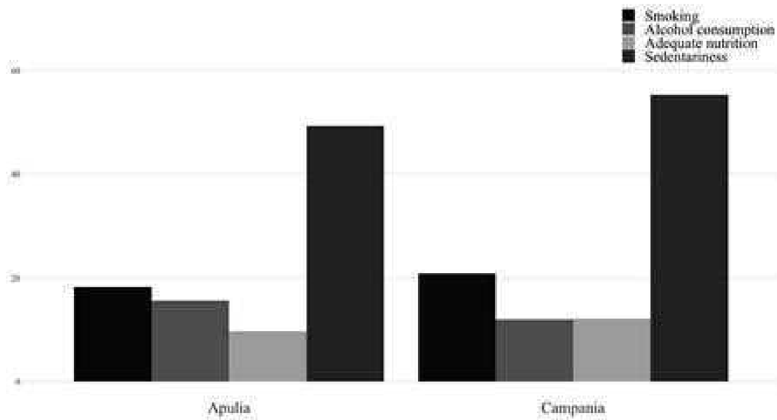


Fig. 5.26. Comparison between basic indicators of Goal 3 risk: Apulia and Campania in 2017.

5.4.4.2. Synthesis by using poset

Figure 5.27 reports the Hasse diagrams of different years. It is immediately evident that there are few comparabilities in the various years, highlighting a high level of fuzziness connected to the phenomenon. Figure 5.28 reports the temporal poset, in which we can observe all regions over the years, the Italian nodes and the embedded scale. Finally, Figure 5.29 reports the national and regional trends of the average height (index numbers at fixed base Italy 2010). The North-South gap is less marked than other situations. Italy and some regions show different trends from those obtained with the composite index (Figure 5.25).

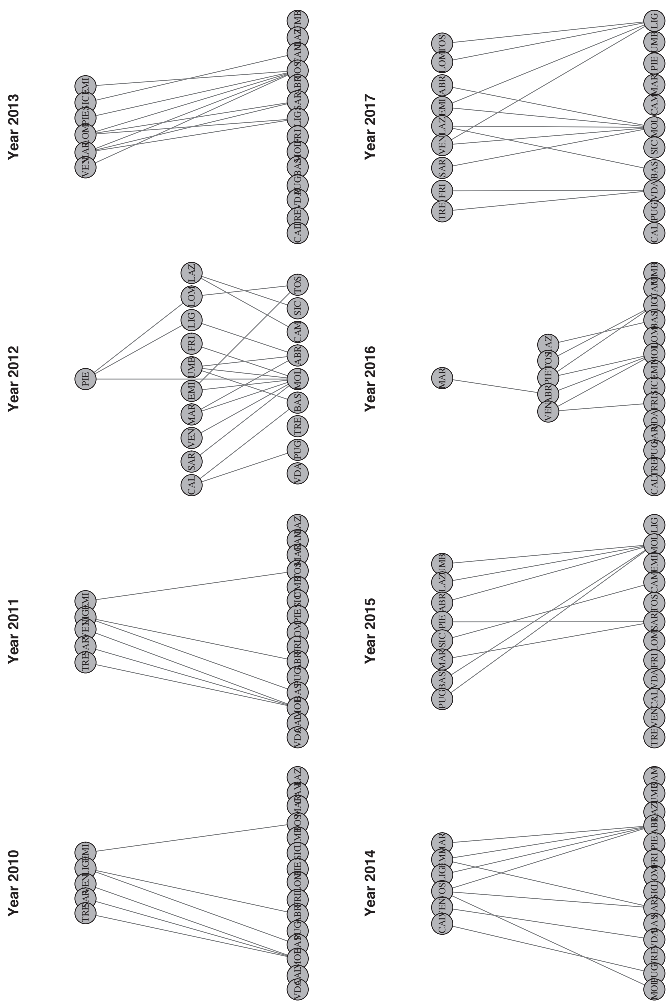


Fig. 5.27. Goal 3 risk: Hasse diagrams from 2010 to 2017.

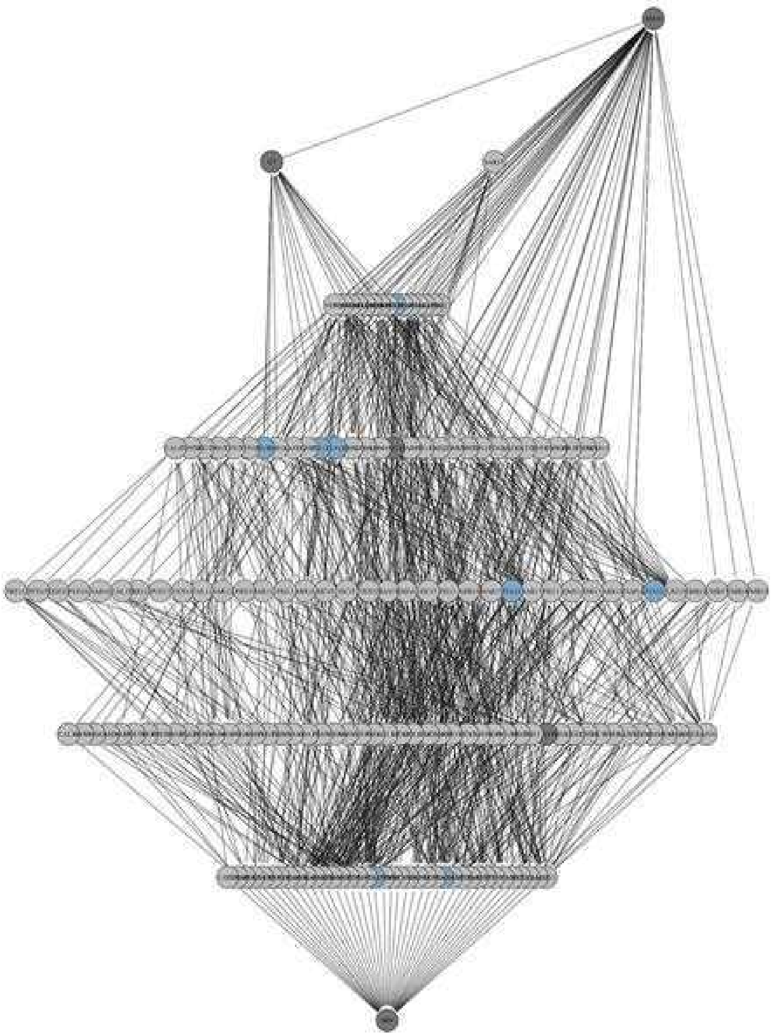


Fig. 5.28. Goal 3 risk: temporal poset years 2010-2017.

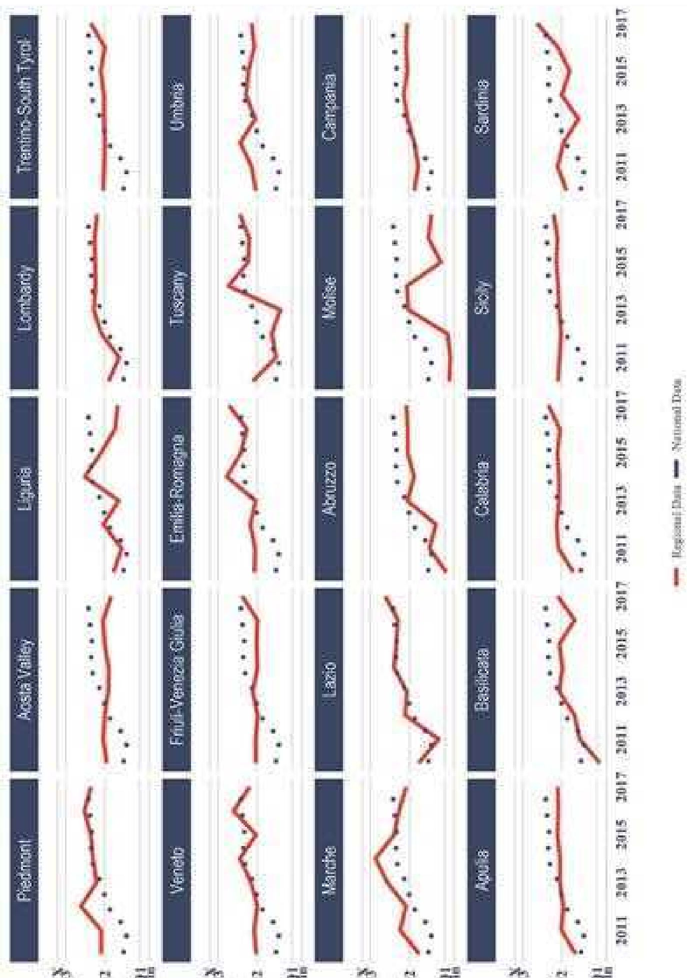


Fig. 5.29. Synthetic index of Goal 3 risk: regional and national data; time series 2010-2017; index numbers from average height distribution; fixed base = Italy 2010.

