# A new methodology for multidimensional poverty measurement based on the capability approach

Vivien Kana Zeumo[a,b,*], Alexis Tsoukiàs[c,*], Blaise Somé[b]

[a] *LAMSADE, Université Paris Dauphine, F-75016 Paris, France*
[b] *LANIBIO, Université de Ouagadougou, 03 B.P. 7021 Ouagadougou 03, Burkina Faso*
[c] *CNRS – LAMSADE, Université Paris Dauphine, F-75016 Paris, France*

**Abstract**

We propose a methodology based on the use of clustering techniques derived from data analysis and multi-attribute decision analysis methods aiming at purposeful multidimensional poverty measurement. Our contribution to methodological knowledge insists on the necessity to build *"meaningful measurements"* for policy making and policy implementation. Our standpoint underlines the necessity to consider the problem of poverty measurement as a decision problem and to tackle its measurement issue with that in mind. We also show that such an exercise can be useful to develop a better operational definition of poverty and to solve the aggregation issues.

*Keywords:* meaningful measurement, capabilities approach, policy making, decision aiding.

## 1. Introduction

The review of the literature on poverty measurement (see Kana et al., 2011) allows us to conclude that measuring poverty is not a representation of an objective situation, it is rather an instrument for pursuing a policy. People may feel poor and not be identified as such. People may be identified as poor and not feel as such. Indeed, poverty is an evolutive, multidimensional, fuzzy and non-objective situation which does not contain anything of numerical, but only the sensation of those who are suffering. We are more or less poor and in many different ways.

Many authors (see Nussbaum, 1987, 2000; Fusco, 2005; Alkire, 2005; Bertin, 2007; Kana et al., 2011) agree that Sen's capability approach (see Sen, 1985) is appropriated as tool aiming at assessing how welfare is distributed among a given population. The reason is that allows to highlight the diversity of relationships between people and goods (commodities), the complex relationships of individuals between themselves (social relations) and of individuals with their environment (institutions, norms, cultures). The strong argument for

*Corresponding author. Tel.: +33 1 44 05 44 01; Fax: +33 1 44 05 40 91
*Email addresses:* `vivien.kana-zeumo@dauphine.fr` (Vivien Kana Zeumo), `tsoukias@lamsade.dauphine.fr` (Alexis Tsoukiàs), `some@univ-ouaga.bf` (Blaise Somé)

the capability approach is based on the postulate that commodities (goods or services) are insufficient to evaluate and describe in a faithful way, the welfare of people. As an example, two people can aspire to different things in terms of welfare, while owning the same resources equivalent to, let's say, \$3500 U.S. This is why Sen (1985) introduces a broad distinction between a person's interests and their fulfilment, respectively called "well-being" and "advantage". Sen argued that *"well-being is concerned with a person's achievement: how 'well' is his or her 'being'? 'Advantage' refers to the real opportunities that the person has, especially compared with others"*. This postulate considers the commodities a mean for improving the quality of life of individuals and advocates to focus on how these individuals will use their resources. This led Sen to develop a broad discussion about the distinctions between commodities, characteristics, functionings and capabilities.

Sen's capabilities approach allows to take into account the notion of freedom that has a person to achieve a certain level of well-being and the assumption of human diversity in the process of poverty measurement. Therefore, while trying to measure poverty we need to take into account several different dimensions of uncertainty. We must select and validate the space of functionings that individuals are able to "do" (doing) or aspire to "be" (being) through their commodities and their characteristics. The choice and validation of the space of functionings can be done in an efficient and realistic way only within a decision aiding setting. This paper shows how we can process the information that is required to implement the capability approach in a way useful for policy design, policy implementation and the assessment of poverty reduction initiatives. We present a new methodology which operationalises Sen's capabilities approach through the development of meaningful multidimensional poverty measurements. The issue of meaningfulness is thus analysed both from a theoretical point of view (measurement theory) and from an operational one (policy effectiveness). The general methodology shown in Figure (1) outlines the different stages allowing to derive the meaningful measurements. The stage concerning supervised learning (schematised as the part appearing in Figure (1) with *"double line"*) will not be developed in this paper, but in a forthcoming one. The paper is organised as follows: Section 2 introduces basic notions concerning our motivations, the concept of meaningful measurement, the concept of capability and the setting. In Section 3 we show how to construct meaningful measurements, while in Section 4 we show how to translate meaningful measurements into concrete actions in terms of policies, programmes and projects for implementation. Concluding comments are given in Section 5 and Section 6 presents an application of our methodology on ASSL 2007 database of Burkina Faso.

## 2. Basics

### *2.1. Our motivations*

Consider a given *client* or *decision maker* with an agenda of poverty alleviation including a certain number of policies that he should like to undertake in a given region of world. This client can be represented, for instance, by the World Bank, the European Union, the UNDP[1],

[1]UNDP: United Nations Development Programme.

the WHO[2] or the NEPAD[3] with a specific poverty reduction policy aiming to support specific categories of citizens through precise actions such as facilitated access to credit, land re-distribution, water supply enhancement programmes, health research programmes, education aid programmes. Note that "precise actions" are a set of actions that our client would like to undertake in a given region of world in order to improve the standard of living of people. In another sense, "precise actions" refer to specific operations targeted on issues to be addressed. Our client is faced to several major problems:

- Know what the situation is and measure it: There are different types of poverty which imply different perspectives between policy maker and subjects. Income is not always representative and the cutting off thresholds are arguable. Measuring poverty has to be an instrument of pursuing a policy. Hence, in order to design interventions best adapted to a given reality, we firstly need to understand the factors and causes determining the present situation. This calls at replying to questions of the type: which elements describe better the specific conditions of the observed population with respect to the precise policy to be pursued? Which elements better characterize the perception of the interested population as members of a specific category? How can we measure it? People being differently poor, how can we construct measurements reflecting different categories of poverty?

- Dealing with different poverties: It is misleading to talk only about "poor" and "not poor", at least as far as a multidimensional perspective of poverty is considered. What we observe in reality are different types of poverty. Various different, though related, questions can be asked: What is the underlying problem that has to be dealt with in priority? What specific objectives are to be pursued in confronting these different poverties? Who are eligible for some policy measure? Who is expected to benefit from such policies? How they should benefit? Is that specific policy efficient? Is this specific policy appropriate for the target group? What is the cost for implementing such a policy? Why? What does it mean fighting poverty?

- Dealing with several different dimensions of uncertainty: Mostly, poverty databases are very large and are formed by mixed variables. Then, the heterogeneous information has to be considered. The challenge consists to identify undiscovered groupings of individuals and establish hidden relationships between them. It is therefore an operation aimed at extracting relevant information from data. This calls at replying to questions of the type: which information is readily available and relevant? Is it useful in order to draw rational conclusions and recommendations? How easy is to assess the missing information?

- Predicting the consequences and valuing the outcomes: Sometimes, policies can be unsuccessful and ineffective without any positive impact in the medium or

[2]WHO: World Health Organization.
[3]NEPAD: New Partnership for Africa's Development.

long-term. This can be due to several reasons such as uncertainties or missing information. Since a policy is considered as a set of actions (or alternatives) that our client would like to undertake in a given region, it is crucial to explore all alternatives of each policy in order to analyse the consequences of the various possible policies which have to be pursued in order to improve the living conditions of households. This leads to assess the effectiveness of various possible policies by putting the best evidence at the heart of research such as to determine whether a particular policy will produce a positive impact on welfare of people in the future when that policy will be translated into specific operations. This involves replying to questions of the type: What are the potential alternatives? What are their consequences? If outcomes are undeterminable or uncertain, what can we decide about a compromise?

We consider the decision aiding process as a distributed cognition process. Nevertheless, our point of view is operational and not cognitive. We make the hypothesis that the participants actively try to create a shared representation. Thus, through these questions we try to analyse the artifacts such a process generates: a representation of the problem situation, a problem formulation, an evaluation model, a final recommendation (Tsoukiàs, 2007, 2008).

To sum up, considering this problem situation, an analyst (or policy maker), a population and some knowledge about it, we are looking for understanding: How to identify the different types of poverty? What are the population needs? Once we have classified the population into different classes of households (which are in reality the target of some policy), how may we derive adequate policies to help "poor people" getting out this situation? How to design and identify potential alternatives or actions such as to highlight preferable alternatives which are *more* important with respect to certain objectives and *less* important with respect to others? How to decide to which alternative policy each individual/household has to be subjected? How to monitor and assess such policies?

To tackle this problem situation, in a way allowing to provide answers to such questions, we have introduced in this paper the concept of *meaningful multidimensional poverty measurement* (MDPM) by combining the capability approach with decision aiding methodology (Tsoukiàs, 2008). This paper outlines (see Figure (1)) the two type of analysis that we have to run in order to derive the meaningful measurements:

- Data driven analysis: Using and evaluating data are important steps to the poverty measurement improvement process. Data are any information about poverty that can be gathered, reviewed and analyzed in order to produce a useful knowledge. Data driven analysis means that decision making in this stage is compelled by data, rather than by intuition or personal experience. Thus, it is the process of identifying target groups based on appropriate analysis of relevant information. It is also the process of constructing meaningful measurements useful for elaborating final recommendation.

- Client driven analysis: a client defines the preferential information and the goals that he/she wishes to prioritize. Analyst will assist client in obtaining the information they need to make well-informed decisions. He will allow client to meet his own needs, to advocate for himself and to access services directly whenever possible. At the client's request, stakeholders will undertake direct advocacy on behalf of the client. Analyst will encourage and invite other actors to collaborate around the well being of the client. Thus, all subjective preferential information is "client driven" which implies interacting with the client (under different form).

*2.2. What is a meaningful measurement?*

A rather complete definition of the term 'measurement' has been given by Mari (2003) who argued that *"measurement is a specific kind of evaluation, i.e. it is an operation aimed at associating an information entity, the result of measurement, with the state of the system under measurement in reference to a given quantity, the measurand"*. We believe that a measurement of poverty should be considered as a set of operations allowing to build a *bridge* (field of subjective human experiences) between the *physical world* (field of physical things) and the *informational world* (field of objective knowledge). The concept of meaningfulness comes from measurement theory (see Suppes, 1959; Krantz et al., 1971). Fred Roberts (1979) presented 'meaningfulness' as an essential condition for a measurement to be well-defined in the meaning of *correctness*, *completeness* and *rationality.* Roberts's standpoint is clearly in the same line of definition given by Stevens (1946) according to which *"measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules"*.

However, in the case of poverty, a measurement is not only performed in order to assign numbers to individuals or households, but it has to help decision makers make well-informed decisions about a particular policy in such a way to design and identify the preferable alternatives with respect to same complex policy issues. It is a decision aiding process run in a suitable way aiming at selecting appropriate policies, laying out the alternatives (or actions), predicting the consequences and valuing the negative and positive outcomes when that policy is being translated into concrete actions, i.e. in terms of strategies, programmes and projects for implementation. This standpoint imposes to define the framework of decision aiding and implies two essential requirements that a poverty measurement has to satisfy beyond measurement meaningfulness: *operationality* and *legitimation.* A poverty measurement is *operational* if it can be used efficiently to recognise actors drawn from some universe it denotes and if it can help decision makers to elaborate well-informed interventions. Otherwise, a poverty measurement is *legitimate* if it takes into account how a final recommendation is presented, implemented and perceived by the other actors besides its precise contents. Note that operationality and legitimation have not been defined explicitly in this paper, so the definitions given here are based on our retrospective analysis and reconstruction (see Bouyssou et al., 2000; Tsoukiàs, 2007).

Therefore, in the field of poverty or welfare, a measurement is *"meaningful"* if it complies to three conditions:

**Theoretical soundness:** poverty measurement needs to be *theoretically sound*, in the sense that the concepts used to construct it are in adequacy with measurement theory;

**Operational completeness:** poverty measurement needs to be *operationally complete*, in the sense that it is useful for policy making, policy implementation and it helps decision makers to make well-informed decisions about policies, programmes or projects.

**Legitimation:** poverty measurement needs to be *legitimated* in the sense that, it should address the intuitive expectation the society and the stakeholders have about what to do for poverty reduction.

**Remark 2.1.** *Operational completeness does not mean addressing all policies, but all operational issues a policy may entail. In other terms a measurement should give sufficient argument to influent a policy.*

A meaningful multidimensional poverty measurement (MDPM) can be defined as follows:

**Definition 2.1.** *A* MDPM *is a meaningful measurement derived from a decision aiding process aiming to improve people's capabilities and their living standards.*

Note that, our postulate concerning the MDPMs is in adequation with the following three positions and Sen's capability approach sketched at the standpoint of its operationalization:

(P1): Measurements are inherent properties of the measured things (see Mari, 2003)

(P2): Measurements are results of operations that preserve the relations observed among measured things (see Mari, 2003; Roberts, 1979)

(P3): Measurements are results of a decision aiding process (see Bouyssou et al., 2000; Tsoukiàs, 2007)

### *2.3. Endowments, commodities, functionings and capabilities*

The origins of the capabilities approach can be found in a series of papers criticising traditional welfare economics, written by Sen in the early 1980s (see Sen, 1976, 1977, 1979, 1981, 1985, 1993) where he developed the concepts of *endowment*, *commodity*, *functioning*, and *capability* for assessing the well-being of individuals. As stated by Alkire (2005), *"the goal of both human development and poverty reduction should be to expand the capability that people have to enjoy 'valuable beings and doings'. They should have access to the positive resources they need in order to have these capabilities. And they should be able to make choices that matter to them*. The capabilities approach appears as the most influential recent attempt for valuing a person's achievement (*'well-being'*), the real opportunities that this person has (*'advantage'*) and the quantity of *'happiness'* generated in this person's life. These purposes are tackled in this paper through four core concepts introduced in Sen's capability approach: endowments, commodities, functionings, capabilities.

Initially, each individual *legally* owns a combination of resources (called *'endowments set'* or more simply *'endowments'*). These *'resources'* include both tangible assets such as land, equipment, animals, etc., and intangibles assets such as natural talents or qualities, labour power, physical abilities, knowledge, skills, etc. Thus, each individual can use the resources of his endowment set to *produce* and to *legally* obtain the set of all possible combinations of goods and services that are exchanged within the society. We call this *'commodities*

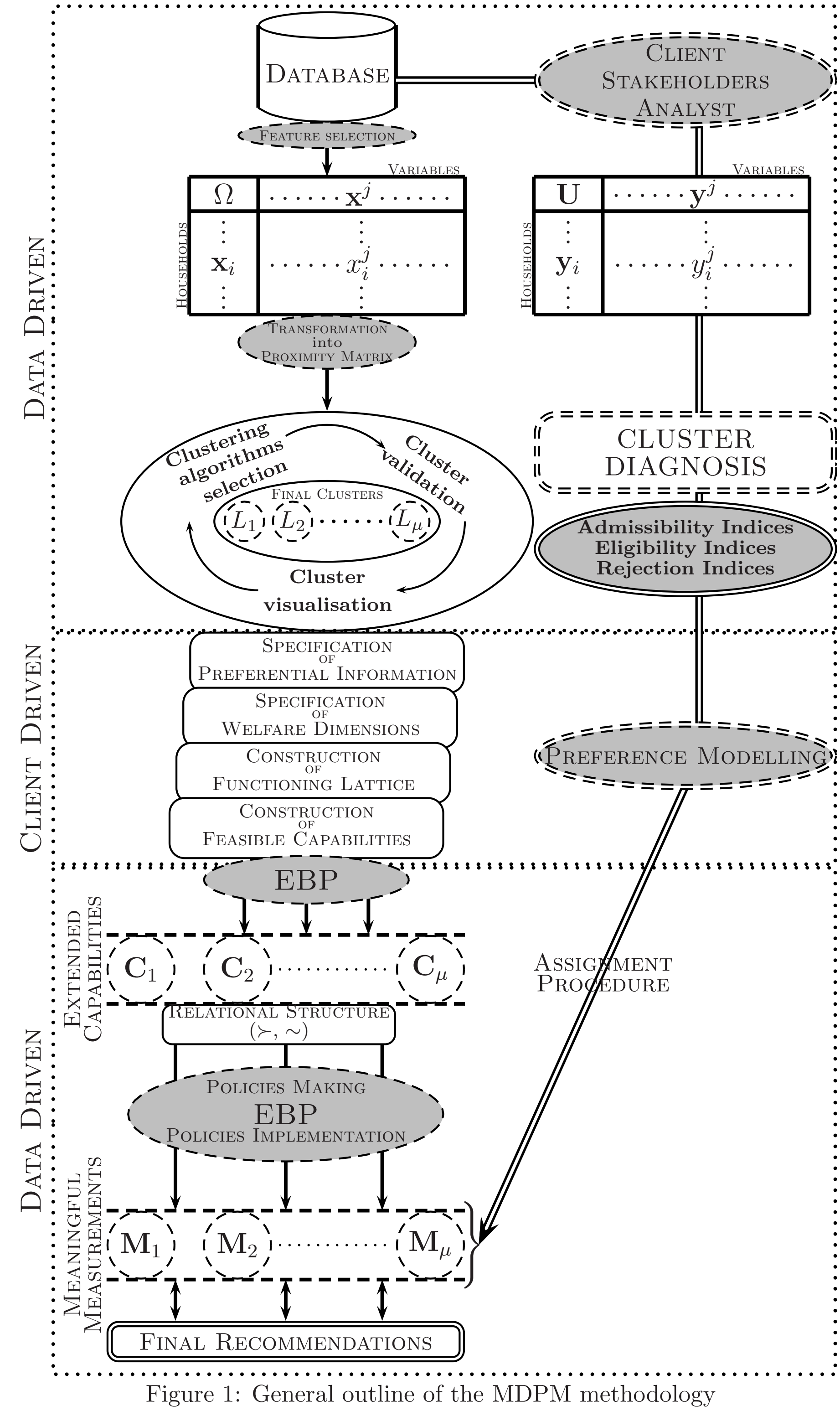


Figure 1: General outline of the MDPM methodology

*set'* or more simply *'commodities'*. Note that, depending on his tastes and preferences, an individual can choose to enjoy only one of such possible combinations. For example[4],

