New tools for the construction of ranking and evaluation indicators in multidimensional systems of ordinal variables

Filomena Maggino[1], Marco Fattore[2]
[1]Università degli Studi di Firenze (Italy), e-mail: filomena.maggino@unifi.it
[2]Università degli Studi di Milano-Bicocca (Italy), e-mail: marco.fattore@unimib.it

Abstract

The problem of ranking and evaluation in multidimensional ordinal datasets is one of the most important issue in applied statistics, particularly in the socio-economic field.
Unfortunately, dealing with ordinal variables raises many conceptual and methodological issues, particularly when consistent and meaningful indicators are to be defined out of qualitative data.
Methodological difficulties rise particularly when single ordinal indicators are to be aggregated into a composite indicator, to get unidimensional scores for comparing and ranking statistical units.
Basically, it can be asserted that the issue of ranking and evaluation in an ordinal setting is still an open problem, since the statistical methodologies, applied in the common practice or proposed at theoretical level, are unsatisfactory in many respects.
Motivated by these issues and by the relevance of the topic, in this paper, we introduce new tools for addressing the construction of indicators for ranking and evaluation purposes in an ordinal context, with the aim to overcome the main problems of the classical composite indicator approach.
The proposed methodology draws upon Partial Order SET (POSET) theory, a branch of discrete mathematics providing concepts and tools, fitting very naturally the needs of ordinal data analysis.
POSET theory provides useful technical tools for addressing the evaluation problem. But more important than this, it helps reformulating the ranking and evaluation problem in such a way that satisfactory solution to the issues outlined above can indeed be worked out, fully respecting the qualitative features of data.

Keywords:
Partial order – Composite indicators – Ordinal variables

1. Introduction

The present debate on well-being measurement is clearly pointing out that a valuable evaluation process has to take into account many different and complementary aspects, in order to get a comprehensive picture of the problem and to effectively support decision-making. Assessing well-being requires sharing a conceptual framework about its determinants and about society and needs the identification of the most consistent and effective methodologies for building indicators and for communicating purposes. From a statistical perspective, one of the critical points concerns the preservation of the true nature of the social economic phenomena to be analyzed. This calls for an adequate methodological approach.

Several socio-economic phenomena have an intrinsic ordinal nature (e.g. material deprivation, democratic development, employment status, and so on) and correspondingly there has been an increasing availability of ordinal datasets. Nevertheless, ordinal data have been often conceived as just a rough approximation of truly numerical and precise, yet non observable, features, as if a numerical latent structure would exist under ordinal appearances. As a result, the search for alternative statistical procedures has been slowed down and many epistemological, methodological and statistical problems regarding ordinal data treatment are still open and unsolved.

1. Methodological approaches: between objectivity, subjectivity and arbitrariness.
The epistemological research of the last century has focused on the role of the subject in knowledge production and has clearly showed how pure objectivism cannot account for the knowledge process, even in scientific disciplines. This is particularly evident when observing and analysing socio-economic phenomena. Given the complexity and the nuances of socio-economic issues, data can often be considered as a (fragmented) ‘’text” to be ‘’read” by the researcher, in search for a ‘’sense” and a structure in it. This ‘’sense structuring” process is not an arbitrary one, but necessarily involves some subjectivity. To make an example, think about the issue of defining poverty thresholds in deprivation studies, both in a monetary and in a multidimensional setting, with the consequences that different choices have in the final picture. Admittedly, in many applied studies subjectivity is generally felt as an issue to be removed and many evaluation procedures are designed to accomplish this task. Ironically, removing subjectivity is not an objective process and often produces arbitrary results. Thus, it is important to distinguish between a necessary ‘’objectivity” of the research methodology (e.g. observation and data collection procedures) and an unavoidable ‘’subjectivity” related, for instance, to the definition and choice of the conceptual framework and the analytical approaches. The real methodological issue is not removing subjectivity; rather, it is building a sound statistical process, where subjective choices are clearly stated and their consequences can be clearly worked out in a formal and unambiguous way.

2. Ordinal data: between accuracy and ambiguity. A great part of the methodological and statistical efforts has been dedicated to the issue of making measures quantitatively more precise. In practice, this has often been turned in applying multivariate statistical tools to ordinal data, after transforming, or interpreting, them in cardinal terms, through more or less sophisticated scaling procedures. These procedures may sometimes lead to useful results, but they are often quite questionable, not being consistent with the intrinsic nature of data. De facto, the efforts for getting more precise measures have the effect of frequently forcing the true nature of socio-economic phenomena. On the contrary, it could be wise to realize that the great part of socio-economic phenomena is characterized by nuances and ‘’ambiguities”, which are not obstacles to be removed, but often represent what really matters.