The rank correlation coefficient is less higher than those of other analyses ($\tau = 0.67$). As shown in Figure 5.30, the variable X12 (smoking) is not correlated with both synthetic measures and X13 (alcohol consumption) shows quite low correlation coefficients. On the basis of these results, consistent with those of explanatory analysis, it could be decided to divide the indicators considered, considering two different dimensions instead only one[11]. Apulia and Campania, having the same composite value, present different average height: Apulia has a value of 92.2 (the index number in equal to 1.85) and Campania 89.8 (the index number is equal to 1.8).

Fig. 5.30. Goal 3 risk: correlation between (CB) average height (AVH), AMPI and basic indicators.

5.5. Final and conclusive remarks

In this Chapter, we started by explaining the concept of sustainable development. The measurement of social phenomena, as we know, starts from their definition and from their inclusion within a theoretical

11 In this case, it was decided not to carry out this further analysis, because the aim is basically to compare the two different methods proposed.

framework that gives them meaning. Sustainable development is a new paradigm. However, there are still some obscure points linked to its definition, which needs to be reconsidered by taking into account its evolution over time. Moreover, many limits are emerging concerning the multidimensional approach, dominant in the literature, restricted to the three-way holistic framework. Even though the rising need to measure and monitor sustainable development urges the definition of a shared framework of goals, targets and indicators in a systemic way, in our opinion, it is not possible to conceive the indicators selected in the SDGs context as representing a system of indicators. Consistently with this reflection and with the multidimensional nature of the concept, we conclude that it makes no sense to seek a single synthetic measure of sustainable development, but that it is more appropriate to consider individual goals separately to fully understand their contribution to the level of well-being in society. A notion not sufficiently clarified is that sustainable development is not the aim, but only a means aimed at achieving well-being for present and future generations. This point is often not considered particularly by policymakers, who perceive sustainable development as part of what needs to be achieved, i.e. the well-being of the society. In policy perspective, limiting the monitoring to the national level is a weak approach, especially in countries with a high level of internal bio-socio-economic-territorial diversity. Consequently, declining SDGs indicators at regional level should not be an option but an unavoidable exercise. We limited ourselves to a restricted number of territorial analyses for each goal, by taking into account that the monitoring of all regions is required at policy level. Obviously, the methodology that has been applied is valid and extendable to all other goals. Moreover, we consider particularly relevant distinguishing between present conditions indicators and risks indicators, especially at policy level. The latter category is not intended as producing actual *predictions*, but could be seen as a good approximation of future well-being. In other words, we think that this approach allows us to take into account the well-being of actual and future generations.

We deal with some methodological issues regarding the synthesis of statistical indicators. We apply concepts analysed in Chapter 3. First, it is important, in view of synthesising indicators, to study each single indicator and its relation with the others. While this is perfectly clear in adopting the reflective model, this also represents a need in the formative one. This study should take into account how indicators (and consequently their mutual relationships) vary over time. In this Chapter, we applied the exploratory analysis (a methodological step often undervalued in the literature) described in paragraph 3.2; we analysed how the correlation *between* and *within* observations are different and both must be considered for a correct understanding of the relationships between indicators. The proposed methodology is an excellent tool for the preliminary study of the phenomenon to be synthesised.

Another important topic addressed was the construction of synthetic measures from multi-indicator systems over time. We applied a methodological proposal, described in paragraph 3.4.3, based on the Partial Order Theory. By using both $AMPI$ and poset, we reached some interesting findings.

- The application of the poset-based method provides results that are consistent with the phenomenon considered and with the synthesis obtained with $AMPI$. The analysis of the rank correlation between the results obtained with the two approaches shows, in all the applications reported, very high coefficients. This supports the methodological robustness of the proposed approach.

- The poset-based methodology offers a number of advantages over the aggregative approach. The analysis of regional structures in individual years through the Hasse diagrams provides, in itself, an immediate synthesis of the phenomenon. We can think of it as a kind of exploratory analysis that gives us a preliminary image of the temporal trend of the phenomenon. For example, by analysing the relationship structure of the regions with respect to Goal 1 risk (Figure 5.13), we can observe how Trentino-South Tyrol is always in a better position than the other regions for all the years considered or, on the contrary, how Sicily and Campania are always in the lowest positions of the Hasse diagrams. We can study the comparabilities present in each poset and analyse the structures in terms of uncertainty of the relations between regions. All this information is very important for the understanding of a phenomenon.

  In order to make inter-temporal comparisons between regions we merged the single-year posets and form a single *temporal* poset. The Hasse diagram of the latter provides us with a bi-dimensional representation of our three-way data time array. In the diagram, the temporal dimension is represented as a set of comparabilities and incomparabilities that are added to those already present in the original posets. The Hasse diagram that is obtained may be useless at first glance[12]. However, the diagram immediately highlights fundamental information about the structure of relationships over time of regions. For instance, let us take into account Figure 5.28. Sardinia, as described in paragraph 5.4.4.1, in 2017 has the highest value. In the Hasse diagram, we can observe how the node of Sardinia in 2017 (SAR17) is in very high position. This gives us immediate information even if we do not know anything about data.

Finally, the measure obtained is more efficient than synthesis through the aggregative approach. The poset-based method defines measures by a profiles analysis. This avoids the implementation of some procedures of data pre-processing (excluding the need to invert polarity if negative): in particular, no normalisation and aggregation of basic indicators are necessary. One of the results is a greater discriminatory capacity of this procedure than that of the aggregative methods, which we have seen in the different examples given in this Chapter. Combinations in different basic indicators correspond to different average heights. Another advantage is the possibility of making better evaluations of phenomena. $AMPI$ is certainly a good tool for synthesis and measurement, however it has a fundamental limit (in addition to those related to the aggregative-compensatory approach as such). The introduction of the reference value in the normalisation procedure allows the inter-temporal comparison of units with respect to this reference, but creates an instrument in which there is no longer an origin of the measurement system. This does not allow to make ratios between obtained values. For example, we can say that a region A with an $AMPI$ of 110 is better than a region B that has a value of 55. It is possible to say that the difference between the two values is equal to 55 points. However, we do not know what these 55 points correspond to, nor is it possible to say that region A has a value twice as high as region B. Average height (defined in paragraph 3.4.2.1) has a defined and closed range of

[12] It should also be noted that the images shown here do not allow the best possible graphic representation, for reasons of space and editorial format.

variation [1, total number of nodes]. We can, therefore, state that if a region A has an average height of 4, it has a higher value than a region B having 2. We can say that the difference between these two regions corresponds to 2 nodes. Moreover, the use of fixed base index numbers makes it possible to relate each average height to the reference one. For instance, in Goal 3 present condition 5.4.3, in 2017 Sicily has an average height of 18, and an index number of 0.30. Thus, it has a value lower than the Italian one in 2010 that is equal to 62. We can also say that the difference is equal to 44 nodes and that Sicily has an average height equal to 0.3 times that of Italy in 2010.

In conclusion, a correct understanding of the phenomena requires the use of procedures that respect the values of each unit's profiles. Aggregating is useful in order to simplify the complexity and allow an immediate representation of the phenomena; however, it does not allow a precise analysis, *crushing* and *flattening* the differences. Our conclusion is that making synthesis through compensative aggregation is not able to render a full understanding of the complexity of social phenomena. The poset-based approach seem to partially overcome this limit.

6. Conclusions

In this thesis I tried to answer a series of questions.

What is complexity? What are the characteristics of complex phenomena? I addressed this topic in the first Chapter. I reconstructed the history of the science of complexity and its relationship with knowledge. Beyond it could be defined as a new paradigm, its impact on the natural and social sciences was strong and undeniable. Complex phenomena have their own characteristics, one of the main ones is their adaptivity, their capacity of learning from experience. Reconstructing this concept in its many facets was a complex operation. Many aspects have probably only been dealt with superficially. The future objective is to deepen this topic and, in particular, its conceptualisation in sociology. Social phenomena can be included within these Complex Adaptive Systems. Their comprehension involves two questions. Understanding these phenomena means measuring them; measuring these phenomena means adopting a synthetic approach. Hence, the other two questions.