[4]This example taken from Osmani's (1993) research paper is a perfect illustration of how the endowments

a farmer may use his land, labour, and other resources to produce the food he wants; a labourer may *'exchange'* his labour power to secure his food; a fisherman may first use his labour, equipment and fishing boat to *'produce'* a catch of fish and then exchange it to get the rice he wants; an unemployed person may use his resource of 'citizenship of a welfare state' to claim a *'transfer'* of state funds in the form of unemployment benefit. In general, the set of exchange resources can be seen as a mapping from a given person's endowment vectors (the set of all possible endowments set) to availability sets of commodity vectors (the set of all possible commodities set). Many authors (see Sen, 1976, 1977; Osmani, 1993) have proposed several procedures for valuing exchange resources. For example, Osmani (1993) suggests to consider: *"for the farmer, the input-output ratios in farm production; for the labourer, the ratio between money wage and the price of food i.e., the real wage rate; for the fisherman, both the input-output ratio in fishing and the relative price of fish and rice; and for the unemployed person, the rate of unemployment benefit"*.

The commodities aren't necessarily translated into well-being but they are rather conceptualised in terms of a person's characteristics. For example, the possession of a vehicle allows the owner to benefit of all properties of the vehicle, which can be used to satisfy mobility, to transport goods, to obtain happiness from travelling and for rental purpose in order to generate income. As mentioned by Sen (1985), the *'characteristics of commodities'* don't tell us what an individual will do with such commodities. For example, a person can possess a vehicle and mechanical skills but, be "poorly transported" because there is no gas station in his village. This latter standpoint led Sen to consider the *'functionings'* of individuals. A functioning is the achievement of an individual i.e. what he succeeds in *doing* and *being* with his commodities and their characteristics given his personal characteristics as well as the social and environmental factors beyond his control. For example, owning a vehicle (commodity), considering the characteristic *'mobility'*, some individuals may achieve the functioning *'adequately transported'*, while others, considering the same commodity and the same characteristic (but without gas station in their village) may achieve the functioning *'poorly transported'*. Note that, a functioning has to be distinguished from the commodities used in order to achieve that precise functioning. For instance, to *'be adequately transported'* must be distinguished from the fact to own a vehicle.

The totality of all the alternative functioning vectors the person can choose from, given the contingent circumstances, reflects the person's capabilities (see Sen, 1985). A *'capability'* reflects a person's *ability* to achieve, through choice, a given functioning among the various alternative functioning bundles. For example, a person may have the ability to buy a vehicle and to have the capability to be *'well transported'*, but he may choose not to do so and continue to get to work by foot or by bike.

The following diagram shows the relationship among the four basic concepts:

$$\textsc{Endowments} \longrightarrow \textsc{Commodities} \longrightarrow \textsc{Functionings} \longrightarrow \textsc{Capabilities}$$

In order to well understand the remaining sections, we introduce the following definitions:

can be transformed into commodities.

**Definition 2.2.** *A* household *is a basic socio-economic unit in which different people, related or not, are living in the same house or concession under the authority of a person named chief of household, putting their commodities and their characteristics together to improve their abilities in* doing *and* being*.*

**Definition 2.3.** *A* cluster *is a set of households which are* similar *or* alike *in term of distribution of commodities.*

**Definition 2.4.** *A* variable *is a characteristic allowing to describe each household.*

**Definition 2.5.** *A* commodity *is a good or service that a household declares to legally attain with his endowments.*

*2.4. The setting*

The basic unit of analysis is a household. However, in practice, our methodology can be conducted both at an individual level such as a person, and at a collective level such as a group or class using the standard device of assuming a *'representative individual'*. Initially, each household $\mathbf{x}_i$ is endowed by a set of resources. These resources are transformed in commodities which can be exchanged within the society. In this paper, we will not deal with the question aiming to know how the endowments are transformed in commodities. We suppose that the input data table represents the set of commodity vectors of households which was determined from a 'Household Living Standards Survey'. Thus, a household's commodities vector can be defined as being the set of all goods and services that he declares to legally attain with his endowments.

We consider a *multidimensional distribution* for a population $\Omega$ of $n$ households on a $n \times m$ data table (1).

$$\Omega = \begin{array}{c} \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_n \end{array} \begin{array}{c} \begin{array}{ccccc} X_1 & \dots & X_j & \dots & X_m \end{array} \\ \begin{bmatrix} x_1^1 & \cdots & x_1^j & \cdots & x_1^m \\ \vdots & & \vdots & & \vdots \\ x_i^1 & \dots & x_i^j & \cdots & x_i^m \\ \vdots & & \vdots & & \vdots \\ x_n^1 & \dots & x_n^j & \cdots & x_n^m \end{bmatrix} \end{array} \tag{1}$$

where $x_i^j$ is the $i$th commodity of household in the $j$th variable, $X = X_1 \times \cdots \times X_m$ and $\Omega \subset X$. The interpretation is as follows: we have households $\mathbf{x}_i$, $i \in \{1, 2, \dots, n\}$, evaluated on a set $\mathcal{J} = \{1, 2, \dots, m\}$ of variables. The set $X_j$ gathers all possible levels that a household can possibly take on the $j$th variable ($j \in \mathcal{J}$). The set $X_j$ is the $j$th set of evaluations on the $n$ households. The household's commodities vector $\mathbf{x}_i$ is the set of evaluations of the $i$th household on the $m$ variables and $\mathbf{x}_i \in X$. The set $X$ is the set of all possible commodities vectors of evaluations on the $m$ variables.

As is customary in data analysis, we consider two characteristics of data: *data type* and *data scale*. *Data type* refers to the degree of quantization in the data and *data scale* indicates the relative significance of numbers. The input data can be typed as *binary* (e.g. "yes-no"), as *discrete* and as *continuous*. The input data can also be scaled as *qualitative* (nominal and ordinal) scales and *quantitative* (interval and ratio) scales. The possible values of a qualitative (nominal or discrete) variable are called the *"modalities of the variable"*.

*2.5. Clustering*

Clustering (see Diday et al., 1982; Jain and Dubes, 1988; Berkhin, 2002, for more details) is used in order to organise data into clusters in such a way that each cluster consists of households that are *similar* in term of commodities distribution between themselves and *dissimilar* to households of other clusters. Initially, we have to select properly the features (*selection of individuals* and *selection of variables*) on which clustering is to be performed so as to encode as much information as possible concerning the task of our interest. Depending on the structure of the data table, it may be necessary to standardize some variables before computing the proximity matrix. Sometimes, standardization of variables is useful to delete the effects of origin and scale in the measurement of the variables (see Anderberg, 1973; Milligan and Cooper, 1988, for more details).

Clustering methods require that a measure of proximity (alikeness or affinity) be established between pairs of households. Thus, this step aims to transform the data table into a *proximity matrix*. A proximity matrix $[\delta(\mathbf{x}_i, \mathbf{x}_k)]$ accumulates the pairwise indices of similarity (or indices of dissimilarity) in a matrix in which each row and column represents a household. The more the $i$th and $k$th household resemble one another, the smaller a dissimilarity index, the larger a similarity index. We assume that an index of proximity is a measure of the dissimilarity $\delta(\mathbf{x}_i, \mathbf{x}_k)$ between the $i$th and $k$th household if it satisfies the following three conditions for all $\mathbf{x}_i,\ \mathbf{x}_k \in \Omega$:

$$(i)\ \delta(\mathbf{x}_i, \mathbf{x}_k) \geq 0; \quad (ii)\ \delta(\mathbf{x}_i, \mathbf{x}_i) = 0; \quad (iii)\ \delta(\mathbf{x}_i, \mathbf{x}_k) = \delta(\mathbf{x}_k, \mathbf{x}_i). \tag{2}$$

The third condition rules out asymmetric indices of proximity and, taken in conjunction with (*ii*), implies that a proximity matrix is fully specified by providing the $n(n-1)/2$ values in its lower triangle.

In this paper, we have chosen a hierarchical clustering method based on Ward's method, also known as the *minimum variance* method. It is one of the most widespread hierarchical clustering methods which is distinct from all other ones because it uses an analysis of variance to evaluate the distances between clusters. In short, this method attempts to minimize the Sum of Squares of any two (hypothetical) clusters that can be formed at each step (see Ward, 1963, for more details concerning this method). In general, this method is regarded as very efficient, however, it tends to create clusters of small size.

## 3. Construction of meaningful measurements

*3.1. Modal-valued matrix and modal-valued criterion*

Cluster description aims to transform the large data table into a summary table in order to gain initial knowledge. Such exercise allows to identify the *'relative importance'* of a variable within a given cluster. For this purpose, we introduce two core concepts: *modal-valued matrix* and *modal-valued criterion*. The *modal-valued matrix* can be considered as a summary table of the large data table where the $j$th column represents the variable $X_j$, $j \in \mathcal{J}$ and the $h$th row denotes the cluster $h \in \{1, \ldots, \mu\}$. The intersection of the $h$th row and the $j$th column denotes the *modal-valued criterion* which is the description of the variable $X_j$ within the cluster $L_h$.

Formally, let $\mathbf{L} = \{L_1, \ldots, L_\mu\}$ be the set of clusters obtained after clustering the population $\Omega$. We consider a multidimensional cluster distribution for a population $\Omega$ of $n$ households with $\mu \times m$ *modal-valued matrix*:

$$\mathbf{X} \equiv [\mathbf{X}^1, \ldots, \mathbf{X}^j, \ldots, \mathbf{X}^m] = \begin{bmatrix} X_1(L_1) & \cdots & X_j(L_1) & \cdots & X_m(L_1) \\ \vdots & & \vdots & & \vdots \\ X_1(L_h) & \ldots & X_j(L_h) & \cdots & X_m(L_h) \\ \vdots & & \vdots & & \vdots \\ X_1(L_\mu) & \ldots & X_j(L_\mu) & \cdots & X_m(L_\mu) \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_h \\ \vdots \\ \mathbf{X}_\mu \end{bmatrix} \quad (3)$$

$X_j(L_h)$ denotes the *modal-valued criterion* of cluster $h \in \{1, 2, \ldots, \mu\}$ in variable $j \in \mathcal{J}$ defined as follows:

$$X_j(L_h) \;=\; \langle(\beta_{j1}, \pi^h_{j1}); \cdots ; (\beta_{js_j}, \pi^h_{js_j})\rangle \quad (4)$$

where $\{\beta_{jk} \,:\, k = 1, \ldots, s_j\}$ is a set of modalities (or states) of $L_h$ over the domain $D_j$ of $X_j$, $j = 1, \ldots, m$; $\pi^h_{jk}$ is a non-negative measure associated with $\beta_{jk}$ using equation (5) and $s_j$ is the number of values actually taken by $D_j$ of $X_j$.

$$\pi^h_{jk} = \frac{|\{\mathbf{x}_i \in L_h \,:\, x^j_i = \beta_{jk}\}|}{|\{L_h\}|}, \quad k = 1, \ldots, s_j \quad (5)$$

Note that the modality $\beta_{jk}$ can be finite or infinite in number, quantitative or categorical in value; and the measure $\pi^h_{jk}$ represents the probability, the proportion or the frequency of value $\beta_{jk}$ within cluster $L_h$. Intuitively, $\pi^h_{jk}$ represents the *'relative importance'* of the modality $\beta_{jk}$ within the cluster $L_h$ i.e. the number of times that a modality $\beta_{jk}$ occurs within the variable $X_j$ of cluster $L_h$.

**Definition 3.1.** *A* modal-valued criterion $X_j(L_h)$ *is a fuzzy subset of the set of modalities (or states)* $\{\beta_{jk} \,:\, k = 1, \ldots, s_j\}$ *of a cluster* $L_h$ *over the domain* $D_j$ *of* $X_j$, $j \in \mathcal{J}$ *defined by:*

$$X_j(L_h) \;=\; \left\{(\beta_{jk}, \pi^h_{jk}) \,:\, k = 1, \ldots, s_j\right\} \quad (6)$$

*where* $\pi^h_{jk}$ *is the weight of the modality* $\beta_{jk}$ *on variable* $j \in \mathcal{J}$ *associated with cluster* $L_h$, $h \in \{1 \ldots, \mu\}$.

The modal-valued matrix of $X$, written $\mathbf{X}$, is the set of all subsets of $X$ such as $\mathbf{X} = \mathbf{X}^1 \times \cdots \times \mathbf{X}^m$ and $\mathbf{X}^j = \{\beta_{jk} \,:\, k = 1, \ldots, s_j\}$ denotes the subset of modalities (or states) over the domain $D_j$ of $X_j$. In the particular case where $X_j \subseteq \mathbb{R}$, we set $\mathbf{X}^j \subseteq \mathbb{R}$. $\mathbf{X}_h$, called '*modal-valued cluster*', represents the *description of cluster* $L_h$ on the modal-valued matrix $\mathbf{X}$ given by equation (7):

$$\mathbf{X}_h \;=\; \langle X_1(L_h); \cdots ; X_m(L_h)\rangle \quad (7)$$

*3.2. An illustrative example*

Let us consider for instance the data table (see Table 2) obtained from a survey of standards of living of households. For this example, we consider 12 households $\Omega = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}, \mathbf{x}_{11}, \mathbf{x}_{12}\}$ evaluated on 7 variables (one quantitative variable and 6 qualitative variables):

| $X_j$ | Description |
|---|---|
| $X_1$ | Monthly income for basic needs of households in euros. |
| $X_2$ | Is there a room equipped for cooking? Yes(Y), No(N). |
| $X_3$ | Type of housing? Villa(V), Apartment Building(A), Single Individual House(S). |
| $X_4$ | Owner of a vehicle, car or lorry? Yes(Y), No(N). |
| $X_5$ | Duration for reaching the nearest public transport (in minutes)?<br>$[0; 14]$ =Very Close(V), $[15; 29]$ =Acceptably Close(A), $[45; 59]$ =Far(F). |
| $X_6$ | Has had problems to meet food needs? Never(N), Sometimes(S), Always(A). |
| $X_7$ | Economic situation? Worse Now(W), Unchanged(U), Better Now(Be). |

Table 1: Description of variables

| | Variables | | | | | | |
|---|---|---|---|---|---|---|---|
| Households | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ |
| $\mathbf{x}_1$ | 250 | N | S | N | F | A | W |
| $\mathbf{x}_2$ | 4500 | Y | V | Y | V | N | Be |
| $\mathbf{x}_3$ | 1500 | Y | A | Y | F | S | U |
| $\mathbf{x}_4$ | 200 | N | S | N | F | A | W |
| $\mathbf{x}_5$ | 800 | N | S | N | F | S | W |
| $\mathbf{x}_6$ | 5000 | Y | V | Y | V | N | Be |
| $\mathbf{x}_7$ | 2500 | Y | A | Y | A | S | U |
| $\mathbf{x}_8$ | 600 | N | S | N | F | A | W |
| $\mathbf{x}_9$ | 2000 | Y | A | Y | A | S | Be |
| $\mathbf{x}_{10}$ | 6500 | Y | V | Y | V | N | U |
| $\mathbf{x}_{11}$ | 1000 | N | A | N | F | A | W |
| $\mathbf{x}_{12}$ | 2800 | Y | A | Y | A | S | Be |

Table 2: Illustrative example

We conducted hierarchical cluster analysis via multiscale bootstrap (number of bootstrap 1000; see Suzuki and Shimodaira, 2006)) using Ward's method (Ward, 1963) and correlation-based dissimilarity matrix. The best number of clusters is three. This result may be confirmed by Calinski and Harabasz's (1974) index which allows to compare the homogeneity of partitions. The visualisation of clusters is possible through the dendrogram of households and multidimensional scaling as shown on Figure 2. The weight values at the vertical axis represent a set of 11 real values (non-decreasing for ultrametric trees). For Ward's method, they are the values of Ward's minimum variance criterion (which minimizes the total within-cluster variance). The points $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$, $\mathbf{x}_5$, $\mathbf{x}_6$, $\mathbf{x}_7$, $\mathbf{x}_8$, $\mathbf{x}_9$, $\mathbf{x}_{10}$, $\mathbf{x}_{11}$, $\mathbf{x}_{12}$ at the horizontal axis represent the households. The best number of clusters is three and is represented by three 'red' block (Figure 2:(a)). Figure 2:(b) shows a two-dimensional representation of the 12 households (see Table 2) provided by non-metric multidimensional scaling for which the stress takes the value 0.009443 (zero is the perfect score). The Coordinate 1 and Coordinate 2 represent the sample scores and correspond to the non-metric multidimensional scaling axes. This chooses a $k$-dimensional ($k = 3$) configuration to minimize the stress, the square root of the ratio of the sum of squared differences between the input distances and those of the configuration to the sum of configuration distances squared (see Gordon, 1999).