3. Ordinal data: technical issues. Transformed or not in quantitative terms, ordinal data are generally submitted to traditional statistical tools, typically designed for quantitative data analysis and usually based on the analysis of linear structures. The results are quite arbitrary and questionable, since the data are forced into a conceptual and technical framework which is ultimately poorly consistent. Although these problems are well-known, and new methodologies are continuously being developed, they are still unsolved. Basically, it can be asserted that the issue of ranking and evaluation in an ordinal setting is still an open problem, even from a pure data treatment point of view.

Motivated by these issues and by the relevance of the topic, in this paper we introduce new tools for ranking and evaluation of ordinal data, with the aim to overcome the main problems of the classical methodologies and, particularly, of the composite indicator approach. We address the evaluation problem through a benchmark approach. Each statistical unit in the population is described in terms of its profile, which is defined in terms of the sequence of scores on the evaluation dimensions; profiles are then assessed against some reference sequences, chosen as benchmarks, to get the evaluation scores. We address the comparison of profiles to benchmarks in a multidimensional setting by using tools and results from partially ordered set theory (POSET theory, for short).

Indeed, through POSET tools, sequences of scores can be assessed without involving any aggregation of the underlying variables, since the evaluation is performed by exploiting the relational structure of the data, which involves solely the partial ordering of the profiles. The remainder of the paper is organized as follows.

Section 2 gives a brief account of the composite indicator approach, highlighting its main criticalities, particularly in the ordinal case. Section 3 introduces a few basic concepts from POSET theory. Section 4 describes the basic evaluation strategy and the procedure to compute the evaluation scores. Section 5 tackle the problem of ‘’weighting” evaluation dimensions. Section 6 concludes. The aim of the paper is primarily methodological, leaving to forthcoming works the systematic application of the evaluation procedure to real data.

2. The composite indicator approach and its critical issues

Addressing the complexity of socio-economic phenomena for evaluation aims is a complex task, often requiring the definition of large systems of indicators. Frequently, the complexity of the indicator system itself leads to the need of computing composite indicators in order to (Noll, 2009):

- answer the call by ‘’policy makers” for condensed information;
- improve the chance to get into the media;
- allow multi-dimensional phenomena to be synthetized;
- allow easier comparisons across time;
- compare cases (e.g. nations, cities, social groups. . .) in a transitive way (e.g. through rankings).

Despite its spreading, the composite indicator approach is currently being deeply criticized as inappropriate and often inconsistent (Freudenberg & Nardo, 2003). Critics point out conceptual, methodological and technical issues, especially concerning the difficulty of conveying into unidimensional measures, all the relevant information pertaining to phenomena which are complex, dynamic, multidimensional and full of ambiguities and nuances. The methodology aimed at constructing composite indicators is very often presented as a process needing specific training, to be performed in a scientific and objective way. Actually the construction procedure, even though scientifically defined, is far from being objective and aseptic. Generally, it comprises different stages (Nardo et al., 2005; Sharpe & Salzman, 2004), each introducing some degree of arbitrariness to make decisions concerning:

- the analytical approach to determine the underlying dimensionality of the available elementary indicators and the selection of those to be used in the evaluation process;
- the choice of the weights used to define the importance of each elementary indicator;
- the aggregation technique adopted to synthesize the elementary indicators into composite indicators.

Indicator selection. Selecting the indicators to be included in the composite represents a fundamental stage in the construction process since it does operationally define the latent concept that the composite is supposed to measure. Selection criteria should consider (Nardo et al., 2005) the issues of reducing redundancies, allowing both comparability among statistical units and over time and should be oriented to obtaining politically relevant results. From a statistical point of view, indicator selection often involves a principal component analysis or a factor analysis, to reveal correlations and associations among evaluation variables and to perform some dimensionality reduction. Irrespective of the statistical tool adopted, dimensionality reduction raises some relevant questions, concerning its consequences on the composite indicator construction. If the concept to be measured turns out to be actually unidimensional, computing a single composite indicator could be justifiable. But when concepts are truly multidimensional, then singling out just one, albeit composite, indicator is very questionable. The nuances and ambiguities of the data would in fact be forced into a conceptual model where all the features conflicting with unidimensionality are considered as noise to be removed. Moreover, synthetic scores could be biased towards a small subset of elementary indicators, failing to give a faithful representation of the data.