What does it mean to measure? What are the characteristics of an efficient measuring instrument? How to develop an efficient instrument? I answered these questions in the second Chapter. The concept of measurement was first addressed in its general meaning, and then in the sociological field. I presented the phases for the realisation and the main characteristics of the instrument for measuring complex social phenomena, the multi-indicator system. It is a complex tool, which needs a specific approach for its correct understanding. Synthesis is the guiding concept.

In the third Chapter, I dealt with the question of the synthesis of multi-indicator system. I approached this topic from a methodological point of view, considering the two typical aspects of synthesis of units and indicators. A large amount of literature exists on the subject; however, the temporal aspect has been poorly addressed. How to synthesize multi-indicator systems over time? What methods and statistical tools exist for this task? Is it possible to think of new methods, which do not have a strong impact on the initial data? All these questions have been addressed in the third Chapter. I formalised the problem of synthesis with respect to three-way data time arrays and highlighted its peculiarities. I proposed a method for the exploratory analysis of basic indicators over time. With respect to the synthesis of units, I presented the different approaches to cluster analysis of multivariate time series. With respect to the synthesis of indicators, I highlighted the main approaches and their strengths and weaknesses. Given some methodological limitations of the aggregative approach, the most used to process cardinal data, I proposed an innovative approach, based on posets.

The fourth and fifth Chapters present two applications that put into practice the process of measurement of complex phenomena. I analysed two important phenomena, well-being and sustainable development. I highlighted all the phases for the correct measurement of complexity, always starting from definition. Several analyses have been carried out. In the fourth Chapter, I used a technique for classifying time series that is particularly suitable for social phenomena. In the fifth Chapter, I compared a robust and widely used aggregative-compensatory method, $AMPI$, with my proposed poset-based procedure.

The future objective, from a methodological point of view, is to extend the poset-based method proposed in this thesis in two directions. The first is the application to ordinal data. The latter are the ideal reference for the use of posets; however, this method is currently not used for synthesising ordinal data over time. The second is the use of other synthetic measures, always derived from posets, that can further increase the level of accuracy and validity of the measurement.

Bibliography

[1] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi, and Teh Ying Wah. Time-series clustering–A decade review. *Information Systems*, 53:16–38, 2015.

[2] Miguel I. Aguirre-Urreta, Mikko Rönkkö, and George M. Marakas. Omission of Causal Indicators: Consequences and Implications for Measurement. *Measurement: Interdisciplinary Research and Perspectives*, 14(3):75–97, 2016.

[3] Leonardo 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.

[4] Leonardo S. Alaimo. Sustainable Development and Territorial Differences: An Italian Analysis of Economic Sustainability. *Working papers of the PhD Course in Applied Social Sciences, Department of Social Sciences and Economics, Sapienza University of Rome*, Paper n.1/2019, 2019.

[5] Leonardo S. Alaimo. Complex systems and complex adaptive systems. In F. Maggino, editor, *Encyclopedia of Quality of Life and Well-being Research*, pages 1–3. Cham: Springer, 2021. DOI: 10.1007/978-3-319-69909-7_104659-1.

[6] Leonardo S. Alaimo. Complexity and knowledge. In F. Maggino, editor, *Encyclopedia of Quality of Life and Well-being Research*,

pages 1–2. Cham: Springer, 2021. DOI: 10.1007/978-3-319-69909-7_104658-1.

[7] Leonardo S. Alaimo, Alberto Arcagni, Marco Fattore, and Filomena Maggino. Synthesis of multi-indicator system over time: A poset-based approach. *Social Indicators Research*, 157(1):77–99, 2021.

[8] Leonardo S. Alaimo, Alberto Arcagni, Marco Fattore, Filomena Maggino, and Valeria Quondamstefano. Measuring Equitable and Sustainable Well-being in Italian Regions. The Non-aggregative Approach. *Social Indicators Research*, 161:711–733, 2022. DOI: 10.1007/s11205-020-02388-7.

[9] Leonardo S. Alaimo, Andrea Ciacci, and Enrico Ivaldi. Measuring sustainable development by non-aggregative approach. *Social Indicators Research*, 157(1):101–122, 2021.

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

[11] Leonardo S. Alaimo, Mariantonietta Fiore, and Antonino Galati. Measuring consumers' level of satisfaction for online food shopping during COVID-19 in Italy using posets. *Socio-Economic Planning Sciences*, 82:101064, 2022. DOI: 10.1016/j.seps.2021.101064.

[12] Leonardo S. Alaimo, Enrico Ivaldi, Stefano Landi, and Filomena 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.

[13] Leonardo S. Alaimo and Filomena Maggino. Sustainable Development Goals Indicators at Territorial Level: Conceptual and Methodological Issues — The Italian Perspective. *Social Indicators Research*, 147(2):383–419, 2020. DOI: 10.1007/s11205-019-02162-4.

[14] Leonardo S. Alaimo and Giulia Nanni. Gender Inequality in Italy: A Territorial Analysis. *RIEDS-Rivista Italiana di Economia,*

Demografia e Statistica-Italian Review of Economics, Demography and Statistics, 72(4):41–52, 2018.

[15] Leonardo S. Alaimo and Luigi M. Solivetti. Territorial Determinants of the Brexit Vote. *Social Indicators Research*, 144(2):647–667, 2019.

[16] Sönke Albers and Lutz Hildebrandt. Methodische Probleme bei der Erfolgsfaktorenforschung—Messfehler, formative versus reflektive Indikatoren und die Wahl des Strukturgleichungs-Modells. *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 58(1):2–33, 2006.

[17] James C. Anderson and David W. Gerbing. Some Methods for Respecifying Measurement Models to Obtain Unidimensional Construct Measurement. *Journal of marketing research*, 19(4):453–460, 1982.

[18] Paola Annoni and Rainer Brüggemann. Exploring Partial Order of European Countries. *Social indicators research*, 92(3):471–487, 2009.

[19] Alberto Arcagni, Elisa Barbiano di Belgiojoso, Marco Fattore, and Stefania M.L. 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.

[20] Alberto Arcagni and Marco Fattore. *parsec: Partial Orders in Socio-Economics*, 2019. R package version 1.2.3.

[21] Henri Atlan. *Entre le cristal et la fumée. Essai sur l'organisation du vivant*. Paris: Edit. du Seuil, 2015.

[22] Richard P. Bagozzi. On the Meaning of Formative Measurement and How it Differs from Reflective Measurement: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2):229–237, 2007.

[23] Jonathan Z. Bakdash and Laura R. Marusich. Repeated Measures Correlation. *Frontiers in Psychology*, 8. Article number 456:1–13, 2017.

[24] Arthur Battram. *Navigating Complexity: The Essential Guide to Complexity Theory in Business and Management*. London: Industrial Society, 1999.

[25] Sascha O. Becker, Thiemo Fetzer, and Dennis Novy. Who voted for Brexit? A comprehensive district-level analysis. *Economic Policy*, 32(92):601–650, 2017.

[26] William Becker, Michaela Saisana, Paolo Paruolo, and Ine Vandecasteele. Weights and Importance in Composite Indicators: Closing the Gap. *Ecological Indicators*, 80:12–22, 2017.

[27] Gleb Beliakov, Ana Pradera, and Tomasa Calvo. *Aggregation Functions: A Guide for Practitioners*, volume 221. Dordrecht: Springer, 2007.

[28] Donald J. Berndt and James Clifford. Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the AAAI-94 Workshop Knowledge Discovery in Databases*, pages 359–370. Seattle, WA, 1994.

[29] Ludwig von Bertalanffy. *General System Theory: Foundations, Devolpment, Applications*. New York: George Braziller, 1968.

[30] Cristoforo S. Bertuglia and Franco Vaio. *Nonlinearity, Chaos, and Complexity: The Dynamics of Natural and Social Systems*. Oxford: Oxford University Press, 2005.

[31] Tugce Beycan, V. Balasubramanian, Rainer Bruggemann, and Christian Suter. Ranking Karnataka Districts by the Multidimensional Poverty Index (MPI) and by Applying Simple Elements of Partial Order Theory. *Social Indicators Research*, 143(1):173–200, 2019.

[32] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA: Kluwer Academic, 1981.

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

[34] Hubert 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.