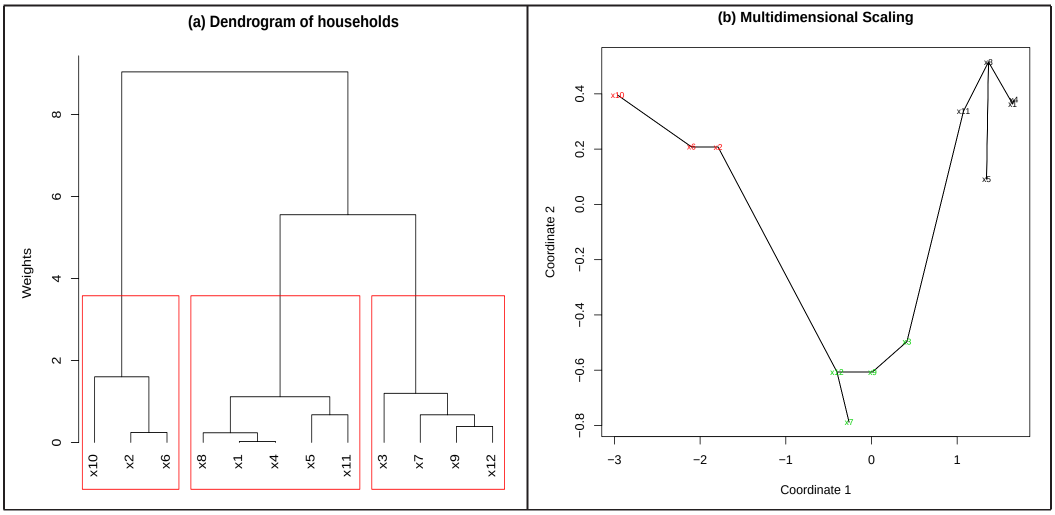


Figure 2: Visualisation of households by clusters

Considering the illustrative example (see Table 2), we obtain clusters $L_1 = \{\mathbf{x}_1, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_8, \mathbf{x}_{11}\}$, $L_2 = \{\mathbf{x}_2, \mathbf{x}_6, \mathbf{x}_{10}\}$, and $L_3 = \{\mathbf{x}_3, \mathbf{x}_7, \mathbf{x}_9, \mathbf{x}_{12}\}$. The *multidimensional distribution* for this population of 12 households is given by the following data table:

$$X = [\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \mathbf{x}^6, \mathbf{x}^7] = \left[\begin{array}{ccccccc} 250.00 & \mathrm{N} & \mathrm{S} & \mathrm{N} & \mathrm{F} & \mathrm{A} & \mathrm{W} \\ 200.00 & \mathrm{N} & \mathrm{S} & \mathrm{N} & \mathrm{F} & \mathrm{A} & \mathrm{W} \\ 800.00 & \mathrm{N} & \mathrm{S} & \mathrm{N} & \mathrm{F} & \mathrm{S} & \mathrm{W} \\ 600.00 & \mathrm{N} & \mathrm{S} & \mathrm{N} & \mathrm{F} & \mathrm{A} & \mathrm{W} \\ 1000.00 & \mathrm{N} & \mathrm{A} & \mathrm{N} & \mathrm{F} & \mathrm{A} & \mathrm{W} \\ \hline 4500.00 & \mathrm{Y} & \mathrm{V} & \mathrm{Y} & \mathrm{V} & \mathrm{N} & \mathrm{Be} \\ 5000.00 & \mathrm{Y} & \mathrm{V} & \mathrm{Y} & \mathrm{V} & \mathrm{N} & \mathrm{Be} \\ 6500.00 & \mathrm{Y} & \mathrm{V} & \mathrm{Y} & \mathrm{V} & \mathrm{N} & \mathrm{U} \\ \hline 1500.00 & \mathrm{Y} & \mathrm{A} & \mathrm{Y} & \mathrm{F} & \mathrm{S} & \mathrm{U} \\ 2500.00 & \mathrm{Y} & \mathrm{A} & \mathrm{Y} & \mathrm{A} & \mathrm{S} & \mathrm{U} \\ 2000.00 & \mathrm{Y} & \mathrm{A} & \mathrm{Y} & \mathrm{A} & \mathrm{S} & \mathrm{Be} \\ 2800.00 & \mathrm{Y} & \mathrm{A} & \mathrm{Y} & \mathrm{A} & \mathrm{S} & \mathrm{Be} \end{array}\right] = \left[\begin{array}{c} \mathbf{x}_1 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \\ \mathbf{x}_8 \\ \mathbf{x}_{11} \\ \hline \mathbf{x}_2 \\ \mathbf{x}_6 \\ \mathbf{x}_{10} \\ \hline \mathbf{x}_3 \\ \mathbf{x}_7 \\ \mathbf{x}_9 \\ \mathbf{x}_{12} \end{array}\right] \quad (8)$$

and its modal-valued matrix is given by:

$$\mathbf{X} = \begin{bmatrix} X_1(L_1) & X_2(L_1) & X_3(L_1) & X_4(L_1) & X_5(L_1) & X_6(L_1) & X_7(L_1) \\ X_1(L_2) & X_2(L_2) & X_3(L_2) & X_4(L_2) & X_5(L_2) & X_6(L_2) & X_7(L_2) \\ X_1(L_3) & X_2(L_3) & X_3(L_3) & X_4(L_3) & X_5(L_3) & X_6(L_3) & X_7(L_3) \end{bmatrix} \equiv \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \quad (9)$$

where each modal-valued criterion $X_j(L_h)$ is: $X_1(L_1) = \langle([200; 1000], 1)\rangle$, $X_2(L_1) = \langle(N, 1.00); (Y, 0.00)\rangle$, $X_3(L_1) = \langle(A, 0.20); (S, 0.80); (V, 0.00)\rangle$, $X_4(L_1) = \langle(N, 1.00); (Y, 0.00)\rangle$, $X_5(L_1) = \langle(A, 0.00); (F, 1.00); (V, 0.00)\rangle$, $X_6(L_1) = \langle(A, 0.80); (N, 0.00); (S, 0.20)\rangle$, $X_7(L_1) = \langle(Be, 0.00); (U, 0.00); (W, 1.00)\rangle$; $X_1(L_2) = \langle([4500; 6500], 1)\rangle$, $X_2(L_2) = \langle(N, 0.00); (Y, 1.00)\rangle$, $X_3(L_2) = \langle(A, 0.00); (S, 0.00); (V, 1.00)\rangle$, $X_4(L_2) = \langle(N, 0.00); (Y, 1.00)\rangle$, $X_5(L_2) = \langle(A, 0.00); (F, 0.00); (V, 1.00)\rangle$, $X_6(L_2) = \langle(A, 0.00); (N, 1.00); (S, 0.00)\rangle$, $X_7(L_2) = \langle(Be, 0.67); (U, 0.33); (W, 0.00)\rangle$; $X_1(L_3) = \langle([1500; 2800], 1)\rangle$, $X_2(L_3) = \langle(N, 0.00); (Y, 1.00)\rangle$, $X_3(L_3) = \langle(A, 1.00); (S, 0.00); (V, 0.00)\rangle$, $X_4(L_3) = \langle(N, 0.00); (Y, 1.00)\rangle$, $X_5(L_3) = \langle(A, 0.75); (F, 0.25); (V, 0.00)\rangle$, $X_6(L_3) = \langle(A, 0.00); (N, 0.00); (S, 1.00)\rangle$, $X_7(L_3) = \langle(Be, 0.50); (U, 0.50); (W, 0.00)\rangle$.
For example, $X_7(L_2)$ should be read as: "considering the variable $X_7$ describing the *'economic situation'* of a household, 67% of households legally own the commodity 'Be' and 33% the commodity 'U', while there is no household with the commodity 'W' within cluster $L_2$".

*3.3. Modalities characterisation*

Sometimes the frequency $\pi^h_{jk}$ of modality $\beta_{jk}$ within cluster $L_h$, such as defined in equation (15), may not accurately reflect how strong is the statement: *''this modality is more relevant in this cluster than in all other clusters"*. This is why we need to characterize the modalities in a probabilistic sense by using *hypothesis testing*. Hypothesis testing is an essential part of statistical inference aiming at determining the probability that a given hypothesis is true. The characterisation used in this paper has been inspired from the principle of statistical characterisation introduced by Morineau (1984). The idea consists at testing if the modality $\beta_{jk}$ of modal-value criterion $X_j(L_h)$ is a relevant characteristic of the cluster $L_h$ through the probability that the hypothesis "the modality $\beta_{jk}$ is *'significantly'* more abundant in the cluster $L_h$ than in the population of $\Omega$" is true. Thus, we set the null hypothesis $H_0$ of random draw (without replacement) of $n_h$ households among the $n$ households of population. $H_0$ ensures that the frequencies $\pi^h_{jk}$ and $p^h_{jk}$ are nearly equal with respect to random fluctuations; where $p^h_{jk}$ is:

$$p^h_{jk} = \frac{|\{x_i \in \Omega \,:\, x^j_i = \beta_{jk}\}|}{|\{\Omega\}|}, \quad k = 1, \ldots, s_j \quad (10)$$

Intuitively, $p^h_{jk}$ represents the *'relative importance'* of the modality $\beta_{jk}$ within the whole population i.e. the number of times that a modality $\beta_{jk}$ occurs within the variable $X_j$ of population $\Omega$.

Let $N$ be a random variable such that $N = n^h_{jk}$ with $n^h_{jk} = n_h \cdot \pi^h_{jk}$. Under hypothesis $H_0$, $N$ follows a hypergeometric distribution, $N \sim \mathcal{H}(n, n_{jk}, n_h)$, with a mean given by $E_h(N) = n_h \cdot p^h_{jk}$, a standard deviation $\sigma^2_h(N) = n_h \cdot \dfrac{n - n_h}{n-1} \cdot p^h_{jk} \cdot \left(1 - p^h_{jk}\right)$ and $n_{jk} = n \cdot p^h_{jk}$.

Hence, the degree of significance $\rho(\beta_{jk})$ under $H_0$ is given by the following equation:

$$\rho(\beta_{jk}) = \mathrm{Prob}_{H_0}\{N > \beta_{jk}\} = \mathrm{Prob}_{H_0}\{t_h(N) > t_h(\beta_{jk})\} \tag{11}$$

where $t_h(N) = \frac{N-E_h(N)}{\sigma_h(N)}$ such that $t_h(N)$ follows a gaussian distribution $\mathcal{N}(0;1)$. In practice, a hypergeometric distribution can be approximated by a gaussian distribution when the number of households in the clusters is sufficiently high (over 30 households for instance).

For each modal-valued criterion $X_j(L_h)$ and each modality $\beta_{jk}$, we define the relevance index $\xi(\beta_{jk})$ by equation (12) as follows:

$$\xi(\beta_{jk}) = \rho(\beta_{jk}) \cdot \pi^h_{jk} + \lambda_h \cdot \zeta^h_{jk} \tag{12}$$

where $\zeta^h_{jk}$ is given by the following equation (13) and $\lambda_h = \dfrac{n_h}{n}$.

$$\zeta^h_{jk} = \frac{|\{x_i \in L_h \,:\, x^j_i = \beta_{jk}\}|}{|\{x_i \in \Omega \,:\, x^j_i = \beta_{jk}\}|} = \frac{n^h_{jk}}{n_{jk}}, \quad k = 1, \ldots, s_j \tag{13}$$

We denote $n_{jk}$ ($n^h_{jk}$ respectively) the number of times that the modality $\beta_{jk}$ occurs within the $j$th variable of population $\Omega$ (of cluster $L_h$ respectively). $\xi(\beta_{jk})$ is nomalised between 0 to 1 interval through equation (14).

$$\widetilde{\xi}(\beta_{jk}) = \frac{\xi(\beta_{jk})}{\max_{\{j,k\}}\left\{\xi(\beta_{jk})\right\}} \tag{14}$$

**Definition 3.2.** *A* modal-valued characterisation $Y_j(L_h)$ *is a fuzzy subset of the set of modalities (or states)* $\{\beta_{jk} \,:\, k = 1, \ldots, s_j\}$ *of a cluster* $L_h$ *over the domain* $D_j$ *of* $X_j$*,* $j \in \mathcal{J}$ *defined by:*

$$Y_j(L_h) \;=\; \left\{\left(\beta_{jk}, \xi(\beta_{jk})\right) \,:\, k = 1, \ldots, s_j\right\} \tag{15}$$

*where* $\xi(\beta_{jk})$ *is the relevance index of the modality* $\beta_{jk}$ *on variable* $j \in \mathcal{J}$ *associated with cluster* $L_h$*,* $h \in \{1 \ldots, \mu\}$*.*

The relevance index $\xi(\beta_{jk})$ appears as a mean showing on a scale between 0 and 1, the level with which the assertion *"$\xi(\beta_{jk})$ is more relevant than $\xi(\beta_{j'k'})$"* is valid. Intuitively, $\xi(\beta_{jk})$ measures the *'degree of connection'* of the modality $\beta_{jk}$ with the cluster $L_h$ considering all other clusters and the whole population i.e. the *'connectedness'* between the modality $\beta_{jk}$ within the variable $X_j$ and a given cluster $L_h$ (as in the case of one causing the other or sharing features with it).

**Remark 3.1.** *The relevance index is a numerical representation which allows comparisons between modalities, but it cannot be treated as a probability because it does not satisfy the 'countable additivity' property. We need these relevance indexes to establish the relation between commodities and welfare dimension levels as we show it in the following sections.*

*3.4. Specification of preferential information*

The decision maker has to specify his preferential information according to policies, programmes or projects that he intends to undertake in a given region. In this paper, we assume that the problem involves at least five types of preferential information (PI): preference information concerning the importance of the variables, preference information concerning welfare dimensions, preference information concerning welfare dimension levels, preference information concerning the ranking of modalities and preference information concerning the specification of households becoming a priority target and precise operations. Then reasons for choosing such types of information are guided by methodological principles (see Colorni and Tsoukiàs, 2013).

*3.4.1. Welfare dimensions*

The decision maker in collaboration with the analyst establishes the set of *"welfare dimensions"*. Welfare dimensions allow to specify and describe the totality of functioning dimensions that the decision maker would like to deal with during the decision process aiming at finding an appropriate final recommendation for improving the living standards of households. This concept is based on the idea that the way through which individuals or households perceive their position or their state within a society is an important aspect to be considered when designing policies which concern them. However, in order to assess the input of policies, we need an analytic structure and for this purpose we introduce the concept of *"welfare dimension"*.