Weighting variables. When constructing composite indicators, particular attention is paid to the weighting process, which gives different importance to the elementary indicators forming the composite. The necessity of choosing weights based on objective principles is frequently asserted (Nardo et al., 2005; Ray, 2008; Sharpe & Salzman, 2004), leading to a preference for statistical tools like correlation analysis, principal component analysis or data envelopment analysis, to mention a few. However, adopting purely statistical methods in the weighting process must be carefully considered. Removing any control over the weighting procedure from the analyst, gives a possibly false appearance of objectivity that is actually difficult to achieve in social measurement (Sharpe & Salzman, 2004). Moreover, since defining weights is often interpreted in the perspective of identifying personal and social values, the procedure should necessarily involve individuals' judgments. If indicators concern societal well-being, their construction turns out to be not just a technical problem, being part of a larger debate aimed at obtaining a larger legitimacy. In this perspective, the weighting issue can be even considered as a leverage of democratic participation to decisions. For example, Hagerty and Land (2007) stresses that building composite indicators should take into account and maximize the agreement among citizens concerning the importance of each elementary indicator. Choosing consistent weighting criteria is thus a critical issue, largely subjective and possibly data independent.

Aggregating indicators. Further criticisms concern the aggregation process (Munda & Nardo, 2008), needed to get unidimensional scores out of multidimensional data, and which raises methodological difficulties when dealing with ordinal data. The process is in fact quite controversial since:

– the indicators to be aggregated are rarely homogeneous and need not share common antecedents (Howell et al., 2007);
– the aggregation technique might introduce implicitly meaningless compensations and trade-offs among evaluation dimensions;
– it is not clear how to combine ordinal variables, using numerical weights.

Even using scaling tools, turning ordinal scores into numerical values, is not satisfactory; it forces the nature of the data and is not definitely a clear process, since different choices of the scaling tools may imply very different final results. Composite indicators represent the mainstream approach to socio-economic evaluation, yet the discussion above shows how many critical issues affect their computation. The difficulties are even greater when ordinal variables are dealt with, since statistical tools based on linear metric structures can be hardly applied to non-numeric data. In a sense, socioeconomic analysis faces an impasse: (i) implicitly or not, it is generally taken for granted that ''evaluation implies aggregation"; thus (ii) ordinal data must be scaled to numerical values, to be aggregated and processed in a (formally) effective way; unfortunately (iii) this often proves inconsistent with the nature of the phenomena and produces results that may be largely arbitrary, poorly meaningful and hardly interpretable. Realizing the weakness of the outcomes based on composite indicator computations, statistical research has focused on developing alternative and more sophisticated analytic procedures1, but almost always assuming the existence of a cardinal latent structure behind ordinal data. The resulting models are often very complicated and still affected by the epistemological and technical issues discussed above. The way out to this impasse can instead be found realizing that evaluation need not imply aggregation and that it can be performed in purely ordinal terms. This is exactly what POSET theory allows to do.

3. Basic elements of partial order theory

A partially ordered set (or a POSET) $P = (X, <)$ is a set X (called the *ground set*) equipped with a partial order relation $<$, that is a binary relation satisfying the properties of *reflexivity*, *antisymmetry* and *transitivity* (Davey and Priestley, 2002):

1. $x < x$ for all x in X (reflexivity);
2. if $x < y$ and $y < x$ then $x = y$ (antisymmetry);
3. if $x < y$ and $y < z$, then $x < z$ (transitivity).

If $x < y$ or $y < x$, then x and y are called *comparable*, otherwise they are said *incomparable* (written $x \| y$). A partial order P where any two elements are comparable is called a *chain* or a *linear order*. On the contrary, if any two elements of P are incomparable, then P is called an *antichain*. A finite POSET P (i.e. a POSET over a finite ground set) can be easily depicted by means of a Hasse diagram, which is a particular kind of directed graph, drawn according to the following two rules: (i) if $s < t$, then node t is placed above node s; (ii) if $s < t$ and there is no other element w such that $s < w < t$ (i.e. if t *covers* s), then an edge is inserted linking node t to node s. By transitivity, $s < t$ (or $t < s$) in P, if and only if in the Hasse diagram there is a descending path linking the corresponding nodes; otherwise, s and t are incomparable.

4. Representing and evaluating ordinal data through POSETs

Let $v_1,\dots,v_k$ be k ordinal evaluation variables. Each possible sequence s of scores on $v_1,\dots,v_k$ defines a different *profile*. Profiles can be (partially) ordered in a natural way, by the dominance criterion given in the following definition:

Definition 1. Let s and t be two profiles over $v_1,\dots,v_k$ we say that t dominates s (written $s \blacktriangleleft t$) if and only if $v_i(s) \leq v_i(t)$ for all $i = 1,\dots,k$ and there is an index j such that $v_j(s) \leq v_j(t)$ where $v_h(s)$ and $v_h(t)$ are the scores of s and t on v_h.