[35] Martin J. Bland and Douglas G. Altman. Calculating Correlation coefficients with Repeated Observations: Part 1—Correlation within Subjects. *British Medical Journal*, 310(6980):446, 1995.

[36] Martin J. Bland and Douglas G. Altman. Calculating Correlation Coefficients with Repeated Observations: Part 2—Correlation between Subjects. *British Medical Journal*, 310(6980):633, 1995.

[37] John Blewitt. *Understanding Sustainable Development*. London: Routledge, 2012.

[38] Brent Bleys. Beyond GDP: Classifying Alternative Measures for Progress. *Social Indicators Research*, 109(3):355–376, 2012.

[39] Herbert Blumer. What is Wrong with Social Theory? *American Sociological Review*, 18:3–10, 1954.

[40] Herbert Blumer. *Symbolic Interactionism: Perspective and Method*. Berkeley and Los Angeles, California: University of California Press, 1969.

[41] Romina Boarini, Åsa Johansson, and Marco Mira d'Ercole. Alternative Measures of Well-Being. *OECD Social, Employment and Migration Working Papers*, 33, 2006.

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

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

[44] Kenneth A. Bollen. Interpretational Confounding is due to Misspecification, not to Type of Indicator: Comment on Howell, Breivik, and Wilcox (2007). *Psychological Methods*, 12(2):219–228, 2007.

[45] Kenneth A. Bollen and Adamantios Diamantopoulos. In Defense of Causal-formative Indicators: A Minority Report. *Psychological Methods*, 22(3):581–596, 2017.

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

[47] Alan J. Bond and Angus Morrison-Saunders. Sustainability Appraisal: Jack of All Trades, Master of None? *Impact Assessment and Project Appraisal*, 27(4):321–329, 2009.

[48] Raymond Boudon and Paul F. Lazarsfeld. *Le vocabulaire des sciences sociales et l'analyse empirique de la causalité*. Paris: Mouton, 1965.

[49] Glenn A. Bowen. Grounded Theory and Sensitizing Concepts. *International Journal of Qualitative Methods*, 5(3):12–23, 2006.

[50] George E.P. Box. Science and Statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

[51] George E.P. Box and Norman R. Draper. *Empirical Model-building and Response Surfaces*. John Wiley & Sons, 1987.

[52] Gaël Brulé and Filomena Maggino. Towards More Complexity in Subjective Well-Being Studies. In G. Brulé and F. Maggino, editors, *Metrics of Subjective Well-Being: Limits and Improvements*, pages 1–17. Cham: Springer, 2017.

[53] Hans Bruyninckx, Sander Happaerts, Karoline Van den Brande, and Karoline van den Brande. *Sustainable Development and Subnational Governments: Policy-making and Multi-level Interactions*. Basingstoke: Palgrave Macmillan, 2012.

[54] Jorge Caiado, Elizabeth A. Maharaj, and Pierpaolo D'Urso. Time-Series Clustering. In C. Henning, M. Meila, F. Murtagh, and R. Rocci, editors, *Handbook of Cluster Analysis*, pages 241–263. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2015.

[55] Ricardo J.G.B. Campello and Eduardo R. Hruschka. A Fuzzy Extension of the Silhouette Width Criterion for Cluster Analysis. *Fuzzy Sets and Systems*, 157(21):2858–2875, 2006.

[56] Giulio Caperna and Giovanna Boccuzzo. Use of Poset Theory with Big Datasets: A New Proposal applied to the Analysis of Life Satisfaction in Italy. *Social Indicators Research*, 136(3):1071–1088, 2018.

[57] Fritjof Capra. *The Web of Life: A New Scientific Understanding of Living Things*. New York: Anchor books, 1996.

[58] Fritjof Capra. *The Hidden Connections*. New York: Doubleday, 2002.

[59] Fritjof Capra and Pier Luigi Luisi. *The Systems View of Life: A Unifying Vision*. Cambridge: Cambridge University Press, 2014.

[60] Lars Carlsen and Rainer Brüggemann. Fragile State Index: Trends and Developments. A partial Order Data Analysis. *Social indicators research*, 133(1):1–14, 2017.

[61] Nicholas R. Chrisman. Rethinking Levels of Measurement for Cartography. *Cartography and Geographic Information Systems*, 25(4):231–242, 1998.

[62] Glenda Chui. "Unified theory" is getting closer, Hawking predicts. *San Jose Mercury News*, 23:29, 2000.

[63] Clifford Cobb, Ted Halstead, and Jonathan Rowe. *The Genuine Progress Indicator: summary of data and methodology*. San Francisco: Redefining Progress, 1995.

[64] Tim Coltman, Timothy M. Devinney, David F. Midgley, and Sunil Venaik. Formative versus Reflective Measurement Models: Two Applications of Formative Measurement. *Journal of Business Research*, 61(12):1250–1262, 2008.

[65] Rina M. Contini. The Paradigm of the Complex Dynamic Systems and Sociological Analysis. *Procedia-Social and Behavioral Sciences*, 92:207–214, 2013.

[66] Renato Coppi and Sergio Bolasco. *Multiway Data Analysis*. Amsterdam: North-Holland Publishing Co., 1989.

[67] Renato Coppi and Pierpaolo D'Urso. Fuzzy *K*-means Clustering Models for Triangular Fuzzy Time trajectories. *Statistical Methods and Applications*, 11(1):21–40, 2002.

[68] Renato Coppi and Pierpaolo D'Urso. Three-way Fuzzy Clustering Models for LR Fuzzy Time Trajectories. *Computational Statistics and Data Analysis*, 43(2):149–177, 2003.

[69] Renato Coppi, Pierpaolo D'Urso, and Paolo Giordani. Fuzzy C-Medoids Clustering Models for Time-varying data. In B. Bouchon-Meunier, G. Coletti, and R.R. Yager, editors, *Modern Information*

Processing: From Theory to Applications, pages 195–206. Amsterdam: Elsevier Science, 2006.

[70] Renato Coppi, Pierpaolo D'Urso, and Paolo Giordani. A Fuzzy Clustering Model for Multivariate Spatial Time Series. *Journal of Classification*, 27(1):54–88, 2010.

[71] Lee J. Cronbach and Paul E. Meehl. Construct Validity in Psychological Tests. *Psychological Bulletin*, 52:281–302, 1955.

[72] Gabor Csardi and Tamas Nepusz. The igraph Software Package for Complex Network Research. *InterJournal*, Complex Systems:1695, 2006.

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

[74] Barry Dalal-Clayton and Barry Sadler. *Strategic Environmental assessment: A Sourcebook and Reference Guide to International Experience*. London: Earthscan, 2005.

[75] Rajesh N. Dave. Validating Fuzzy Partitions Obtained through C-shells Clustering. *Pattern Recognition Letters*, 17(6):613–623, 1996.

[76] Brian A. Davey and Hilary A. Priestley. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press, 2002.

[77] Alberto F. De Toni and Luca Comello. *Prede o ragni: Uomini e organizzazioni nella ragnatela della complessità*. Torino: Utet libreria, 2005.

[78] Alberto F. De Toni and Luca Comello. *Journey Into Complexity*. Venezia: Marsilio Editori, 2010.

[79] Dimitri Devuyst. Linking Impact Assessment and Sustainable Development at the Local Level: The Introduction of Sustainability Assessment Systems. *Sustainable development*, 8(2):67–78, 2000.

[80] Enrico di Bella, Luca Gandullia, Lucia Leporatti, Marcello Montefiori, and Patrizia Orcamo. Ranking and Prioritization of Emergency Departments Based on Multi-indicator Systems. *Social Indicators Research*, 136(3):1089–1107, 2018.

[81] Adamantios Diamantopoulos, Petra Riefler, and Katharina P. Roth. Advancing Formative Measurement Models. *Journal of Business Research*, 61(12):1203–1218, 2008.

[82] Adamantios Diamantopoulos and Judy A. Siguaw. Formative versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration. *British Journal of Management*, 17(4):263–282, 2006.

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

[84] Hermann Diels and Walther Kranz. *Die Fragmente der Vorsokratiker. Volume 1*. Berlin: Weidmann, 1951.

[85] Ed Diener. Assessing Subjective Well-being: Progress and Opportunities. *Social Indicators Research*, 31(2):103–157, 1994.

[86] Ed Diener and Martin E.P. Seligman. Beyond Money: Toward an Economy of Well-being. *Psychological Science in the Public Interest*, 5(1):1–31, 2004.