**Definition 3.3.** *A* welfare dimension *describes the set of all possible functionings of a household under the form of achievable levels of welfare on this dimension.*

Formally, let $\mathcal{J}$ be a set of variables. For any non empty subset $\mathcal{J}_u$ of $\mathcal{J}$ we denote by $F_{\mathcal{J}_u}$ (resp. $F_{-\mathcal{J}_u}$) the set $\prod_{j\in\mathcal{J}_u} X_j$ (resp. $\prod_{j\notin\mathcal{J}_u} X_j$) such that $\mathcal{J} = \mathcal{J}_1 \cup \cdots \cup \mathcal{J}_p$ and $\mathcal{J}_1 \cap \cdots \cap \mathcal{J}_p = \emptyset$. $F_{\mathcal{J}_u}$ is a potential welfare dimension describing $\mathcal{J}_u$ functioning and $F = \{F_{\mathcal{J}_u} : u = 1, \ldots, p\}$ denotes the set of all potential welfare dimensions. We will write $F_{\mathcal{J}_u}(L_h)$ if it is related to cluster $L_h$.

Remark that, each $F_{\mathcal{J}_u}$ is perceived as the set of all commodities related to the set $\prod_{j\in\mathcal{J}_u} X_j$ within the whole population $\Omega$ and $F_{\mathcal{J}_u}(L_h)$ is perceived as the set of all commodities related to the set $\prod_{j\in\mathcal{J}_u} X_j$ within the cluster $L_h$. For instance, the welfare dimension can be *'Mobility'* or *'Social Integration'* in the area of social studies and *'Air Pollution'* or *'Risk'* in the area of pollution studies.

*3.4.2. Welfare dimension levels*

The decision maker has to define a set of levels associated to each welfare dimension. A *welfare dimension level* allows, in a more general case, to distinguish the possible intensity degrees within a given welfare dimension. In the case of poverty, it allows to distinguish the degree of necessity of households in term of policy intervention on this welfare dimension. For instance, *'Bad'*, *'Average'* and *'Good'* can be three levels associated to the welfare dimension 'Mobility' or 'Social Integration'; *'Low'*, *'Acceptably Low'*, *'Relatively Low'*, *'High'* and *'Very High'* can be five levels associated to the welfare dimensions 'Air Pollution' or 'Risk'.

Formally, for the $u$th potential welfare dimension $F_{\mathcal{J}_u}$, we obtain a set of levels

$$F_{\mathcal{J}_u} = \{\xi_v^u \,:\, v = 1, \ldots, t_u\} \tag{16}$$

where $u \in \{1, \ldots, p\}$, $\xi_v^u$ denotes the $v$th level of the $u$th potential welfare dimension $F_{\mathcal{J}_u}$ and $t_u$ is the number of possible level(s) actually taken by $F_{\mathcal{J}_u}$.

### *3.5. Construction of capabilities*

#### *3.5.1. Computing of indexes* $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ *and* $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$

For the purpose of constructing the capabilities set associated to each cluster, we introduce two indexes: the index $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and the index $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$. We denote by $\mathbf{M} = \big\{\beta_{jk} : j \in \mathcal{J} \text{ and } k \in \{1, \ldots, s_j\}\big\}$ the set of all modalities where $s_j$ is the number of values actually taken by $X_j$. The idea consists to directly ask the decision maker to partition the set of all modalities in three subsets $\langle \mathbf{S}, \mathbf{N}, \mathbf{U}\rangle$. We interpret the subset $\mathbf{S}$ as containing **S**atisfactory modalities, $\mathbf{N}$ contains **N**eutral modalities while $\mathbf{U}$ contains **U**nsatisfactory ones. Formally, $\mathbf{S} = \big\{\beta_{jk} \,:\, \beta_{jk} > \beta_{jk'} \text{ with } k \neq k' \text{ and } k, k' \in \{1, \ldots, s_j\},\ j \in \mathcal{J}\big\}$, $\mathbf{U} = \big\{\beta_{jk} \,:\, \beta_{jk} < \beta_{jk'} \text{ with } k \neq k' \text{ and } k, k' \in \{1, \ldots, s_j\},\ j \in \mathcal{J}\big\}$ and $\mathbf{N} = \big\{\beta_{jk} : \text{not}(\beta_{jk} > \beta_{jk'}) \text{ and not}(\beta_{jk} < \beta_{jk'}) \text{ with } k \neq k' \text{ and } k, k' \in \{1, \ldots, s_j\},\ j \in \mathcal{J}\big\}$.

Hence, for each $F_{\mathcal{J}_u}$ (with $u = 1 \ldots, p$) we set $\mathbf{S}_{\mathcal{J}_u} = \{\beta_{jk} : j \in \mathcal{J}_u \text{ and } \beta_{jk} \in \mathbf{S}\}$ and $\mathbf{U}_{\mathcal{J}_u} = \{\beta_{jk} : j \in \mathcal{J}_u \text{ and } \beta_{jk} \in \mathbf{U}\}$. Remark that $\mathbf{S} = \bigcup_{1 \leq u \leq p} \mathbf{S}_{\mathcal{J}_u}$ and $\mathbf{U} = \bigcup_{1 \leq u \leq p} \mathbf{U}_{\mathcal{J}_u}$. Then, the twofold partition $\langle \mathbf{S}, \mathbf{U}\rangle$ allows us to evaluate and pairwise rank the clusters according to a welfare dimension level. The evaluation of the cluster $L_h$ is given by the strongness of its **S**atisfactory modalities (see equation 17) and the weakness of its **U**nsatisfactory modalities (see equation 18).

$$\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}) = \sum_{j \in \mathcal{J}_u} w_j \Big( \sum_{k \in \mathbf{S}_{\mathcal{J}_u}} \xi(\beta_{jk}) \Big) \tag{17}$$

$$\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u}) = \sum_{j \in \mathcal{J}_u} w_j \Big( \sum_{k \in \mathbf{U}_{\mathcal{J}_u}} \xi(\beta_{jk}) \Big) \tag{18}$$

where $h = 1, \ldots \mu$, $u = 1, \ldots, p$ and $w_j$ denote the $j$th positive weight representing the importance of the variable $X_j$. Intuitively, $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ (respectively, $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$) measures the *'degree of connection'* (respectively, *'degree of disconnection'*) of the higher (respectively, the lower) welfare dimension level of $F_{\mathcal{J}_u}(L_h)$ with the cluster $L_h$ considering all other clusters and the whole population.

#### *3.5.2. Functioning vectors*

A *'functioning vector'* is a collection of attainable levels on each welfare dimension. A *'component'* of functioning vector is also called *'achievement level'* on a particular welfare dimension. Formally, we end getting the $l$th functioning vector of cluster $L_h$:

$$\mathbf{F}_l(L_h) \quad = \quad \langle F_{\mathcal{J}_1}(L_h); \cdots ; F_{\mathcal{J}_p}(L_h)\rangle_l \tag{19}$$

where $h \in \{1, \ldots, \mu\}$, $l \in \{1, \ldots, l_h\}$ and $l_h$ represents the number of functioning vectors actually taken by $L_h$.

**Remark 3.2.** *Trivially, if each potential welfare dimension $\mathcal{J}_u$ has $N_u$ welfare dimension levels (with $u = 1, \ldots, p$), the number of all possible functioning vectors $N$ s.t.:*

$$N \;=\; N_1 \times N_2 \times \cdots \times N_p \tag{20}$$

*3.5.3. Functioning lattice*

A *functioning lattice* is the visualisation of the totality of functioning vectors partially ordered by natural dominance. Any further relation is established by the client following his own private preferences. In general, the functioning vectors are built independently from the clusters. However, given a cluster we may establish the *functioning threshold* which denotes the *'functioning vectors frontier'* within the functioning lattice.

*3.5.4. Construction of functioning thresholds*

An obvious numerical representation amounts to associate a real number to each subset $\mathbf{C}_h$ of the set of all possible functioning vectors in such a way that the comparison between these numbers faithfully reflects the preference relation $\succsim$ on the various potential welfare dimensions where $\succ$ refers to strict preference relation (asymmetric part) and $\sim$ indifference relation (symmetric part). Note that, $\mathbf{C}_a \succsim \mathbf{C}_b$ can be interpreted as "the standard living (or welfare) offered to households belonging to cluster $L_a$ is considered to be *at least as preferable as* the standard living (or welfare) offered to households belonging to cluster $L_b$". Hence, we associate to each MDPM defined on $\mathcal{J}_u$, noted by $\mathbf{M}_h(\mathcal{J}_u)$, a sequence of alternatives which concretely represents the appropriate interventions to the households belonging to cluster $L_h$. $\mathbf{M}_h$ is then given by the following equation:

$$\mathbf{M}_h(\mathcal{J}_u) = \vartheta_u\big[\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}), \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})\big] \tag{21}$$

where $\vartheta_u$ is the $u$th real-valued function which allows to aggregate $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$. This leads to numerically recoding the value judgment between $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$ on the various potential welfare dimensions $F_{\mathcal{J}_u}$, in such a way that the subsets $\mathbf{C}_a$ and $\mathbf{C}_b$ can simply be compared taking the sum of these functions as follows:

$$\mathbf{C}_a \succsim \mathbf{C}_b \iff \sum_{u=1}^{p} \mathbf{M}_a(\mathcal{J}_u) \;\geq\; \sum_{u=1}^{p} \mathbf{M}_b(\mathcal{J}_u) \tag{22}$$

$$\iff \sum_{u=1}^{p} \vartheta_u\big[\mathbf{C}_a(\mathbf{S}_{\mathcal{J}_u}), \mathbf{C}_a(\mathbf{U}_{\mathcal{J}_u})\big] \;\geq\; \sum_{u=1}^{p} \vartheta_u\big[\mathbf{C}_b(\mathbf{S}_{\mathcal{J}_u}), \mathbf{C}_b(\mathbf{U}_{\mathcal{J}_u})\big] \tag{23}$$

We now construct a *relation judgement* $\mathcal{R}_h(\mathcal{J}_u)$ which is a formal mechanism for linking a specific functioning vector $\xi^u_v,\ v = 1, \ldots, t_u$ to the relevant potential welfare dimension $F_{\mathcal{J}_u},\ u = 1, \ldots, p$. The decision maker (or the client) has to define the attainable levels of each potential welfare dimension $F_{\mathcal{J}_u}$ and the cut-offs $\varepsilon_{vu}$. We suppose that the attainable levels of each potential welfare dimension $F_{\mathcal{J}_u}$ are ordered i.e. $\xi^u_1 > \xi^u_2 > \cdots > \xi^u_{t_u}$, for all

$u \in \{1, \ldots, p\}$.

$$\mathcal{R}_h(\mathcal{J}_u) = \begin{cases} \xi_1^u, & \text{if } \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{1u} \\ \xi_2^u, & \text{if } \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{2u} \\ \vdots & \vdots \\ \xi_{t_u}^u, & \text{if } \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{t_u u} \end{cases} \tag{24}$$

$\mathbf{M}_h(\mathcal{J}_u)$ is given by equation (21) and $\mathbf{M}_h(\varepsilon_u) \geq \varepsilon_{1u} > \varepsilon_{2u} > \cdots > \varepsilon_{t_u u} \geq -\mathbf{M}_h(\varepsilon_u)$ with $\phi_h[\mathcal{J}_u] = \min\left\{\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u});\, \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})\right\}$; $\psi_h[\mathcal{J}_u] = \mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}) + \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$ and $\mathbf{M}_h(\varepsilon_u) = 1 - \dfrac{\phi_h[\mathcal{J}_u]}{\psi_h[\mathcal{J}_u]}$ for all $h \in \{1, \ldots, \mu\}$ and $u \in \{1, \ldots, p\}$.

Hence, given each cluster, we may establish the *sup-functioning vectors*. We call these *functioning thresholds*. Formally, the functioning threshold $\mathbf{F}_{\sup}(L_h)$ is defined as follows:

$$\mathbf{F}_{\sup}(L_h) = \langle \mathcal{R}_h(\mathcal{J}_1), \mathcal{R}_h(\mathcal{J}_2), \ldots, \mathcal{R}_h(\mathcal{J}_p) \rangle \tag{25}$$

*3.5.5. Feasible capabilities*

A *'feasible capability'* is a subset of all possible functioning vectors compatible with a given set of commodities. Formally, from each $l$th functioning vector of cluster $L_h$ denoted by $\mathbf{F}_l(L_h)$, we define the subset $\mathbf{C}_h$ of the set of all possible functioning vectors. $\mathbf{C}_h$ represents the feasible capability of all households within $L_h$ as follows.

$$\mathbf{C}_h \;=\; \left\{\mathbf{F}_l(L_h) \in \mathbb{C} \;:\; \mathbf{F}_l(L_h) \leq \mathbf{F}_{\sup}(L_h)\right\} \tag{26}$$

where $h \in \{1, \ldots, \mu\}$ and $\mathbb{C}$ denotes the non-empty set of all possible functioning vectors. Remark that, $l \in \{1, \ldots, l_h\}$ and $l_h$ represents the number of functioning vectors actually taken by $L_h$. Thus, we introduce the following definition.

**Definition 3.4.** *A* "core poor" *is a household who has a* 'bad score' *on all conflictual commodities taken into account simultaneously to evaluate its* 'feasible capability' *in the society in which he lives, whose measurement of his poverty is in conformity with his self-perception of situation.*

Note that, each feasible capability associated to a given cluster reflects the *'ability'* of households within this cluster to achieve a subset of functionings among the various alternatives functioning bundles.

*3.5.6. Extended capabilities*

While generally commodities are uniquely associated to one specific welfare dimension, some commodities (such as income) are instead associated to the whole set of welfare dimensions. We denote them as *'generic commodities'*.

**Definition 3.5.** *A* generic commodity *is a commodity which can improve any welfare dimension.*

Feasible capabilities can be extended using the generic commodities which allow to increase some welfare dimension levels of some of the functioning vectors. Thus, feasible capabilities are transformed in order to obtain what we call *'extended capabilities'*.

Formally, let $\xi_1^u > \xi_2^u > \cdots > \xi_{t_u}^u$ be the ordered attainable levels for all $u \in \{1, \ldots, p\}$. Let $\mathbf{F}_{\sup}(L_h)$ be the *functioning threshold* of feasible capability $\mathbf{C}_h$ such that we have $\mathbf{F}_{\sup}(L_h) = \langle \xi_{\mathbf{k}_1}^1, \xi_{\mathbf{k}_2}^2, \ldots, \xi_{\mathbf{k}_p}^p \rangle$. Hence, $\xi_{\mathbf{k}_u}^u \in \{\xi_1^u, \xi_2^u, \cdots, \xi_{t_u}^u\}$ for all $\mathbf{k}_u \in \{1, \ldots, t_u\}$ and $u \in \{1, \ldots, p\}$. We set $\xi_{\mathbf{k}_u}^u = \max\limits_{v \in \{1, \ldots, t_u\}} \{\xi_v^u\}$. A extended capability $\widetilde{\mathbf{C}}_h$ is defined as follows for all $h \in \{1, \ldots, \mu\}$:

$$Ext(\mathbf{C}_h) = \widetilde{\mathbf{C}}_h = \{\mathbf{F}_l(L_h) \in \mathbb{C} : \mathbf{F}_l(L_h) \leq \widetilde{\mathbf{F}}_{\sup}(L_h) \text{ with } l = 1, \ldots, l_h\} \tag{27}$$

such that

$$\begin{aligned} \widetilde{\mathbf{F}}_{\sup}(L_h) &= \mathbf{F}_{\sup}(L_h) \cup \langle Q_{\mathbf{k}_1}^1, Q_{\mathbf{k}_2}^2, \ldots, Q_{\mathbf{k}_p}^p \rangle && (28) \\ &= \langle \xi_{\mathbf{k}_1}^1, \xi_{\mathbf{k}_2}^2, \ldots, \xi_{\mathbf{k}_p}^p \rangle \cup \langle Q_{\mathbf{k}_1}^1, Q_{\mathbf{k}_2}^2, \ldots, Q_{\mathbf{k}_p}^p \rangle && (29) \\ &= \langle \Xi_{\mathbf{k}_1}^1, \Xi_{\mathbf{k}_2}^2, \ldots, \Xi_{\mathbf{k}_p}^p \rangle && (30) \end{aligned}$$

where $\Xi_{\mathbf{k}_u}^u = \begin{cases} \xi_{\mathbf{k}_u}^u \cup Q_{\mathbf{k}_u}^u, & \text{if } Q_{\mathbf{k}_u}^u \neq \emptyset; \\ \xi_{\mathbf{k}_u}^u, & \text{else} \end{cases}$ for all $u \in \{1, \ldots, p\}$.
$Q_{\mathbf{k}_u}^u = \{\xi_{\mathbf{k}_v}^v : \xi_{\mathbf{k}_v}^v > \xi_{\mathbf{k}_u}^u \text{ and } \mathbf{k}_v \in \{1, \ldots, t_u\} \backslash \{\mathbf{k}_u\}\}$.

