Abstract: Clustering & Hierarchical Disjoint Non-negative Factor Analysis (CHDNFA) is a new latent factor model here proposed to model a Composite Indicator (CI) and simultaneously, cluster individuals according to the CI.

In general, CIs are multidimensional concepts described by at least a theoretical construct (factor) related to a set of measured variables.

Frequently, CIs have a hierarchical structure, i.e., they are characterized by a set of factors -representing specific CIs (also named dimensions)- each one loading exclusively on a disjoint subsets of variables; then there is a general CI loading on all specific CIs. Clustering & Hierarchical Disjoint Factor Analysis is here proposed to model the hierarchical structure of factors that is supposed in part or totally unknown, identifying a reduced set of factors each one related to a disjoint subset of variables.

Simultaneously the best clustering of the units is required according to the identified CI. The clustering is useful to identify classes of units that have significant different CI. The ranking of clusters allows to find those classes of units that clearly differ from each other.

Furthermore, in a definition of a CI each subset of measured variables used to describe a multidimensional concept must be internally consistent and reliable, that is, variables related to the factor measure "consistently" a unique theoretical construct. This means that variables are concordant with

the related factor and loadings must be positive. This last requirement is included as a constraint in the new methodology that is named Clustering and Hierarchical Disjoint Non-negative Factor Analysis (CHDNFA).

Properties are discussed for CHDNFA. Cross-loadings can also be estimated to increase the fit of the factor model starting from the best CHDNFA solution.

CHDNFA has also the option to constraint a variable to load on a pre-specified factor in order to hypothesize, a priori, some relations between variables and loadings.

A simulation study shows performances of CHDNFA and an application to optimally identify the dimensions of well-being is used to illustrate the characteristics of the new methodology. A final discussion completes the paper.

Keywords: Clustering & Hierarchical Factor Analysis model, Disjoint Factor

Analysis, Sparse loading matrix.