[87] Ed Diener and William Tov. National Accounts of Well-being. In K.C. Land, A.C. Michalos, and M.J. Sirgy, editors, *Handbook of Social Indicators and Quality of Life Research*, pages 137–157. Dordrecht: Springer, 2012.

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

[89] Émile Durkheim. *Le suicide*. Paris: Hachette Livre, 1897.

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

[91] Pierpaolo D'Urso. Dissimilarity Measures for Time Trajectories. *Statistical Methods & Applications*, 9(1-3):53–83, 2000.

[92] Pierpaolo D'Urso. Fuzzy Clustering. In C. Henning, M. Meila, F. Murtagh, and R. Rocci, editors, *Handbook of Cluster Analysis*, pages 545–573. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2015.

[93] Pierpaolo D'Urso, Livia De Giovanni, and Riccardo Massari. Robust fuzzy clustering of multivariate time trajectories. *International Journal of Approximate Reasoning*, 99:12–38, 2018.

[94] Pierpaolo D'Urso and Elizabeth A. Maharaj. Wavelets-based Clustering of Multivariate Time series. *Fuzzy Sets and Systems*, 193:33–61, 2012.

[95] Pierpaolo D'Urso, Leonardo S. Alaimo, Livia De Giovanni, and Riccardo Massari. Well-being in the Italian regions over time. *Social Indicators Research*, 161:599–627, 2022. doi: 10.1007/s11205-020-02384-x.

[96] Pierpaolo D'Urso, Dario Di Lallo, and Elizabeth A. Maharaj. Autoregressive Model-based Fuzzy Clustering and Its Application for Detecting Information Redundancy in Air Pollution Monitoring Networks. *Soft Computing*, 17(1):83–131, 2013.

[97] Pierpaolo D'Urso and Elizabeth A. Maharaj. Autocorrelation-based Fuzzy Clustering of Time Series. *Fuzzy Sets and Systems*, 160(24):3565–3589, 2009.

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

[99] Jeffrey R. Edwards and Richard P. Bagozzi. On the Nature and Direction of Relationships between Constructs and Measures. *Psychological Methods*, 5(2):155–174, 2000.

[100] Jennifer Elliott. *An Introduction to Sustainable Development*. London: Routledge, 2012.

[101] Lee H. Endress. Scarcity, Security, and Sustainable Development. In U. Chakravorty, A. Balisacan, and M.–L. Ravago, editors, *Sustainable Economic Development: Resources, Environment, and Institutions*, pages 49–66. San Diego: Academic Press, 2015.

[102] Péter Érdi. *Complexity Explained*. Berlin: Springer, 2008.

[103] Andrea Falcon. Aristotle on causality. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Stanford: Stanford University Press, 2008.

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

[105] Marco Fattore. Synthesis of Indicators: The Non-aggregative Approach. In F. Maggino, editor, *Complexity in Society: From Indicators Construction to their Synthesis*, pages 193–212. Cham: Springer, 2017.

[106] Marco Fattore. Non-aggregated Indicators of Environmental Sustainability. *Silesian Statistical Review/Slaski Przeglad Statystyczny*, 16(22):7–22, 2018.

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

[108] Richard P. Feynman. *The Character of Physical Law*. London: Penguin Books, 1992.

[109] Mariantonietta Fiore, Leonardo S. Alaimo, and Nino Chkhartishvil. The Amazing Bond among Wine Consumption, Health and Hedonistic Well-being. *British Food Journal*, 122(8):2707–2723, 2020. DOI: 10.1108/BFJ-05-2019-0344.

[110] Michael Freudenber. Composite Indicators of Country Performance. A Critical Assessment. *OECD Science, Technology and Industry Working Papers*, page 13, 2003. DOI: 10.1787/405566708255.

[111] Sylvia Frühwirth-Schnatter. Panel Data Analysis: A Survey on Model-based Clustering of Time Series. *Advances in Data Analysis and Classification*, 5(4):251–280, 2011.

[112] Elisa Fusco, Francesco Vidoli, and Biresh K. Sahoo. Spatial Heterogeneity in Composite Indicator: A Methodological Proposal. *Omega*, 77:1–14, 2018.

[113] Antonino Galati, Leonardo S. Alaimo, Teresa Ciaccio, Demetris Vrontis, and Mariantonietta Fiore. Plastic or not plastic? that's the problem: analysing the italian students purchasing behavior of mineral water bottles made with eco-friendly packaging. *Resources, Conservation and Recycling*, 172:106060, 2022.

[114] Xiaoyu Gan, Ignacio C. Fernandez, Jie Guo, Maxwell Wilson, Yuanyuan Zhao, Bingbing Zhou, and Jianguo Wu. When to Use What: Methods for Weighting and Aggregating Sustainability Indicators. *Ecological Indicators*, 81:491–502, 2017.

[115] Alberto Gandolfi. *Formicai, imperi, cervelli: Introduzione alla scienza della complessità*. Bellinzona: Edizioni Casagrande, 1999.

[116] Des Gasper. Human Well-Being: Concepts and Conceptualizations. In M. McGillivray, editor, *Human Well-Being. Studies in Development Economics and Policy*, pages 23–64. London: Palgrave Macmillan, 2007.

[117] Christophe Genolini and Bruno Falissard. KmL: K-means for Longitudinal Data. *Computational Statistics*, 25(2):317–328, 2010.

[118] Robert B. Gibson, Selma Hassan, and James Tansey. *Sustainability Assessment: Criteria and Processes*. London; Washington: Earthscan, 2005.

[119] Toni Giorgino. Computing and visualizing dynamic time warping alignments in r: the dtw package. *Journal of statistical Software*, 31:1–24, 2009.

[120] Enrico Giovannini and Tommaso Rondinella. Measuring Equitable and Sustainable Well-Being in Italy. In F. Maggino and G. Nuvolati, editors, *Quality of Life in Italy. Research and Reflections*, pages 9–25. Cham: Springer, 2012.

[121] Matthew J. Goodwin and Oliver Heath. The 2016 referendum, Brexit and the left behind: An aggregate-level analysis of the result. *The Political Quarterly*, 87(3):323–332, 2016.

[122] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. *Aggregation Functions*, volume 127. Cambridge: Cambridge University Press, 2009.

[123] Ulrich Grober. The Discovery of Sustainability: The Genealogy of a Term. In J.C. Enders and M. Remig, editors, *Theories of Sustainable Development*, pages 6–25. London: Routledge, 2016.

[124] Mònica Guillen-Royo. *Sustainability and Wellbeing: Human-scale Development in Practice*. New York: Earthscan, Routledge, 2016.

[125] Hermann Haken. *Synergetics: An Introduction*. Heidelberg: Springer-Verlag, 1983.

[126] Hermann Haken. L'approccio della sinergetica al problema dei sistemi complessi. In G. Bocchi and M. Ceruti, editors, *La sfida della complessità*, pages 194–206. Milano: Feltrinelli, 1985.

[127] Gerhard Hartmuth, Katja Huber, and Dieter Rink. Operationalization and Contextualization of Sustainability at the Local Level. *Sustainable Development*, 16(4):261–270, 2008.

[128] Takehiro Hatakeyama. Sustainable Development Indicators: Conceptual Frameworks of Comparative Indicators Sets for Local Administrations in Japan. *Sustainable Development*, 26(6):683–690, 2018.

[129] HDTR. Human Development Report 2014 Team. *Human Development Report 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience*, 2014. New York: United Nations Development Programme.

[130] Christian Hennig and Marina Meila. Cluster Analysis: An Overview. In C. Henning, M. Meila, F. Murtagh, and R. Rocci, editors, *Handbook of Cluster Analysis*, pages 1–20. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2015.

[131] Marta Herva, Amaya Franco, Eugenio F. Carrasco, and Enrique Roca. Review of Corporate Environmental Indicators. *Journal of Cleaner Production*, 19(15):1687–1699, 2011.

[132] Max Hocutt. Aristotle's four becauses. *Philosophy*, 49(190):385–399, 1974.

[133] Robert V. Horn. *Statistical Indicators: For the Economic and Social Sciences*. Cambridge: Cambridge University Press, 1993.

[134] Harold Hotelling. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of educational psychology*, 24(6):417–441, 1933.

[135] Roy D. Howell, Einar Breivik, and James B. Wilcox. Reconsidering Formative Measurement. *Psychological Methods*, 12(2):205–218, 2007.