In practice, we can use evidence-based policy (EBP) to tackle this problem. This implies the use of data collection on welfare for experimenting, quantitative and qualitative analysing, the use of poverty knowledge, expert knowledge, existing national and international research, existing statistics, stakeholder skills to judge how and the extent to which generic commodities can be used to increase some welfare dimension levels of some of the functioning vectors. Intuitively, $\widetilde{\mathbf{F}}_{\sup}(L_h)$ represents the maximum 'amount' of functioning vector that a household in $L_h$ can achieve under extended capability $\widetilde{\mathbf{C}}_h$.

### *3.6. An illustrative example*

### *3.7. Modalities characterisation*

Considering our previous illustrative example (Table 2) where $(\lambda_1, \lambda_2, \lambda_3) = (0.42, 0.25, 0.33)$ for each cluster $L_1$, $L_2$ and $L_3$, we obtain the Table (3) using the R software[5] and each modal-valued characterisation $Y_j(L_h)$ with $\xi(\beta_{jk})$ is nomalised between 0 to 1 using equation (14) is: $Y_2(L_1) = \langle (N, 1.00); (Y, 0.00) \rangle$, $Y_3(L_1) = \langle (A, 0.06); (S, 0.85); (V, 0.00) \rangle$, $Y_4(L_1) = \langle (N, 1.00); (Y, 0.00) \rangle$, $Y_5(L_1) = \langle (A, 0.00); (F, 0.95); (V, 0.00) \rangle$, $Y_6(L_1) = \langle (A, 0.85); (N, 0.00); (S, 0.06) \rangle$, $Y_7(L_1) = \langle (Be, 0.00); (U, 0.00); (W, 1.00) \rangle$; $Y_2(L_2) = \langle (N, 0.00); (Y, 0.76) \rangle$, $Y_3(L_2) = \langle (A, 0.00);$

[5]R is a free software programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland (New Zealand) and is currently developed by the R Development Core Team. Free downloadable on http://www.r-project.org (see R Development Core Team, 2011)

$(S, 0.00)$; $(V, 1.00)\rangle$, $Y_4(L_2) = \langle(N, 0.00); (Y, 0.76)\rangle$, $Y_5(L_2) = \langle(A, 0.00); (F, 0.00); (V, 1.00)\rangle$, $Y_6(L_2) = \langle(A, 0.00); (N, 1.00); (S, 0.00)\rangle$, $Y_7(L_2) = \langle(Be, 0.51); (U, 0.17); (W, 0.00)\rangle$; $Y_2(L_3) = \langle(N, 0.00); (Y, 0.89)\rangle$, $Y_3(L_3) = \langle(A, 1.00); (S, 0.00); (V, 0.00)\rangle$, $Y_4(L_3) = \langle(N, 0.00); (Y, 0.89)\rangle$, $Y_5(L_3) = \langle(A, 0.85); (F, 0.05); (V, 0.00)\rangle$, $Y_6(L_3) = \langle(A, 0.00); (N, 0.00); (S, 1.00)\rangle$, $Y_7(L_3) = \langle(Be, 0.37); (U, 0.48); (W, 0.00)\rangle$. Remark that there is a distinction between the modal-valued criterion $X_j(L_h)$ and the modal-valued characterisation $Y_j(L_h)$. Consider for example $X_7(L_3)$ and $Y_7(L_3)$. $X_7(L_3)$ shows that: "for the variable $X_7$ describing the *'economic situation'* of a household, 50% of households *legally* own the commodity 'Be' and the commodity 'U' within cluster $L_2$", whereas $Y_7(L_3)$ shows that: "for the variable $X_7$ describing the *'economic situation'* of a household, the degree of connection of the commodity 'U' (equal to 0.48) is greater than the degree of connection of the commodity 'Be' (equal to 0.37) within cluster $L_3$." Then, the commodity 'U' (equal to 0.48) is *more connected* to cluster $L_3$ than the commodity 'Be' (equal to 0.37) in the same cluster. The notion of relevance is not revealed by the modal-valued criterion $X_j(L_h)$ but only by the modal-valued characterisation $Y_j(L_h)$.

| $F_{\mathcal{J}_u}$ | Housing ($\mathcal{J}_1$) | | | | | Mobility ($\mathcal{J}_2$) | | | | | Nutrition ($\mathcal{J}_3$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X$ | $X_2$ | | $X_3$ | | | $X_4$ | | $X_5$ | | | $X_6$ | | | $X_7$ | | |
| $\beta_{jk}$ | N | Y | A | S | V | N | Y | A | F | V | A | N | S | Be | U | W |
| $\rho(\beta_{jk})$ | 1.00 | 0.00 | 0.03 | 0.99 | 0.16 | 1.00 | 0.00 | 0.16 | 0.99 | 0.16 | 0.99 | 0.16 | 0.03 | 0.07 | 0.16 | 1.00 |
| $\zeta^1_{jk}$ | 1.00 | 0.00 | 0.20 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.83 | 0.00 | 1.00 | 0.00 | 0.20 | 0.00 | 0.00 | 1.00 |
| $\pi^1_{jk}$ | 1.00 | 0.00 | 0.20 | 0.80 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.80 | 0.00 | 0.20 | 0.00 | 0.00 | 1.00 |
| $p^1_{jk}$ | 0.42 | 0.58 | 0.42 | 0.33 | 0.25 | 0.42 | 0.58 | 0.25 | 0.50 | 0.25 | 0.33 | 0.25 | 0.42 | 0.33 | 0.25 | 0.42 |
| $\xi(\beta_{jk})$ | 1.00 | 0.00 | 0.06 | 0.85 | 0.00 | 1.00 | 0.00 | 0.00 | 0.95 | 0.00 | 0.85 | 0.00 | 0.06 | 0.00 | 0.00 | 1.00 |
| $\rho(\beta_{jk})$ | 0.16 | 0.84 | 0.16 | 0.26 | 1.00 | 0.16 | 0.84 | 0.38 | 0.09 | 1.00 | 0.26 | 1.00 | 0.16 | 0.76 | 0.38 | 0.16 |
| $\zeta^2_{jk}$ | 0.00 | 0.43 | 0.00 | 0.00 | 1.00 | 0.00 | 0.43 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.50 | 0.33 | 0.00 |
| $\pi^2_{jk}$ | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.67 | 0.33 | 0.00 |
| $p^2_{jk}$ | 0.42 | 0.58 | 0.42 | 0.33 | 0.25 | 0.42 | 0.58 | 0.25 | 0.50 | 0.25 | 0.33 | 0.25 | 0.42 | 0.33 | 0.25 | 0.42 |
| $\xi(\beta_{jk})$ | 0.00 | 0.76 | 0.00 | 0.00 | 1.00 | 0.00 | 0.76 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 | 0.51 | 0.17 | 0.00 |
| $\rho(\beta_{jk})$ | 0.07 | 0.93 | 0.99 | 0.14 | 0.26 | 0.07 | 0.93 | 0.98 | 0.03 | 0.26 | 0.14 | 0.26 | 0.99 | 0.59 | 0.76 | 0.07 |
| $\zeta^3_{jk}$ | 0.00 | 0.57 | 0.80 | 0.00 | 0.00 | 0.00 | 0.57 | 1.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.80 | 0.50 | 0.67 | 0.00 |
| $\pi^3_{jk}$ | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.75 | 0.25 | 0.00 | 0.00 | 0.00 | 1.00 | 0.50 | 0.50 | 0.00 |
| $p^3_{jk}$ | 0.42 | 0.58 | 0.42 | 0.33 | 0.25 | 0.42 | 0.58 | 0.25 | 0.50 | 0.25 | 0.33 | 0.25 | 0.42 | 0.33 | 0.25 | 0.42 |
| $\xi(\beta_{jk})$ | 0.00 | 0.89 | 1.00 | 0.00 | 0.00 | 0.00 | 0.89 | 0.85 | 0.05 | 0.00 | 0.00 | 0.00 | 1.00 | 0.37 | 0.48 | 0.00 |
| $n_{jk}$ | 5 | 7 | 5 | 4 | 3 | 5 | 7 | 3 | 6 | 3 | 4 | 3 | 5 | 4 | 3 | 5 |

Table 3: Computation of some indexes of modality characterisation

### *3.7.1. Welfare dimensions*

According to the illustrative example, we consider $F = \langle$Housing, Mobility, Nutrition$\rangle$. We associate subsets of commodities to dimensions of welfare. For example, we associate $X_2$ and $X_3$ to "Housing", $X_4$ and $X_5$ to "Mobility", $X_6$ and $X_7$ to "Nutrition".

**Remark 3.3.** *In the illustrative example the unique generic commodity is income given by criterion $X_1$.*

*3.7.2. Welfare dimension levels*

Considering our example, we can associate the three levels '**B**ad', '**A**verage' and '**G**ood' to all welfare dimensions in such a way that $F_{\mathcal{J}_1}(L_h) = F_{\mathcal{J}_2}(L_h) = F_{\mathcal{J}_3}(L_h) = \{\mathbf{G}, \mathbf{A}, \mathbf{B}\}$, where $F_{\mathcal{J}_1} =$ 'Housing', $F_{\mathcal{J}_2} =$ 'Mobility', $F_{\mathcal{J}_3} =$ 'Nutrition'.

*3.7.3. Specification of subsets* **S** *and* **U**

Considering the illustrative example, the decision maker can split the set of **S**atisfactory modalities (and the set of **U**nsatisfactory modalities respectively) of potential dimensions of welfare $\mathcal{J}_1$, $\mathcal{J}_2$ and $\mathcal{J}_3$ (called 'Housing', 'Mobility' and 'Nutrition' respectively) as follows: $\mathbf{S}_{\mathcal{J}_1} = \{X_2.\mathrm{Y}; X_3.\mathrm{V}; X_3.\mathrm{A}\}$, $\mathbf{S}_{\mathcal{J}_2} = \{X_4.\mathrm{Y}; X_5.\mathrm{V}; X_5.\mathrm{A}\}$, $\mathbf{S}_{\mathcal{J}_3} = \{X_6.\mathrm{N}; X_7.\mathrm{Be}\}$, and $\mathbf{U}_{\mathcal{J}_1} = \{X_2.\mathrm{N}; X_3.\mathrm{S}\}$, $\mathbf{U}_{\mathcal{J}_2} = \{X_4.\mathrm{N}; X_5.\mathrm{A}; X_5.\mathrm{F}\}$, $\mathbf{U}_{\mathcal{J}_3} = \{X_6.\mathrm{A}; X_6.\mathrm{S}; X_7.\mathrm{W}\}$ respectively. The **N**eutral modality is given by $\mathbf{N} = \{X_7.\mathrm{U}\}$ but for empirical reasons we add it to $\mathbf{S}_{\mathcal{J}_3}$ and $\mathbf{U}_{\mathcal{J}_3}$. To be more precise: since these subsets are used in order to assign "new individuals" to **S** or **U**, excluding the neutrals from **S** and **U** would imply that "new individuals" *similar* to "neutral" could be directly assigned **N**. In practice, we can assign "new individuals" *similar* to "neutral" both in **S** and **U** in such a way to delete the subset **N**.

*3.7.4. Construction of functionings thresholds*

In the example (see Table 5), we consider that $\vartheta_u$ is given by equation (31).

$$\mathbf{M}_h(\mathcal{J}_u) = \vartheta_u\big[\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}), \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})\big] = \frac{\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}) - \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})}{\max\limits_{h=1,\dots,\mu}\big\{\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}), \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})\big\}} \tag{31}$$

The cut-off $\mathbf{F}_{\sup}(L_h)$ in Figure 3 represents the *'functioning threshold'* that we have to determine. Using equation (17) and equation (18), we compute $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$ for each cluster $L_h$, $h \in \{1, 2, 3\}$ and each welfare dimension $u \in \{1, 2, 3\}$ with $w_j = 1$. These are provided by Table 4.

| | Housing ($\mathcal{J}_1$) | | Mobility ($\mathcal{J}_2$) | | Nutrition ($\mathcal{J}_3$) | |
|---|---|---|---|---|---|---|
| Clusters | $\mathbf{S}_{\mathcal{J}_1}$ | $\mathbf{U}_{\mathcal{J}_1}$ | $\mathbf{S}_{\mathcal{J}_2}$ | $\mathbf{U}_{\mathcal{J}_2}$ | $\mathbf{S}_{\mathcal{J}_3}$ | $\mathbf{U}_{\mathcal{J}_3}$ |
| $\mathbf{C}_1$ | 0.06 | 1.85 | 0.00 | 1.95 | 0.00 | 1.92 |
| $\mathbf{C}_2$ | 1.76 | 0.00 | 1.76 | 0.00 | 1.68 | 0.17 |
| $\mathbf{C}_3$ | 1.89 | 0.00 | 1.74 | 0.90 | 0.85 | 1.48 |

Table 4: Evaluation of $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$

$$\mathcal{R}_h(\mathcal{J}_u) = \begin{cases} \mathbf{G}\text{ood}, & \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{1u} \\ \mathbf{A}\text{verage}, & \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{2u} \\ \mathbf{B}\text{ad}, & \mathbf{M}_h(\mathcal{J}_u) \geq \varepsilon_{3u} \end{cases} \tag{32}$$

$\mathbf{M}_h(\mathcal{J}_u)$ is given by equation (31) and $-\mathbf{M}_h(\varepsilon_u) \leq \varepsilon_{3u} < \varepsilon_{2u} < \varepsilon_{1u} \leq \mathbf{M}_h(\varepsilon_u)$ with $\phi_h[\mathcal{J}_u] = \min\left\{\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u});\, \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})\right\}$; $\psi_h[\mathcal{J}_u] = \mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u}) + \mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$ and $\mathbf{M}_h(\varepsilon_u) = 1 - \dfrac{\phi_h[\mathcal{J}_u]}{\psi_h[\mathcal{J}_u]}$. With $\varepsilon_{1u} = 0.75$ and $\varepsilon_{2u} = 0.50$, we compute the meaningful measurements given by Table 5.

| | $\mathcal{J}_1$ | $\varepsilon_1$ | $\mathcal{J}_2$ | $\varepsilon_2$ | $\mathcal{J}_3$ | $\varepsilon_3$ |
|---|---|---|---|---|---|---|
| $\mathbf{M}_1$ | -0.95 | 0.97 | -1.00 | 1.00 | -1.00 | 1.00 |
| $\mathbf{M}_2$ | 0.93 | 1.00 | 0.91 | 1.00 | 0.79 | 0.91 |
| $\mathbf{M}_3$ | 1.00 | 1.00 | 0.43 | 0.66 | -0.33 | 0.64 |

Table 5: Example of meaningful measurements

Thus we obtain $\mathbf{F}_{\sup}(L_1) = \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle$, $\mathbf{F}_{\sup}(L_2) = \langle \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle$ and $\mathbf{F}_{\sup}(L_3) = \langle \mathbf{G}, \mathbf{B}, \mathbf{B} \rangle$.

#### *3.7.5. Functioning vectors*

Considering our example where we have associated the three levels '**B**ad', '**A**verage' and '**G**ood' to all welfare dimensions s.t. $F_{\mathcal{J}_1}(L_h) = F_{\mathcal{J}_2}(L_h) = F_{\mathcal{J}_2}(L_h) = \{\mathbf{G},\, \mathbf{A},\, \mathbf{B}\}$ for $F_{\mathcal{J}_1}$ = 'Housing', $F_{\mathcal{J}_2}$ = 'Mobility', $F_{\mathcal{J}_3}$ = 'Nutrition', the $l$th functioning vector associated to cluster $h$ can be given by $\mathbf{F}_l(L_h) = \langle \mathbf{A}; \mathbf{B}; \mathbf{G} \rangle$ and an achievement level on welfare dimension 'Nutrition' is '**G**'.