[136] Istat Italian National Institute of Statistics. *Rapporto BES 2015: il benessere equo e sostenibile in Italia*. Rome: Istat, 2015.

[137] Istat Italian National Institute of Statistics. Indicators for the Measurement of Equitable and Sustainable Well-being (BES), 2018. Data retrieved from Istat datawarehouse, https://www.istat.it/en/well-being-and-sustainability/the-measurement-of-well-being/indicators.

[138] Istat Italian National Institute of Statistics. Indicators for the United Nations Sustainable Development Goals, 2018. Data retrieved from Istat datawarehouse, https://www.istat.it/en/well-being-and-sustainability/sustainable-development-goals/istat-indicators-for-sustainable-development.

[139] Istat Italian National Institute of Statistics. Territorial Indicators for Development Policies, 2018. Data retrieved from Istat datawarehouse, https://www.istat.it/it/archivio/16777.

[140] Hesam Izakian, Witold Pedrycz, and Iqbal Jamal. Fuzzy Clustering of Time series Data Using Dynamic Time Warping Distance. *Engineering Applications of Artificial Intelligence*, 39:235–244, 2015.

[141] Young-Seon Jeong, Myong K. Jeong, and Olufemi A. Omitaomu. Weighted Dynamic Time Warping for Time Series Classification. *Pattern Recognition*, 44(9):2231–2240, 2011.

[142] Ravi Kanbur, Ebrahim Patel, and Joseph Stiglitz. Sustainable Development Goals and Measurement of Economic and Social Progress. In J. Stiglitz, J. Fitoussi, and M. Durand, editors, *For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, pages 33–48. Paris: OECD Publishing, 2018.

[143] Alboukadel Kassambara and Fabian Mundt. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2017. R package version 1.0.5.

[144] Maurice G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1/2):81–93, 1938.

[145] Robert F. Kennedy. Speech at the University of Kansas, 3 March 18, 1968.

[146] Adalbert Kerber. Evaluation, Considered as Problem Orientable Mathematics over Lattices. In M. Fattore and R. Brüggemann, editors, *Partial Order Concepts in Applied Sciences*, pages 87–103. Dordrecht: Springer, 2017.

[147] Adalbert Kerber and Rainer Brüggemann. Problem Driven Evaluation of Chemical Compounds and Its Exploration. *MATCH Commun Math Comput Chem*, 73:577–618, 2015.

[148] Henk A.L. Kiers and Iven Van Mechelen. Three-way Component Analysis: Principles and Illustrative Application. *Psychological methods*, 6(1):84–110, 2001.

[149] Julia 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.

[150] Katarina Košmelj. A Two-step Procedure for Clustering Time Varying Data. *Journal of Mathematical Sociology*, 12(3):315–326, 1986.

[151] Katarina Košmelj and Vladimir Batagelj. Cross-sectional Approach for Clustering Time Varying Data. *Journal of Classification*, 7(1):99–109, 1990.

[152] Pieter M. Kroonenberg. *Three-mode Principal Component Analysis: Theory and Applications*. Leiden: DSWO press, 1983.

[153] Pieter M. Kroonenberg. *Applied Multiway Data Analysis*, volume 702. Hoboken, New Jersey: John Wiley & Sons, 2008.

[154] Rudolf Kruse, Christian Döring, and Marie-Jeanne Lesot. Fundamentals of Fuzzy Clustering. In J. V. De Oliveira and W. Pedrycz, editors, *Advances in Fuzzy Clustering and Its Applications*, pages 3–30. Chichester, England: John Wiley and Sons, 2007.

[155] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press, 1970.

[156] Simon Kuznets. National Income, 1929-1932. In *National Income, 1929-1932*, pages 1–12. NBER, 1934.

[157] Francesco La Camera and Edo Ronchi. *Sviluppo sostenibile: Origini, teoria e pratica*. Roma: Editori riuniti, 2005.

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

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

[160] Ervin Laszlo. *Evoluzione*. Milano: Feltrinelli, 1986.

[161] Serge Latouche. *Farewell to Growth*. Cambridge: Polity Press, 2009.

[162] Paul F. Lazarsfeld. Evidence and Inference in Social Research. *Daedalus*, 87(4):99–130, 1958.

[163] Amedeo Lepore. Il divario Nord-Sud dalle origini a oggi. Evoluzione storica e profili economici. *Rivista economica del Mezzogiorno*, 26(3):389–410, 2012.

[164] Hugo Letiche, Michael Lissack, and Ron Schultz. *Coherence in the Midst of Complexity: Advances in Social Complexity Theory*. Basingstoke: Palgrave Macmillan, 2012.

[165] Manuel Lima. *Visual Complexity: Mapping Patterns of Information*. New York: Princeton Architectural Press, 2013.

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

[167] Filomena Maggino. *La misurazione nella ricerca sociale: teorie, strategie, modelli*. Collana Astris. Applicazioni statistiche nella ricerca sociale. Firenze: Firenze University Press, 2004.

[168] Filomena Maggino. Assessing the Subjective Wellbeing of Nations. In W. Glatzer, L. Camfield, V. Møller, and M. Rojas, editors, *Global Handbook of Quality of Life*, pages 803–822. Dordrecht: Springer, 2015.

[169] Filomena Maggino. Subjective Well-being and Subjective Aspects of Well-being: Methodology and Theory. *Rivista internazionale di scienze sociali*, 1:89–121, 2015.

[170] Filomena Maggino. Challenges, Needs and Risks in Defining Wellbeing Indicators. In F. Maggino, editor, *A Life Devoted to Quality of Life*, pages 209–233. Cham: Springer, 2016.

[171] Filomena Maggino. Dealing with Syntheses in a System of Indicators. In F. Maggino, editor, *Complexity in Society: From Indicators Construction to their Synthesis*, pages 115–137. Cham: Springer, 2017.

[172] Filomena Maggino. Developing Indicators and Managing the Complexity. In F. Maggino, editor, *Complexity in Society: From Indicators Construction to their Synthesis*, pages 87–114. Cham: Springer, 2017.

[173] Filomena Maggino and Leonardo S. Alaimo. Complexity and wellbeing: measurement and analysis. In L. Bruni, A. Smerilli, and D. De Rosa, editors, *A Modern Guide to the Economics of Happiness*, pages 113–128. Edward Elgar Publishing, 2021.

[174] Filomena Maggino and Leonardo S. Alaimo. Measuring complex socio-economic phenomena. conceptual and methodological issues. In S. Valaguzza and M. A. Hughes, editors, *Interdisciplinary Approaches to Climate Change for Sustainable Growth*, pages 43–59. Cham: Springer, 2022.

[175] Filomena Maggino, Rainer Bruggemann, and Leonardo S. Alaimo. Indicators in the framework of partial order. In Rainer Bruggemann, Lars Carlsen, Tugce Beycan, Christian Suter, and Filomena Maggino, editors, *Measuring and Understanding Complex Phenomena*, pages 17–29. Springer, 2021.

[176] Elizabeth A. Maharaj. Cluster of Time Series. *Journal of Classification*, 17(2):297–314, 2000.

[177] Elizabeth A. Maharaj, Pierpaolo D'Urso, and Jorge Caiado. *Time series Clustering and Classification*. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2019.

[178] Elizabeth A. Maharaj and Pierpaolo D'Urso. Fuzzy Clustering of Time Series in the Frequency domain. *Information Sciences*, 181(7):1187–1211, 2011.

[179] Elizabeth A. Maharaj, Pierpaolo D'Urso, and Don U.A. Galagedera. Wavelet-based Fuzzy Clustering of Time Series. *Journal of Classification*, 27(2):231–275, 2010.

[180] Nic Marks, Andrew Simms, San Thompson, and Saamah Abdallah. The (Un)Happy Planet Index: An index of human well-being and environmental impact., 2006.

[181] Margaret Masterman. The Nature of a Paradigm. In I. Latakos and A. Musgrave, editors, *Criticism and the Growth of Knowledge*, pages 59–90. Cambridge: Cambridge University Press, 1970.

[182] Humberto R. Maturana and Francisco J. Varela. *Autopoiesis and Cognition: the Realization of the Living*. Dordrecht, Holland: D. Reidel Pub. Co, 1980.

[183] Claudio Mazziotta, Matteo Mazziotta, Adriano Pareto, and Francesco Vidoli. La sintesi di indicatori territoriali di dotazione infrastrutturale: metodi di costruzione e procedure di ponderazione a confronto. *Rivista di Economia e Statistica del Territorio*, 1(1):7–33, 2010.

[184] Matteo Mazziotta. Composite Indicators for Measuring Wellbeing of Italian Municipalities. *Unpublished Doctoral Dissertation, University of Rome" La Sapienza", Rome, Italy*, 2019.

[185] Matteo Mazziotta and Adriano Pareto. Comparing Two Noncompensatory Composite Indices to Measure Changes Over Time: A Case study. *Statistika-Statistics and Economy Journal*, 95(2):44–53, 2015.

[186] Matteo Mazziotta and Adriano Pareto. On a Generalized Noncompensatory Composite Index for Measuring Socio-economic Phenomena. *Social indicators research*, 127(3):983–1003, 2016.

[187] Matteo Mazziotta and Adriano 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.

[188] Peter McCullagh and John A. Nelder. *Generalized Linear Models*. London, New York: Chapman and Hall, 1983.

[189] Donella H. Meadows. *Thinking in Systems: A Primer*. London: Earthscan, 2009.

[190] Donella H. Meadows, Dennis L. Meadows, Jorgen Randers, and William W. III Behrens. *Limits to Growth*. New York: Universe Book, 1972.

[191] Robert K. Merton. *Social Theory and Social Structure*. New York: The Free Press, 1957.

[192] Robert K. Merton. *On the Shoulders of Giants: A Shandean Postscript*. San Diego (Calif.): Harcourt Brace Jovanivich, 1985.

[193] Robert K. Merton and Matilda W. Riley. *Sociological Traditions from Generation to Generation: Glimpses of the American Experience*. Michigan University: Ablex Publishing, 1980.

[194] Joel Michell. Bertrand Russell's 1897 Critique of the Traditional Theory of Measurement. *Synthese*, 110(2):257–276, 1997.

[195] Joel Michell. *Measurement in Psychology: A Critical History of a Methodological Concept*. Cambridge: Cambridge University Press, 1999.

[196] James G. Miller. *Theory of Living Systems*. Niwot, Colorado: University of Colorado Press, 1995.

[197] John H. Miller and Scott E. Page. *Complex Adaptive systems: An Introduction to Computational Models of Social Life*, volume 17. Princeton: Princeton University Press, 2009.

[198] Marc Miringoff. Index of Social Health. *Institute for Innovation in Social Policy, Fordham University, Tarrytown, New York*, 1995.

[199] Bedrich Moldan and Arthur L. Dahl. Challenges to Sustainability Indicators. In T. Hák, B. Moldan, and A.L. Dahl, editors, *Sustainability Indicators: A Scientific Assessment*, pages 1–24. Washington: Island Press, 2012.

[200] Sara Moreno Pires, Teresa Fidélis, and Tomás B. Ramos. Measuring and Comparing Local Sustainable Development Through Common Indicators: Constraints and Achievements in Practice. *Cities*, 39:1–9, 2014.

[201] Edgar Morin. *The Red and the White: Report from a French Village*. New York: Pantheon Books, 1970.

[202] Edgar Morin. *La méthode. Tome I. La nature de la nature*. Paris: Le Seuil, 1977.

[203] Edgar Morin. Le vie della complessità. In G. Bocchi and M. Ceruti, editors, *La sfida della complessità*, pages 49–60. Milano: Feltrinelli, 1985.

[204] Edgar Morin. Restricted Complexity, General Complexity. In C. Gershenson, D. Aerts, and B. Edmonds, editors, *Worldviews, Science and Us: Philosophy and Complexity*, pages 1–25. Singapore: World Scientific, 2007.

[205] Edgar Morin. *On Complexity*. Chesskill (New Jersey): Hampton Press, 2008.

[206] Enrico Morselli. *Il suicidio. Saggio di statistica morale comparata*. Milano, Dumolard, 1879.

[207] Frederick Mosteller and John W. Tukey. *Data Analysis and Regression*. Boston: Addison-Wesley, 1977.

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

[209] Michela Nardo, Michaela Saisana, Andrea Saltelli, and Stefano Tarantola. *Tools for Composite Indicators Building.* European Commission. Institute for the Protection and Security of the Citizen, JRC: Ispra, Italy, 2005.

[210] Joseph Neggers and Hee S. Kim. *Basic Posets*. Singapore: World Scientific Publishing, 1998.

[211] Jum C. Nunnally. *Psychometric Theory. Second edition*. New York: McGraw-Hill, 1978.

[212] Keith Nurse. Culture as the Fourth Pillar of sustainable Development. *Small States: Economic Review and Basic Statistics*, 11:28–40, 2006.

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

[214] Leslie E. Papke and Jeffrey M. Wooldridge. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632, 1996.

[215] Oliver Parodi. The Missing Aspect of Culture in Sustainability Concepts. In J.C. Enders and M. Remig, editors, *Theories of Sustainable Development*, pages 169–187. London: Routledge, 2016.

[216] Artur Pawłowski. How Many Dimensions Does sustainable Development Have? *Sustainable Development*, 16(2):81–90, 2008.

[217] Karl Pearson. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901.

[218] Brian E. Perron and David F. Gillespie. *Key Concepts in Measurement*. New York: Oxford University Press, 2015.

[219] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A Global Averaging Method for Dynamic Time Warping, with Applications to Clustering. *Pattern Recognition*, 44(3):678–693, 2011.

[220] Domenico Piccolo. A Distance Measure for Classifying ARIMA Models. *Journal of Time Series Analysis*, 11(2):153–164, 1990.

[221] Maurizio Pisati. *SPMAP. Stata module to visualize spatial data*. Boston: Boston College Department of Economics, 2018.

[222] Michael E. Porter, Scott Stern, and Michael Green. *Social Progress Index 2014*. Washington, DC: Social Progress Imperative, 2014.

[223] Theodore M. Porter. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press, 1996.

[224] Robert Prescott-Allen. *The Wellbeing of Nations: A Country-by-country Index of Quality of Life and the Environment*. Washington, DC: Island Press, 2001.

[225] Hilary A. Priestley. Ordered Sets and Complete Lattices. In R. Backhouse, R. Crole, and J. Gibbon, editors, *Algebraic and Coalgebraic Methods in the Mathematics of Program Construction. International Summer School and Workshop, Oxford, April 10-14, 2000, revised lectures*, pages 21–78. Dordrecht: Springer, 2002.

[226] Ilya Prigogine. *Time, Structure and Fluctuations*. The Nobel Foundation: Nobel Lecture, 8 December 1977, 1977.

[227] Ilya Prigogine. *From Being to Becoming: Time and Complexity in the Physical Sciences*. San Francisco, California: W.H. Freeman, 1982.

[228] Ilya Prigogine and Isabelle Stengers. *La nuova alleanza, trad. it*. Torino: Einaudi, 1981.

[229] Ilya Prigogine, Isabelle Stengers, and with a foreword by Alvin Toffler. *Order Out of Chaos: Man's New Dialogue with Nature*. New York: Bantam Books, 1984.

[230] Lubomir S. Prytulak. Critique of S. S. Stevens' Theory of Measurement scale Classification. *Perceptual and Motor Skills*, 41(1):3–28, 1975.

[231] Sangeeta Rani and Geeta Sikka. Recent Techniques of Clustering of Time Series Data: A Survey. *International Journal of Computer Applications*, 52(15):1–9, 2012.

[232] Chotirat A. Ratanamahatana and Eamonn Keogh. Everything you know about dynamic time warping is wrong. In *Third Workshop on Mining Temporal and Sequential Data*, 2004.

[233] Giovanni Reale. *Storia della filosofia greca e romana*. Milano: Giunti, 2018.

[234] Hans Reichenbach. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: The University of Chicago Press, 1938.

[235] Nicole Richardt. Gender Gap. In William A. Jr. Darity, editor, *International Encyclopedia of the Social Sciences.*, volume 3, pages 277–279. USA: MacMillan, 2008.

[236] Andrew Rutherford. *ANOVA and ANCOVA: A GLM Approach*. Hoboken, New Jersey: John Wiley & Sons, 2011.

[237] Rocco Sacconaghi. Building Knowledge. Between Measure and Meaning: A Phenomenological Approach. In F. Maggino, editor, *Complexity in Society: From Indicators Construction to their Synthesis*, pages 51–68. Cham: Springer, 2017.

[238] Jeffrey D. Sachs. *The Age of Sustainable Development*. Columbia University: Columbia University Press, 2015.