#### *3.7.6. Functioning lattice*

The totality of all possible functioning vectors can be visualized through the functioning lattice given by Figure (3) which shows the functioning vectors from the 'best' ones (given by $\langle \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle$) to the worst (given by $\langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle$). If we suppose for example that the symbol $\rightarrow$ denotes *preference relation* $\succsim$, $\langle \mathbf{G}, \mathbf{A}, \mathbf{G} \rangle \rightarrow \langle \mathbf{G}, \mathbf{B}, \mathbf{A} \rangle$ can be interpreted as "the standard living (or welfare) offered by $\langle \mathbf{G}, \mathbf{A}, \mathbf{G} \rangle$ is considered to be *at least as preferable as* that the standard living (or welfare) offered by $\langle \mathbf{G}, \mathbf{B}, \mathbf{A} \rangle$". The notation introduced in Figure (3) refers to the usual dominance dominance existing between functionings. In other terms,

$$\mathbf{F}_{\mathbf{a}} \rightarrow \mathbf{F}_{\mathbf{b}} \quad \text{iff} \quad \langle F_{\mathcal{J}_1}; \cdots; F_{\mathcal{J}_p} \rangle_{\mathbf{a}} \rightarrow \langle F_{\mathcal{J}_1}; \cdots; F_{\mathcal{J}_p} \rangle_{\mathbf{b}} \tag{33}$$

$$\text{iff} \quad \langle \xi^1_{\mathbf{a}_1}, \xi^2_{\mathbf{a}_2}, \ldots, \xi^p_{\mathbf{a}_p} \rangle \rightarrow \langle \xi^1_{\mathbf{b}_1}, \xi^2_{\mathbf{b}_2}, \ldots, \xi^p_{\mathbf{b}_p} \rangle \tag{34}$$

$$\text{iff} \quad \forall\, u,\ \xi^u_{\mathbf{a}_u} \succsim \xi^u_{\mathbf{b}_u} \text{ and } \exists\, w,\ \xi^w_{\mathbf{a}_w} \succ \xi^w_{\mathbf{b}_w}. \tag{35}$$

where $\xi^u_{\mathbf{a}_u}, \xi^u_{\mathbf{b}_u} \in \{\xi^u_1, \xi^u_2, \cdots, \xi^u_{t_u}\}$ for all $\mathbf{a}_u, \mathbf{b}_u \in \{1, \ldots, t_u\}$ and $u \in \{1, \ldots, p\}$. $\xi^u_{\mathbf{a}_u}$ (resp. $\xi^u_{\mathbf{b}_u}$) denotes the $\mathbf{a}_u$th (resp. $\mathbf{b}_u$th) level of the $u$th potential welfare dimension $F_{\mathcal{J}_u}$ and $t_u$ is the number of possible level(s) actually taken by $F_{\mathcal{J}_u}$. However, some cases (such as $\langle \mathbf{A}, \mathbf{G}, \mathbf{B} \rangle$ vs $\langle \mathbf{G}, \mathbf{B}, \mathbf{A} \rangle$) can be conflictual and difficult to interpret in term of preference relation. In this situation, the client has to use his own private preferences to decide.

#### *3.7.7. Feasible capabilities*

The feasible capabilities are given in Table 6. A quick analysis of households from each cluster shows that the 'worst' feasible capabilities $\mathbf{C}_1$ ($\mathbf{F}_{\sup}(L_1) = \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle$) are characterised by the fact that the heads of households are in the majority described by

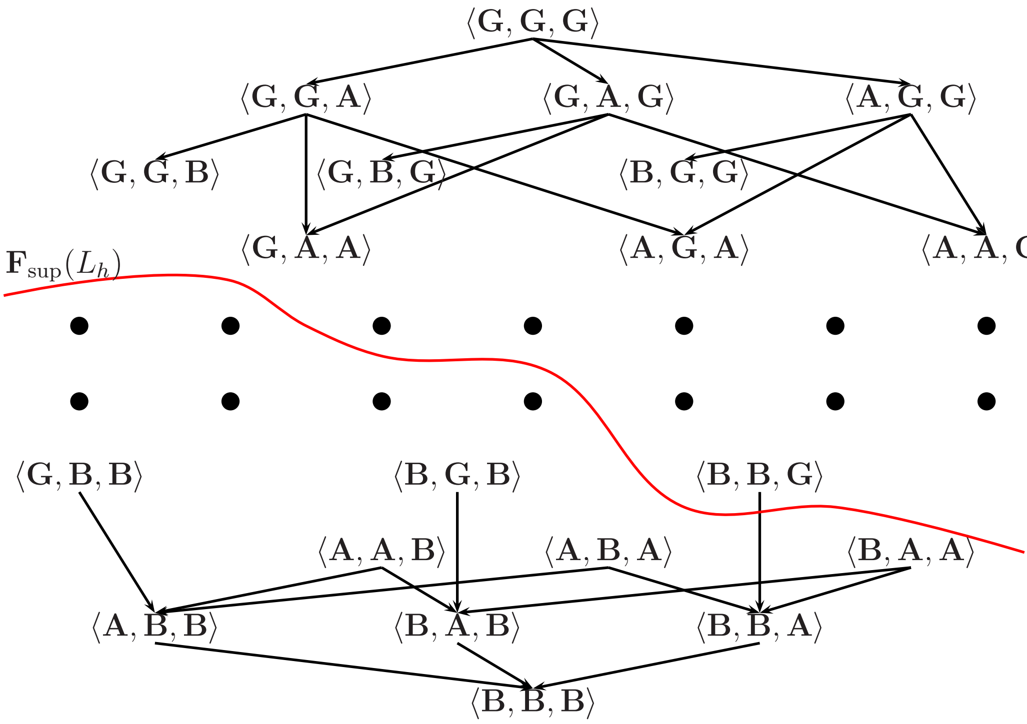


Figure 3: Functioning lattice (from the example)

**U**nsatisfactory modalities $\mathbf{U}_{\mathcal{J}_1}$, $\mathbf{U}_{\mathcal{J}_2}$ and $\mathbf{U}_{\mathcal{J}_3}$. They are those we have called 'core poor' (see definition 3.4) in this paper. The 'best' ones $\mathbf{C}_2$ ($\mathbf{F}_{\text{sup}}(L_2) = \langle \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle$) are characterised by the fact that the heads of households are in the majority described by the **S**atisfactory modalities $\mathbf{S}_{\mathcal{J}_1}$, $\mathbf{S}_{\mathcal{J}_2}$ and $\mathbf{S}_{\mathcal{J}_3}$. The feasible capabilities $\mathbf{C}_3$ ($\mathbf{F}_{\text{sup}}(L_3) = \langle \mathbf{G}, \mathbf{B}, \mathbf{B} \rangle$) is characterised by households which are in majority described by **S**atisfactory and **U**nsatisfactory modalities $\mathbf{S}_{\mathcal{J}_1}$, $\mathbf{U}_{\mathcal{J}_2}$ and $\mathbf{U}_{\mathcal{J}_3}$. Thus, we look for the presence and absence of modalities in each cluster. It is important to note that, the words 'worst' and 'best' do not contain anything of numerical but only reflect the household's well-being and advantage.

$$\mathbf{C}_1 = \big\{ \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle \big\}$$
$$\mathbf{C}_2 = \big\{ \langle \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle; \langle \mathbf{G}, \mathbf{G}, \mathbf{A} \rangle; \langle \mathbf{G}, \mathbf{A}, \mathbf{G} \rangle; \langle \mathbf{A}, \mathbf{G}, \mathbf{G} \rangle; \cdots \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle \big\}$$
$$\mathbf{C}_3 = \big\{ \langle \mathbf{G}, \mathbf{B}, \mathbf{B} \rangle; \langle \mathbf{A}, \mathbf{B}, \mathbf{B} \rangle; \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle \big\}$$

Table 6: The set of feasible capabilities

### *3.7.8. Extended capabilities*

The feasible capabilities are transformed to extended capabilities using the generic commodities (such as income on the illustrative example) which allow to increase some of the attainable functionings. This raises two main questions: How to increase some of the attainable functionings while avoiding contradictions and inconsistencies? How to deal with legitimacy and rationality of such a procedure?

The analyst can compute the attainable levels of each welfare dimension which he can improve (or increase) using the preferential information provided by the decision maker. Considering the set of feasible capabilities shown in Table (6), we note that it is possible to improve the feasible capability $\mathbf{C}_3$ on the welfare dimension $\mathcal{J}_2$ representing 'Mobility'. In fact, a quick observation shows that $\mathbf{M}_3(\mathcal{J}_3) < \mathbf{M}_3(\mathcal{J}_2) < \varepsilon_v = 0.50$. Thus, one can be tempted to conclude that "the standard living (or welfare) offered by $\mathbf{C}_1$ and $\mathbf{C}_3$ on the welfare dimension $\mathcal{J}_2$ is *'equivalent to'* the standard living (or welfare) offered by $\mathbf{C}_1$ and $\mathbf{C}_3$ on the welfare dimension $\mathcal{J}_3$". The modal-valued matrix given by equation (9) shows that $X_1(L_1) = [200, 1000]$ while $X_1(L_3) = [1500, 2800]$. As $\mathbf{M}_3(\mathcal{J}_2) > \mathbf{M}_3(\mathcal{J}_3)$ and $X_1(L_3) > X_1(L_1)$ (i.e $\mathbf{M}_3(\mathcal{J}_2) > \mathbf{M}_3(\mathcal{J}_3)$ in Table 5 for the generic commodity 'income' belonging in interval [1500, 2800]), we can intuitively increase the attainable level of welfare dimension $\mathcal{J}_2$ of $L_3$ from $\mathbf{B}$ to $\mathbf{A}$. For feasible capability $\mathbf{C}_3$, we have

$$\widetilde{\mathbf{F}}_{\sup}(L_3) = \langle \xi^1_{\mathbf{k}_1}, \xi^2_{\mathbf{k}_2}, \xi^3_{\mathbf{k}_3} \rangle \cup \langle Q^1_{\mathbf{k}_1}, Q^2_{\mathbf{k}_2}, Q^3_{\mathbf{k}_3} \rangle \tag{36}$$

$$= \langle \Xi^1_{\mathbf{k}_1}, \Xi^2_{\mathbf{k}_2}, \Xi^3_{\mathbf{k}_3} \rangle \tag{37}$$

where $\Xi^1_{\mathbf{k}_1} = \xi^1_{\mathbf{k}_1} = \mathbf{G}$ and $Q^1_{\mathbf{k}_1} = \emptyset$; $\Xi^2_{\mathbf{k}_2} = \xi^2_{\mathbf{k}_2} \cup Q^2_{\mathbf{k}_2} = \{\mathbf{B}\} \cup \{\mathbf{A}\} = \mathbf{A}$ and $Q^1_{\mathbf{k}_1} = \{\mathbf{A}\}$; $\Xi^3_{\mathbf{k}_3} = \xi^3_{\mathbf{k}_3} = \mathbf{B}$ and $Q^3_{\mathbf{k}_3} = \emptyset$. Thus, the extended capability of the feasible capability $\mathbf{C}_3$ is given in Table 7. Note that a such transformation from feasible capabilities to extended capabilities depends on preferential information provided by the decision maker and scientific evidence from empirical experiences. We do not see any procedural legitimacy issue here. The whole procedure is based on merging data driven analysis with client (value) driven analysis. If the method is expected to be used by decision maker it should reflect his/her values. There is no trade-off issue either. Experiences are used as a basis for "learning" preferences and values which the decision maker may find difficult to express directly.

| |
|---|
| $\widetilde{\mathbf{C}}_1 = \{\langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle\}$ |
| $\widetilde{\mathbf{C}}_2 = \{\langle \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle; \langle \mathbf{G}, \mathbf{G}, \mathbf{A} \rangle; \langle \mathbf{G}, \mathbf{A}, \mathbf{G} \rangle; \langle \mathbf{A}, \mathbf{G}, \mathbf{G} \rangle;\ \cdots\ \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle\}$ |
| $\widetilde{\mathbf{C}}_3 = \{\langle \mathbf{G}, \mathbf{A}, \mathbf{B} \rangle; \langle \mathbf{G}, \mathbf{B}, \mathbf{B} \rangle; \langle \mathbf{A}, \mathbf{A}, \mathbf{B} \rangle; \langle \mathbf{A}, \mathbf{B}, \mathbf{B} \rangle; \langle \mathbf{B}, \mathbf{A}, \mathbf{B} \rangle; \langle \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle\}$ |

Table 7: The set of extended capabilities

## 4. Taking action

The reader should note that meaningful measurement is carried out before policy making and policy implementation, since the analysis of the positive and negative consequences of potential policies is expected to be done exactly in order to supply the appropriate information to decision maker. These information are crucial for the decision makers in the sense that they help them to select adequate policies for each cluster of households. In other words, the reason to being of meaningful measurement is related to its capacity to help in designing and identifying of preferable alternatives with respect to complex policy issues. The meaningful measurement allows to make analytics among alternatives. Analytics, in our context, implies highlighting the implicit values and judgments in adequation with the preference information of the decision maker and identifying clusters which are in relation with them. Clusters which preclude implication for alternative policy choices are not worth to be considered.

Suppose that the decision maker establishes a set of actions, let say: house improvement programme (such as sanitation infrastructure supply, water supply, power supply), nutrition improvement programme, support for micro credit, etc. Considering our illustrative example, analytics imply to conclude that the policy established by the decision maker only concerns poorly housed and poorly nourished households i.e. $L_1$, and poorly nourished and acceptably transported households i.e. $L_2$. Remark that the households from $L_2$ are subject to any particular improvement programme. Thus, identifying the clusters which are in relation with the policy established by the decision maker can lead to the following policy implementation: households from $L_1$ are subject to alternative (i), households from $L_3$ are subject to alternative (ii) and households from $L_2$ are subject to alternative (iii) as specified in the following:

(i) poorly housed and poorly nourished households are subject to a specific house improvement programme (such as sanitation infrastructure supply, water supply, power supply) and a nutrition improvement programme totally in charge of government.

(ii) poorly nourished and acceptably transported households are subject to a specific nutrition improvement programme totally in charge of government if they are also poorly housed, but they have to contribute up to 70% of costs if they are acceptably or well housed.

(iii) well-housed, well-transported and well nourished households are not subject to any particular policy but they can benefit of tax deduction if they financially contribute or support the government in applying concrete actions in (i) and (ii).

## 5. Conclusion

We have introduced a new methodology for multidimensional poverty measurement based on Sen's capability approach under a decision aiding perspective. The methodology is based on two stages.

The first stage consists to an unsupervised classification of population samples who aims at finding a convenient and valid segmentation of the population in clusters with homogenous socio-economic commodities. To achieve this goal, we use a statistical classification technique for discovering whether the households of a population fall into different groups by making quantitative comparisons of such commodities. Then, the input data table representing the set of commodity vectors of households is organised into clusters in such a way that each cluster consists of households that are similar in term of commodities distribution between themselves and dissimilar to households of the others clusters. We have established the concept of *"cluster description"* aiming at transforming the large data table into a summary table in order to gain initial knowledge. Cluster description gives a summary of the original data table in an explanatory way (i.e., close to the initial language of the user) by identifying the *'relative importance'* of variable within a given cluster and by providing descriptions based on properties concerning the initial variables. On the other hand, we have shown how to start with more complex data and derive meaningful measurements in order to build a *bridge* (subjective perception of poverty) between the *physical world* (what is that poverty) and the *informational world* (objective knowledge on poverty and policy effectiveness). For this purpose, we have constructed the *'relevance indexes'* by measuring the *'connectedness'*(or *'degree of connection'*) between each modality $\beta_{jk}$ within the variable $X_j$ and a given cluster $L_h$ (as in the case of one causing the other or sharing features with it). The relevance index is a numerical representation (it cannot be treated as a probability because it does not satisfy the 'countable additivity' property) aiming at constructing the relation between commodities and welfare dimension levels, from which we derive the functionings and the feasible capabilities corresponding to each cluster. Each feasible capability associated to a given cluster reflects the ability of households within this cluster to achieve a subset of functionings among the various alternative functioning bundles. Depending on generic commodities, the feasible capability of a cluster can be improved in such a way to obtain an extended capability. This allows to take into account in the meaningful multidimensional poverty measurement (MDPM) process some commodities such as income, occupational status, marital status, age, size of household, etc. Specifically, MDPM is presented as an instrument that helps a decision maker in making well-informed decisions about policies, programmes and projects by putting the best available evidence from research at the heart of policy development and implementation. Through meaningful measurements (easy to interpret and close to the initial language of the user), decision makers are able to build himself a vivid picture of different types of poverty, both across countries, regions and the world and within countries by ethnic group, urban/rural location, or other key household features. By being independent of the initial data table, decision makers are able to target resources and design policies more effectively.

The second stage aims at formulating an assignment procedure of 'new' households to one or several clusters by examining the commodities vector of each household and by referring to diagnosis clusters, admissibility indices, eligibility indices and rejection indices that we have broadly developed in Kana (2012). We suppose that clusters are *not ordered*. They are rather described/characterised by one or several 'central' modalities grouped into subsets such as each subset, called *'diagnosis cluster'*, be associated to their corresponding cluster. Hence, we proceed to assign the new households to specific clusters using diagnosis of clusters. To achieve this goal, we define an algorithm based on the idea of supervised classification methods. As we mentioned in Introduction, this stage (schematised as the part appearing in Figure (1) with *"double line"*) has not been developed in this paper, but in a forthcoming. In this forthcoming paper, we introduce a new multicriteria filtering method allowing to the decision maker getting a nuanced and reliable view about degree of membership of each household to different clusters, and the potential of the method about "predicting the consequences and valuing the outcomes".

## References


S. Alkire. *Valuing Freedoms: Sen's Capability Approach and Poverty Reduction.* Oxford University Press, New York, 2005.

R.M. Anderberg. *Cluster analysis for applications.* Academic Press, New York, 1973.

P. Berkhin. Survey of clustering data mining techniques. Technical report, Accrue Software, 2002.

A. Bertin. *Pauvreté Monétaire, Pauvreté non Monétaire: une Analyse des Intéractions Appliquée à la Guinée.* PhD thesis, Université Montesquieu-Bordeaux IV, 2007.

D. Bouyssou, T. Marchant, M. Pirlot, P. Perny, A. Tsoukiàs, and Ph. Vincke. *Evaluation and Decision Models: a Critical Perspective.* Kluwer Academic, Boston, 2000.

A. Colorni and A. Tsoukiàs. What is a decision problem? preliminary statements. In *Algorithmic Decision Theory*, volume 8176 of *Lecture Notes in Computer Science*, pages 139–153. Springer, 2013.

E. Diday, J. Lemaire, J. Pouget, and F. Testu. *Éléments d'analyse des données.* Bordas, Paris, 1982.

A. Fusco. *La Contribution des Analyses Multidimensionnelles à la Compréhension et à la Mesure du Concept de Pauvreté: Application Empirique au Panel Communautaire des Ménages.* PhD thesis, Université de Nice-Sophia Antipolis, 2005.

A.D. Gordon. *Classification.* 2nd Edition, Chapman & Hall/CRC, New York, 1999.

A.K. Jain and C.R. Dubes. *Algorithms for Clustering.* Prentice-Hall, Inc, Englewood Cliffs, New Jersey 07632, 1988.

V.Z. Kana. *Construction d'Indices: Mesure Multidimensionnelle de la Pauvreté.* PhD thesis, Université Paris Dauphine, 2012.

V.Z. Kana, B. Somé, and A. Tsoukiàs. A survey on multidimensional poverty measurement: a decision aiding perspective. Cahier du LAMSADE N° 313, LAMSADE, CNRS, 2011.

D.H. Krantz, R.D. Luce, P. Suppes, and A. Tversky. *Foundation of Measurement: Additive and Polynomial Representations.* Academic Press, vol. 1, New York, 1971.

L. Mari. Epistemology of measurement. *Measurement, Elsevier Science Ltd*, 34(1):17–30, 2003.

G. Milligan and M. Cooper. A study of standardization of variables in cluster analysis. *Journal of Classification*, 5(2):181–204, 1988.

A. Morineau. Note sur la caractérisation statistique d'une classe et les valeurs tests. *Bulletin Technique Centre Statistique Informatique Appliquées*, 2(1-2):20–27, 1984.

M.C. Nussbaum. Nature, function and capability: Aristotle on political distribution. Working Papers 31, UNU-WIDER, 1987.

M.C. Nussbaum. *Women and Human Development: the Capabilities Approach.* Cambridge University Press, Cambridge, 2000.

S.R. Osmani. The entitlement approach to famine: An assessment. Research paper, World Institute for Development Economics Research, 1993.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2011.

F.S. Roberts. *Measurement Theory with Applications to Decisionmaking, Utility, and the Social Sciences.* Encyclopedia of Mathematics and its Applications (vol. 7), Addison-Wesley, 1979.

A.K. Sen. Famines as failures of exchange entitlements. *Economic and Political Weekly*, 11(31/33):1273–1280, 1976.

A.K. Sen. Starvation and exchange entitlements: A general approach and its application to the great Bengal famine. *Cambridge Journal of Economics*, 1(1):33–59, 1977.

A.K. Sen. Equality of what? The Tanner lecture on human values, Stanford University, 1979.

A.K. Sen. *Poverty and Famines: an Essay on Entitlement and Deprivation.* Clarendon Press, Oxford, 1981.

A.K. Sen. *Commodities and Capabilities.* Elsevier Science Publishers B.V., North-Holland, 1985.

A.K. Sen. *Capability and Well-being.* in Nussbaum M. and Sen A.(eds), The Quality of Life, Oxford University Press, 1993.

S.S. Stevens. On the theory of scales of measurement. *American Association for the Advancement of Science*, 103(2684):677–680, 1946.

P. Suppes. *Measurement empirical meaningfulness and three-valued logic.* in Churchman and Ratoosh (eds.), Measurement: Definition and Theories, John Wiley & Sons, New York, 1959.

R. Suzuki and H. Shimodaira. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics (Oxford, England)*, 22(12):1540–1542, 2006.

A. Tsoukiàs. On the concept of decision aiding process: an operational perspective. *Annals of Operations Research*, 154(1):3–27, 2007.

A. Tsoukiàs. From decision theory to decision aiding methodology. *European Journal of Operational Research*, 187(1):138–161, 2008.

UNDP. Indicateurs pour le suivi des Objectifs du Millénaire pour le Développement - Définitions, Raison d'être, Concepts et Sources de données. United nations development report, UNDP, 2003.

J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

## 6. Appendix: Application on ASSL 2007 database of Burkina Faso

### *6.1. Databases*

We start presenting succinctly the whole database that we use for a broad illustration of our methodology. It is about the Annual Survey of Standard of Living of Households for year 2007 (ASSL) from the Burkina Faso's National Institute of Statistics and Demography (INSD[6]). Burkina Faso implementation of this survey follows standardized guidelines and receives technical assistance, in terms of Unified Questionnaire of Basic Indicators of Well-being (QUIBB 2007 in french), the sampling procedure and training of the enumerators, so that within the survey there is greater homogeneity and comparability than between other national multi-topic household surveys. The annual survey monitoring the living conditions of households was conducted from the perspective of a better understanding of poverty in Burkina Faso and for better tracking its manifestations. It aims at providing useful data to refine the analysis within the various sectoral and thematic groups of institutional arrangements for monitoring of the implementation of the Strategic Framework to Fight against Poverty. It will allow all stakeholders of the Poverty Reduction Strategy Papers (PRSP) to obtain information to determine the trends of poverty in Burkina Faso by updating the indicators. An indicator can be defined as a measurement that helps us to understand where we are, where we are going and how far we are from the goal. It allows to summarize the characteristics of systems or highlight what is happening in a system. Note that we have made the hypothesis in this paper that the variables describe the commodities of households.

The survey dataset (1255 households on 48 variables) used in order to illustrate our methodology is a national representative sample of households living in Ouagadougou (the capital of Burkina Faso). The ASSL provides information on the dimensions of poverty that we have grouped in six potential welfare dimension characterizing a cluster $L_h$ of the population $\Omega$ as follows: nutrition, education, water and sanitation, housing, health and transportation. Note that, the decision about the number of potential welfare dimension and the choice of variables to associate to each of them depend on the decision maker (or policy maker) and its interest field. The choices we did in this paper, concerning the six potential welfare dimensions, are just for illustrative purposes.

[6]INSD (Institut National de la Statistique et de la Démographie) in French.

**Remark 6.1.** *This survey has been done in the framework of the Millennium Development Goals (MDG). The MDG and targets come from the Millennium Declaration, signed by 189 countries, including 147 heads of State and Government, in September 2000 (see UNDP, 2003).*

- **Nutrition:** ASSL provides nutritional information for each household member. We use three dimensions of poverty to identify whether a household is deprived or vulnerable in term of nutrition (poorly nourished) by measuring directly whether a household has had problems satisfying his basic needs and his economic situation.

- **Education:** ASSL provides information on the years of education and access to school for each household member. Years of schooling capture the level of knowledge and understanding of each household. Like the first one, it doesn't capture the quality of education nor the level of skills but, we have considered it in this paper as a relative good indicator of functionings.

- **Water and Sanitation (WS):** The factor *water and sanitation* uses six dimensions of poverty which, in combination, represent the deprivation situation of each cluster of households in terms of access to clean drinking water and to adequate sanitation. It includes two standard MDG indicators (clean drinking water and improved sanitation) which provide some rudimentary indications of the quality of water and sanitation services for the households.

- **Housing**: The factor *housing* uses twenty-one dimensions of poverty which, in combination, represent the housing poverty situation of each cluster. It includes one standard MDG indicators (the use of clean cooking fuel) and two non-MDG indicators (electricity and flooring material). Both of them provide some rudimentary indications of the quality of housing for each cluster.

- **Health**: The factor *health* uses one dimension of poverty which evaluates access to health service no matter what mode of transportation is required to access it.

- **Mobility**: This factor covers the ownership of some consumer goods for transportation such as bicycle, motorbike, car and the access to public service transportation.

Remark that, each $F_{\mathcal{J}_u}$, $u \in \{1, 2, 3, 4, 5, 6\}$ is perceived as the set of all commodities related to the set $\prod_{j\in\mathcal{J}_u} X_j$ within the whole population $\Omega$ and $F_{\mathcal{J}_u}(L_h)$ is perceived as the set of all commodities related to the set $\prod_{j\in\mathcal{J}_u} X_j$ within the cluster $L_h$.

The common problem of missing data can happen during cluster analysis. To solve this problem, we only deal with valid values. Thus, this methodology can be also applied to others datasets such that the Demographic and Health Surveys (DHS) and World Health Surveys (WHS). Our choice for ASSL 2007 was only motivated by the needs of the illustration.

### 6.2. *Application*

For a broad illustration of our methodology, let us consider the survey dataset of standard of living of 1255 households evaluated on 48 variables related to six potential dimensions of welfare (see section 6.1). We conducted hierarchical cluster analysis via multiscale bootstrap (number of bootstrap 1000; (see Suzuki and Shimodaira, 2006)) using the Ward method (Ward, 1963) and a correlation-based dissimilarity matrix.

#### 6.2.1. *Cognitive competence*

The best number of clusters was obtained for eight clusters. This result can be also confirmed by Calinski and Harabasz's (1974) index which allows to compare the homogeneity of partitions. Using the R software with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8) = (0.063, 0.272, 0.224, 0.164, 0.09, 0.118, 0.028, 0.042)$ and $\mathbf{L} = \{L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8\}$, we have obtained the following tables. Table 8 gives a summary of the original data table in an explanatory way

| | Housing | | Mobility | | Nutrition | | WS | | Health | | Education | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Clusters | $\mathbf{S}_{\mathcal{J}_1}$ | $\mathbf{U}_{\mathcal{J}_1}$ | $\mathbf{S}_{\mathcal{J}_2}$ | $\mathbf{U}_{\mathcal{J}_2}$ | $\mathbf{S}_{\mathcal{J}_3}$ | $\mathbf{U}_{\mathcal{J}_3}$ | $\mathbf{S}_{\mathcal{J}_4}$ | $\mathbf{U}_{\mathcal{J}_4}$ | $\mathbf{S}_{\mathcal{J}_5}$ | $\mathbf{U}_{\mathcal{J}_5}$ | $\mathbf{S}_{\mathcal{J}_6}$ | $\mathbf{U}_{\mathcal{J}_6}$ |
| $\mathbf{C}_1$ | 0.87 | 0.04 | 0.82 | 0.05 | 0.92 | 0.07 | 0.87 | 0.07 | 0.72 | 0.08 | 0.86 | 0.11 |
| $\mathbf{C}_2$ | 0.16 | 0.95 | 0.24 | 0.67 | 0.53 | 0.59 | 0.18 | 1.00 | 0.26 | 0.63 | 0.37 | 0.78 |
| $\mathbf{C}_3$ | 0.37 | 0.72 | 0.70 | 0.41 | 0.66 | 0.36 | 0.44 | 0.78 | 0.45 | 0.35 | 0.38 | 0.51 |
| $\mathbf{C}_4$ | 0.11 | 1.00 | 0.08 | 0.81 | 0.42 | 0.54 | 0.08 | 0.80 | 0.28 | 0.35 | 0.21 | 0.65 |
| $\mathbf{C}_5$ | 0.74 | 0.10 | 0.74 | 0.06 | 0.84 | 0.16 | 0.82 | 0.09 | 0.86 | 0.03 | 0.77 | 0.16 |
| $\mathbf{C}_6$ | 0.60 | 0.22 | 0.62 | 0.17 | 0.84 | 0.14 | 0.72 | 0.24 | 0.89 | 0.05 | 0.63 | 0.22 |
| $\mathbf{C}_7$ | 0.99 | 0.05 | 1.00 | 0.03 | 1.00 | 0.15 | 0.88 | 0.04 | 1.00 | 0.00 | 1.00 | 0.01 |
| $\mathbf{C}_8$ | 0.94 | 0.03 | 0.91 | 0.02 | 0.69 | 0.15 | 0.98 | 0.02 | 0.87 | 0.01 | 0.93 | 0.02 |

Table 8: Evaluation of $\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$ and $\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$

(i.e. close to the language of the decision maker) by identifying the *'relative importance'* of each welfare dimension within a given cluster and by providing descriptions based on **S**atisfactory modalities ($\mathbf{C}_h(\mathbf{S}_{\mathcal{J}_u})$) and **U**nsatisfactory modalities ($\mathbf{C}_h(\mathbf{U}_{\mathcal{J}_u})$).