[239] Wolfgang Sachs. *Planet Dialectics: Essays on Ecology, Equity, and the End of Development*. London: Zed Books, 1999.

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

[241] Mika Sato and Yoshiharu Sato. On a Multicriteria Fuzzy Clustering Method for 3-way Data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 02(02):127–142, 1994.

[242] Berndt Schröder. *Ordered Set. An Introduction*. Boston: Birkäuser, 2002.

[243] Erwin Schrödinger. *What is life? The Physical Aspect of the Living Cell and Mind*. Cambridge: Cambridge University Press, 1944.

[244] Lucas Seghezzo. The Five Dimensions of Sustainability. *Environmental Politics*, 18(4):539–556, 2009.

[245] Amartya Sen. *Commodities and Capabilities*. Oxford: Oxford University Press, 1999.

[246] Amartya Sen. *Development as Freedom*. New York: Anchor Books, 2000.

[247] Ashish Singhal and Dale E. Seborg. Clustering Multivariate Time-series Data. *Journal of Chemometrics*, 19(8):427–438, 2005.

[248] Luigi M. Solivetti. *Sociologia come ricerca*. Roma: Carocci editore, 2001.

[249] Isabelle Stengers. Perché non puó esserci un paradigma della complessità. In G. Bocchi and M. Ceruti, editors, *La sfida della complessità*, pages 61–83. Milano: Feltrinelli, 1985.

[250] Stanley S. Stevens. On the Theory of Scales of Measurement. *Science*, 103(2684):677–680, 1946.

[251] Joseph E. Stiglitz, Amartya Sen, and Jean P. Fitoussi. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: Internet resource, 2009. www.stiglitz-sen-fitoussi.fr.

[252] Aki Tonami and Akihisa Mori. Sustainable Development in Thailand: Lessons from Implementing Local Agenda 21 in Three Cities. *The Journal of Environment & Development*, 16(3):269–289, 2007.

[253] Howell Tong and P. Dabas. Cluster of Time Series Models: An Example. *Journal of Applied Statistics*, 17(2):187–198, 1990.

[254] Edward R. Tufte. *The Visual Display of Quantitative Information*, volume 2. Cheshire, Connecticut: Graphics Press, 2001.

[255] John R. Turner and Rose M. Baker. Complexity Theory: An Overview with Potential Applications for the Social Sciences. *Systems*, 7(1):4, 2019.

[256] Robert K. Turner, David W. Pearce, and Ian Bateman. Environmental Economics: An Elementary Introduction, 1993.

[257] UNCED. United Nations Conference on Environment and Development. *Rio Declaration on Environment and Development*, 1992.

[258] UNGA. United Nations General Assembly. *In Larger Freedom: Towards Development, Security and Human Rights for All. Report of the Secretary-General*, 2005. UN doc. A/59, 2005.

[259] UNGA. United Nations General Assembly. *The Future We Want* (*Resolution Adopted by the General Assembly on 27 July 2012*), 2012.

[260] UNGA. United Nations General Assembly, 70 Session. *Transforming our World: The 2030 Agenda for Sustainable Development*, 2015.

[261] Arthur Upgren. *Weather: How it Works and why it Matters*. Reading, MA : Basic Books, 2008.

[262] Michalis Vazirgiannis. Clustering Validity. In L. Liu and M.T. Özsu, editors, *Encyclopedia of Database Systems*, pages 388–393. Boston, MA: Springer, 2009.

[263] Paul F. Velleman and Leland Wilkinson. Nominal, Ordinal, Interval, and Ratio Typologies are Misleading. *The American Statistician*, 47(1):65–72, 1993.

[264] Mitchell M. Waldrop. *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York: Simon and Schuster, 1992.

[265] Weina Wang and Yunjie Zhang. On Fuzzy Cluster Validity Indices. *Fuzzy Sets and Systems*, 158(19):2095–2117, 2007.

[266] WCED. United Nations World Commission on Environment and Development. *Our Common Future. The Brundtland Report*, 1987.

[267] WEF. World Economic Forum. *Global Gender Gap*, 2018. Geneva: World Economic Forum.

[268] Taiyun Wei and Viliam Simko. corrplot: Visualization of a correlation matrix. *R package version 0.73*, 230(231):11, 2013.

[269] Sooksiri Wichaisri and Apichat Sopadang. Trends and Future Directions in Sustainable Development. *Sustainable Development*, 26(1):1–17, 2018.

[270] Hadley Wickham. *ggplot2: Elegant graphics for data analysis*. Cham: Springer, 2016.

[271] James B. Wilcox, Roy D. Howell, and Einar Breivik. Questions about Formative Measurement. *Journal of Business Research*, 61(12):1219–1228, 2008.

[272] Xuanli L. Xie and Gerardo Beni. A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(8):841–847, 1991.

[273] Yimin Xiong and Dit-Yan Yeung. Time Series Clustering with ARMA Mixtures. *Pattern Recognition*, 37(8):1675–1689, 2004.

[274] Lotfi A. Zadeh. Fuzzy Sets. *Information and Control*, 8(3):338–353, 1965.

[275] Wolfgang Zapf. Systems of Social Indicators: Current Approaches and Problems. *International Social Science Journal*, 27(3):479–498, 1975.

[276] Richard A. Zeller and Edward G. Carmines. *Measurement in the social Sciences. The Link between Theory and Data*. Cambridge: Cambridge University Press, 1980.

Collana Studi e Ricerche

Per informazioni sui volumi precedenti della collana, consultare il sito: www.editricesapienza.it | *For information on the previous volumes included in the series, please visit the following website: www.editricesapienza.it*

118. Tubulin and Microtubules as Drug Targets for Potential Cancer Chemotherapy and CNS-Directed Therapies
Ludovica Monti
119. Simulations of RF Beam Manipulations Including Intensity Effects for CERN PSB and SPS Upgrades
Danilo Quartullo
120. Multi-drug resistant Klebsiella pneumoniae strains circulating in hospital setting
Whole-genome sequencing and Bayesian phylogenetic analysis for outbreak investigations
Eleonora Cella
121. Agave negatively regulates YAP and TAZ transcriptionally and post-translationally in osteosarcoma cell lines
A promising strategy for Osteosarcoma treatment
Maria Ferraiuolo
122. Trigeminal Neuralgia
From clinical characteristics to pathophysiological mechanisms
Giulia Di Stefano
123. Le geometrie del Castello di Anet
Il ‘pensiero’ stereotomico di Philibert de l’Orme
Antonio Calandriello
124. Towards Recognizing New Semantic Concepts in New Visual Domains
Massimiliano Mancini
125. La distribuzione spaziale dei reperti come base per un’interpretazione dei livelli subappenninici di Coppa Nevigata (Manfredonia, FG) in termini di aree di attività
Enrico Lucci
126. Costruire, violare, placare: riti di fondazione, espiazione, dismissione tra fonti storiche e archeologia
Attestazioni a Roma e nel *Latium Vetus* dall’VIII a.C. al I d.C.
Silvia Stassi
127. Complexity of Social Phenomena
Measurements, Analysis, Representations and Synthesis
Leonardo Salvatore Alaimo
128. Etica ebraica e spirito del capitalismo in Werner Sombart
Ilaria Iannuzzi

Studi e Ricerche

▶ Scienze sociali

What is complexity? What are the characteristics of a complex system? What does it mean to measure? How to develop an efficient tool for measuring socio-economic phenomena? What is synthesis? What are the main statistical tools for synthesis of multi-indicator systems? These are the research questions that are attempted to be answered in this book, the result of the author's research work during his PhD. The book offers a conceptual and methodological analysis of the topic of synthesis of complex social phenomena, also proposing interesting applications to real cases.

Leonardo Salvatore Alaimo is Research Fellow of Social Statistics (SECS-S/05) at the Department of Social Sciences and Economics of the University of Rome La Sapienza. He holds a Ph.D. in Social Statistics at the University of Rome La Sapienza in 2020. He took part in many international conferences and he is the author of several papers published in a wide range of international peer-reviewed journals. His research interests are mainly the measurement of complex phenomena, the synthesis of systems of statistical indicators, multi-level models, composite indicators, the multivariate statistical methods, sustainability, quality of life, and well-being.

ISBN 978-88-9377-251-8

9 788893 772518

Winner of the Competition "Prize for PhD Thesis 2020"
arranged by Sapienza University Press.

www.editricesapienza.it