From equation (22), we obtain the ordered representation (38) which shows that the feasible capability $\mathbf{C}_4$ is the poorest ones and then the highest priority in term of policies intervention or policies implementation.

$$\mathbf{C}_7 \succ \mathbf{C}_8 \succ \mathbf{C}_1 \succ \mathbf{C}_5 \succ \mathbf{C}_6 \succ \mathbf{C}_3 \succ \mathbf{C}_2 \succ \mathbf{C}_4 \tag{38}$$

From Table 9, we can derive the *functioning thresholds* $\mathbf{F}_{\sup}(L_h)$ which are given by the first functioning vector of the feasible capabilities in Table 10. For instance, we have $\mathbf{F}_{\sup}(L_1) = \langle \mathbf{G}, \mathbf{G}, \mathbf{G}, \mathbf{G}, \mathbf{A}, \mathbf{G} \rangle$ and $\mathbf{F}_{\sup}(L_2) = \mathbf{F}_{\sup}(L_4) = \langle \mathbf{B}, \mathbf{B}, \mathbf{B}, \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle$, $\mathbf{F}_{\sup}(L_3) = \langle \mathbf{B}, \mathbf{A}, \mathbf{A}, \mathbf{B}, \mathbf{B}, \mathbf{B} \rangle$, $\mathbf{F}_{\sup}(L_5) = \mathbf{F}_{\sup}(L_6) = \langle \mathbf{A}, \mathbf{A}, \mathbf{A}, \mathbf{A}, \mathbf{G}, \mathbf{A} \rangle$, $\mathbf{F}_{\sup}(L_7) = \langle \mathbf{G}, \mathbf{G}, \mathbf{G}, \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle$, $\mathbf{F}_{\sup}(L_8) = \langle \mathbf{G}, \mathbf{G}, \mathbf{A}, \mathbf{G}, \mathbf{G}, \mathbf{G} \rangle$. Since it makes more sense to have eight different clusters of households being differently poor, the data table is organised into clusters in such a way that each classe consists of households that are *similar* in term of commodities distribution between themselves and *dissimilar* to households of other clusters. It is likely to have more effective policies if these are correctly targeted. The ordered representation (38) shows that the cluster $L_4$ (with feasible capability $\mathbf{C}_4$) is the poorest one i.e. the highest priority in terms of policies intervention or policies implementation. This is confirmed by the result in Table 9: $\mathbf{M}_4(\mathcal{J}_u, \varepsilon_u) = \big\{(-0.89, 0.90);\ (-0.73, 0.91); (-0.11, 0.56); (-0.72, 0.91); (-0.06, 0.55); (-0.44, 0.76)\big\}$, where $u \in \{1, 2, 3, 4, 5, 6\}$ and $(-1.00, 1.00)$ (resp. $(1.00, 1.00)$) is the perfect score characterizing the cluster of poorest (resp. richest) households according to a given welfare dimension. Considering the representation (38), the households belonging to the cluster $L_4$ are more poor than the households belonging the cluster $L_2$. Looking at Table 9 , we note that the households belonging the cluster $L_2$ are more deprivated on welfare dimension 'Health' than all others ($\mathbf{M}_4(\mathcal{J}_5, \varepsilon_5) = (-0.36, 0.70)$). Therefore, one of the main advantages of our methodology is its capacity to determine among the clusters of households, the welfare dimensions on which people are most deprivated in order to target resources and design policies more effectively. For instance, considering the cluster of poorest households $L_4$, the highest priority in term of policies intervention has to be focused on welfare dimension 'housing'. It is true that the households of the clusters $L_4$ are deprivated on all welfare dimensions, but they are more deprivated on welfare dimension 'Housing' than all other welfare dimensions. Indeed, the welfare dimensions can be ordered (in term priority of policies intervention) as follows: Housing$\succ$Mobility$\succ$WS$\succ$Education$\succ$Nutrition$\succ$Health.

| | $\mathcal{J}_1$ | $\varepsilon_1$ | $\mathcal{J}_2$ | $\varepsilon_2$ | $\mathcal{J}_3$ | $\varepsilon_3$ | $\mathcal{J}_4$ | $\varepsilon_4$ | $\mathcal{J}_5$ | $\varepsilon_5$ | $\mathcal{J}_6$ | $\varepsilon_6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{M}_1$ | 0.84 | 0.96 | 0.78 | 0.95 | 0.85 | 0.93 | 0.80 | 0.93 | 0.65 | 0.90 | 0.75 | 0.89 |
| $\mathbf{M}_2$ | -0.78 | 0.85 | -0.43 | 0.74 | -0.06 | 0.53 | -0.82 | 0.85 | -0.36 | 0.70 | -0.41 | 0.68 |
| $\mathbf{M}_3$ | -0.35 | 0.66 | 0.29 | 0.63 | 0.29 | 0.64 | -0.34 | 0.64 | 0.10 | 0.56 | -0.13 | 0.57 |
| $\mathbf{M}_4$ | -0.89 | 0.90 | -0.73 | 0.91 | -0.11 | 0.56 | -0.72 | 0.91 | -0.06 | 0.55 | -0.44 | 0.76 |
| $\mathbf{M}_5$ | 0.64 | 0.88 | 0.68 | 0.93 | 0.68 | 0.84 | 0.73 | 0.90 | 0.84 | 0.97 | 0.61 | 0.83 |
| $\mathbf{M}_6$ | 0.38 | 0.73 | 0.45 | 0.79 | 0.70 | 0.85 | 0.48 | 0.75 | 0.84 | 0.94 | 0.42 | 0.74 |
| $\mathbf{M}_7$ | 0.94 | 0.95 | 0.97 | 0.97 | 0.85 | 0.87 | 0.84 | 0.95 | 1.00 | 1.00 | 0.99 | 0.99 |
| $\mathbf{M}_8$ | 0.92 | 0.97 | 0.89 | 0.98 | 0.55 | 0.83 | 0.96 | 0.98 | 0.86 | 0.99 | 0.90 | 0.98 |

Table 9: Example of meaningful measurements

*6.2.2. Analytical competence*

The analytical stage aims at understanding the commodities determining this situation. It leads to the description and characterization of different clusters which have been discovered during the cognitive stage. The analytical stage involves an analysis of data whose primary purpose is the highlighting of existing relationships between clusters, between the commodities that characterize them, and between clusters and variables which are generally difficult to detect in very large databases. For instance, considering the ASSL 2007 database of Burkina Faso (see Annexe 6), the 'worst' feasible capabilities (e.g. $\big\{\mathbf{C}_3;\ \mathbf{C}_2;\ \mathbf{C}_4\big\}$) are characterised by the fact that the head of household is *Farmer*, while the 'best' ones (e.g $\big\{\mathbf{C}_7;\ \mathbf{C}_8;\ \mathbf{C}_1\big\}$) are characterised by the fact that the head of household is *Public Sector Employees*. It is important to note that, the words 'worst' and 'best' do not contain anything of numerical but only reflect the household's well-being and advantage. This is crucial since it allows to understand each cluster from the household's commodities perspective, identifying possible deprivations and targeting the clusters to which we may tailor adequate policies. For instance the clusters $L_2$, $L_3$ and $L_4$ (which represent 66% of the population) are totally deprivated on three welfare dimensions (Housing, WS and Education). This means that, approximately 66% of the population need an improvement of the living conditions firstly on three welfare dimensions (i.e. 'Housing', 'WS' and 'Education').

$$\begin{aligned}
&\mathbf{C}_1 = \big\{\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_2 = \big\{\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_3 = \big\{\langle \mathbf{B},\mathbf{A},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{B},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_4 = \big\{\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_5 = \big\{\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{B}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_6 = \big\{\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{B}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_7 = \big\{\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}\\
&\mathbf{C}_8 = \big\{\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{G}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\big\}
\end{aligned}$$

Table 10: The set of feasible capabilities

Consider now the generic commodity 'income'[7] and the feasible capabilities $\mathbf{C}_5$ and $\mathbf{C}_6$ (see Table 10) that we would like to increase some of the attainable levels of functioning vector. A quick observation shows that the functioning thresholds $\mathbf{F}_{\sup}(L_5) = \mathbf{F}_{\sup}(L_6)$. One can be tempted to conclude that "the standard living (or welfare) offered by $\mathbf{C}_5$ is considered to be *equivalent* to the standard living (or welfare) offered by $\mathbf{C}_6$". However, Table 9 shows that only $\mathbf{M}_5(\mathcal{J}_5)$ and $\mathbf{M}_6(\mathcal{J}_5)$ are equal in terms of meaningful measurement with $X_2(L_5) = [90000,\ 110000]$ and $X_2(L_6) = [63000,\ 90000]$. Thus, we can intuitively increase the attainable levels of welfare dimension $\mathcal{J}_3$ from $\mathbf{A}$ to $\mathbf{G}$, i.e where $\mathbf{M}_5(\mathcal{J}_3) < \mathbf{M}_6(\mathcal{J}_3)$ in Table 9. The extended capabilities of the feasible capabilities $\mathbf{C}_5$ and $\mathbf{C}_6$ are then given by the sets described in 39 and 40.

$$\widetilde{\mathbf{C}}_5 = \big\{\langle \mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\ \cdots\cdots\cdots\big\} \tag{39}$$

$$\widetilde{\mathbf{C}}_6 = \big\{\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{B}\rangle;\ \cdots\cdots\cdots\big\} \tag{40}$$

Assuming that the decision maker has imposed some levels of control in such a way that the extended capabilities of the feasible capabilities $\mathbf{C}_7$ and $\mathbf{C}_8$ are given by

$$\widetilde{\mathbf{C}}_7 = \big\{\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{G}\rangle\big\}$$

and

$$\widetilde{\mathbf{C}}_8 = \big\{\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{G}\rangle\big\}.$$

$$\begin{aligned}
&\widetilde{\mathbf{C}}_1 = \{\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A}\rangle;\ \cdots\ \langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\}\\
&\widetilde{\mathbf{C}}_2 = \{\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\}\\
&\widetilde{\mathbf{C}}_3 = \{\langle \mathbf{B},\mathbf{A},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{B},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{A},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle;\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\}\\
&\widetilde{\mathbf{C}}_4 = \{\langle \mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B},\mathbf{B}\rangle\}\\
&\widetilde{\mathbf{C}}_5 = \{\langle \mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\ \cdots\cdots\cdots\}\\
&\widetilde{\mathbf{C}}_6 = \{\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A}\rangle;\langle \mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{A},\mathbf{B}\rangle;\ \cdots\cdots\cdots\}\\
&\widetilde{\mathbf{C}}_7 = \{\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{G}\rangle\}\\
&\widetilde{\mathbf{C}}_8 = \{\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{G}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{G},\mathbf{G},\mathbf{A}\rangle;\langle \mathbf{G},\mathbf{G},\mathbf{A},\mathbf{A},\mathbf{G},\mathbf{G}\rangle\}
\end{aligned}$$

Table 11: The set of extended capabilities

Thus, the extended capabilities are given as follows:

Therefore, considering the ASSL 2007 database of Burkina Faso and suppose that the decision maker establishes a set of actions, let say: house improvement programme (such as sanitation infrastructure supply, water supply, power supply), education programme, nutrition improvement programme, health improvement programme, mobility improvement programme, support micro credit. The clusters of households $\{L_2, L_4\}$ (with respectively the extended capabilities $\{\widetilde{\mathbf{C}}_2, \widetilde{\mathbf{C}}_4\}$) may benefit from all improvement programmes, while the cluster of households $\{L_3\}$ (with extended capabilities $\{\widetilde{\mathbf{C}}_3\}$) does not need the mobility improvement programme, the nutrition improvement programme and the micro credit support. Indeed, the households of cluster $\{L_3\}$ are acceptably transported and acceptably nourished households according to the results of Table 11. A similar analysis can be easily widened on all other clusters and all welfare dimensions in such a way to assess the effectiveness of various possible policies evaluating whether a particular policy will produce a positive impact on the welfare of the people in the future when that policy will be translated into concrete actions.

[7] Note that 'income' in ASSL 2007 is given in CFA franc. The CFA franc is the name of two currencies used in Africa: 1 euro = 655.957 CFA.

| $X_i$ | **Description: range** |
|---|---|
| $X_1$ | Size of Household: 0, 1, 2, . . . |
| $X_2$ | Monthly Income for Basic Needs of Household: $> 0$ |
| $X_3$ | Age (in years): $\geq 0$ |
| $X4$ | HhWEIGHT: Household Weight ($> 0$) |
| $X_5$ | Occupational Status of Household Head: Public Sector Employee(Pu), Private Sector Employee(Pr), Employee of the Informal Private(I), Farmer(F), Cotton Farmer(C), Other Type of Agriculture (O), Familial help-Volunteer-Apprentice(FVA), Inactive(N), Jobless(J) |
| $X_6$ | Gender: Male(M), Female(F) |
| $X_7$ | Marital Status: Single(S), Monogamous(M), Polygamous(P), Widower(W), Divorced(D), Common-Law(C) |
| $X_{36}$ | General Economic Situation of the Community: Worse Now(W), Bad Now(Ba), Unchanged(U), Better Now(Be), More Better Now(M) |
| $X_{37}$ | How many people most contribute to the Household Expenses: 0, 1, 2, . . . |
| | **Housing** |
| $X_8$ | Housing Tenure Status: Homeowner(H), Leaseholder(L), occupier rent free(O) |
| $X_9$ | Number of Separate Rooms: 0, 1, 2, . . . |
| $X_{10}$ | Has a Untreated Net: Yes(Y), No(N) |
| $X_{11}$ | Has a Treated Net: Yes(Y), No(N) |
| $X_{12}$ | Has an Iron: Yes(Y), No(N) |
| $X_{13}$ | Has an Improved Stove: Yes(Y), No(N) |
| $X_{14}$ | Has a Fridge/Freezer: Yes(Y), No(N) |
| $X_{15}$ | Has a Television: Yes(Y), No(N) |
| $X_{16}$ | Has a Bed/Mattress: Yes(Y), No(N) |
| $X_{17}$ | Has a Modern Cooker: Yes(Y), No(N) |
| $X_{18}$ | Has a Computer: Yes(Y), No(N) |
| $X_{19}$ | Has Electricity: Yes(Y), No(N) |
| $X_{20}$ | Is There a Room Equipped for Cooking?: Yes(Y), No(N) |
| $X_{21}$ | Has a Landline: Yes(Y), No(N) |
| $X_{22}$ | Has a Mobile Phone: Yes(Y), No(N) |
| $X_{23}$ | Area of Residence: Developed Area(D), Undeveloped Area(U) |
| $X_{24}$ | Type of Housing: Apartment Building(A), Villa(V), Single Individual House(S), Multiple Occupancy Building(M), Traditional House(T), Other(O) |
| $X_{25}$ | Soil Materials of the Main Building: Tile(T), Cement(C), Sand/Clay(S), Other(O) |
| $X_{26}$ | Wall Materials of the House: Mud Brick/Earth(E), Stones(S), Bricks(B), Cement/Concrete(C), Wood/Bamboo(W), Metal Sheet(M), Wall Straw(WS), Other(O) |
| $X_{27}$ | Materials of the Roof of the House: Earth(E), Straw(S), Bricks(B), Metal Sheet(M), Cement/Concrete(C), Tiles(T) |
| $X_{28}$ | Main Energy Sources for Cooking: Firewood with Improved Stove(FI), Firewood with Single Stove(FS), Charbon with Improved Stove(CI), Charbon with Single Stove(CS), Kerosene/Oil(K), Gas(G), Electricity(E), Harvest Residue/Sawdust(H), Animal Waste(A), Other(O) |
| $X_{29}$ | Main Energy Sources for Lighting: Kerosene/Oil(K), Gas(G), Electricity(E), Solar Energy(S), Generator(Ge), Battery Torch(B), Lamp Loadable/Batteries(L), Candles(C), Other(O) |
| | **Mobility** |
| $X_{30}$ | Owner a Bicycle: Yes(Y), No(N) |
| $X_{31}$ | Owner of a Motorcycle: Yes(Y), No(N) |
| $X_{32}$ | Owner of a Car or a Lorry: Yes(Y), No(N) |
| $X_{33}$ | Time Taken to Reach the Nearest Public Transport (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| | **Nutrition** |
| $X_{34}$ | Has had Problems to Meet Food Needs: Never(N), Rarely(R), Sometimes(S), Often(O), Always(A) |
| $X_{35}$ | General Economic Situation of the Household: Worse Now(W), Bad Now(Ba), Unchanged(U), Better Now(Be), More Better Now(M) |
| $X_{38}$ | Time Taken to Reach the Market of Food Nearest (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| | **Water and Sanitation (WS)** |
| $X_{39}$ | Main Source of Drinking Water: Barrage-River-Lake(B), Ordinary Wells(O), Closed Wells with Nozzles(C), Single Wells with Nozzles(S), Drilling(D), Public Fountain(P), Indoor Tap(I), Indoor Tap Shared (T), Other(O) |
| $X_{40}$ | Type of Toilets used: Septic Tank(S), Traditional Latrines Rehabilitated(T), VIP Tank(V), Latrine or Toilet Flush(L), Ordinary Latrines(OL), In the Nature(I), Other(O) |
| $X_{41}$ | Mode of Garbage Disposal: Dustbin(removal by service)(D), Public Garbage(P), Individual Garbage(I), Pits(Pi), Ferry(F), Road/Street(R), Other(O) |
| $X_{42}$ | Mode of Sewage Disposal: In the Courtyard(C), In the Street/Road(S), In the Pit(P), Into Public Drains(D), In the Nature(I), Other(O) |
| $X_{43}$ | Time Taken to Reach the Closest Source of Cooking Water (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| $X_{44}$ | Time Taken to Reach the Closest Source of Drinking Water (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| | **Health** |
| $X_{45}$ | Time Taken to Reach the Closest Health Service (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| | **Education** |
| $X_{46}$ | Time Taken to Reach the Closest Primary School (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| $X_{47}$ | Time Taken to Reach the Closest Secondary School (in minutes): $[0;14]$ =Very Close(V), $[15;29]$ =Acceptably Close(A); $[30;44]$ = Close(C); $[45;59]$ =Far(F), 60+ =Far Away(FA) |
| $X_{48}$ | Level of Education of Household Head: Has Never Been to School(N), Primary Not Completed(PN), Primary Completed(PC), Secondary Not Completed(SN), Secondary Completed(SC), Higher School(H), Adult Literacy(A) |

Table 12: Description of criteria