Clearly, not all the profiles can be ordered based on the previous definition; as a result, the set of profiles gives rise to a POSET (the *profile* POSET). Figure 1 reproduces the Hasse diagram of a profile POSET built on 5 binary variables, referring to the ownership of 5 different goods (0 – ownership; 1 – non-ownership). The top node (i.e. profile 11111) represents complete deprivation; on the opposite, the bottom node (i.e. profile 00000) represents no deprivation at all.

Since ordinal phenomena cannot be measured against an absolute scale, the evaluation scores to be assigned to the profiles are computed comparing them against some reference profiles selected as benchmarks. With reference to the deprivation POSET of Figure 1, let us assume that profile 01000 represents a “certainly non-deprived” configuration, so that its deprivation score is set to 0 (the minimum deprivation degree). As a consequence, also profile 00000 will be scored as 0. At the opposite, suppose that both profile 01110 and profile 11001 represent deep deprivation configurations, receiving deprivation score 1 (the maximum deprivation degree). Then, also profiles 11110, 01111, 11101, 11011 and 11111 will get score 1 (since they dominate 01110 or 11001). The set W of profiles receiving score 0 and the set D of profiles receiving score 1 are depicted as black nodes in the Hasse diagram of Figure 1. As it can be seen, benchmarks share a role similar to that of thresholds in classical evaluation studies, as explained in mode details in Fattore et al. (2011). Thus, we will call w = (01000) and d = (01110, 11001) the *inferior* and the *superior* threshold respectively.

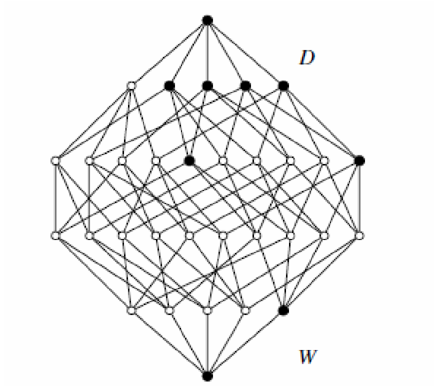

Figure 1. Hasse diagram of the deprivation POSET built on 5 binary variables.

Profiles not belonging to D or W are not “certainly deprived”, neither “certainly non-deprived”, so they will consistently be assigned deprivation scores in (0,1). The scores are computed based on the “position” of the profiles in the profile POSET, with respect to the benchmarks. To perform such a computation, the notion of *linear extension* of a POSET is to be introduced. A linear extension of a POSET P is a linear ordering of the elements of P which is consistent with the constraints given by the partial order relation. For example, if P is composed of three elements x, y and z, with $y < x$, $z < x$ and $y \parallel z$, only two linear extensions are possible, namely $z < y < x$ and $y < z < x$, since x is greater than both y and z in P. The set of all the linear extensions of a POSET P comprises all the linear orders compatible with P and identifies uniquely the partial order structure (Neggers & Kim, 1998; Schroeder, 2003). Profiles in D are ranked, in any linear extension of the profile POSET, above at least one element of the superior threshold d. Similarly, profiles in W are ranked, in any linear extension of the profile POSET, below at least one element[1] of the inferior threshold w. All other profiles are ranked, in the linear extensions, over elements of d, below elements of w or in between the thresholds, with different frequencies, depending on their position in the profile POSET. If a profile s is ranked, in a linear extension, over elements of d, in *that* linear extension it receives score equal to 1; if it is ranked below elements of w, it receives score equal to 0; if it is ranked in between, it receives score 0.5, to account for the assessing uncertainty. Averaging[2] over the set of all the linear extensions of the profile POSET, the final evaluation score is computed for profile s and, similarly, for any profile in P. The corresponding evaluation function for the POSET depicted in Figure 1, given the thresholds w = (01000), d = (01110, 11001), is reported in Figure 2(a). As it can be seen, the evaluation function is not just a function of the number of deprivation in the profile. The scores reproduce the nuances of deprivation more faithfully than a simple counting approach as in Cerioli and Zani (1990) and this shows how the POSET approach is effective in extracting information out of the data structure.

[1]In general, any threshold can be composed of more than one profile, even if in the example discussed in the text the inferior threshold reduces to a singleton.
[2] The choice of a simple average can be maintained based on the classical axiomatic theory of means.

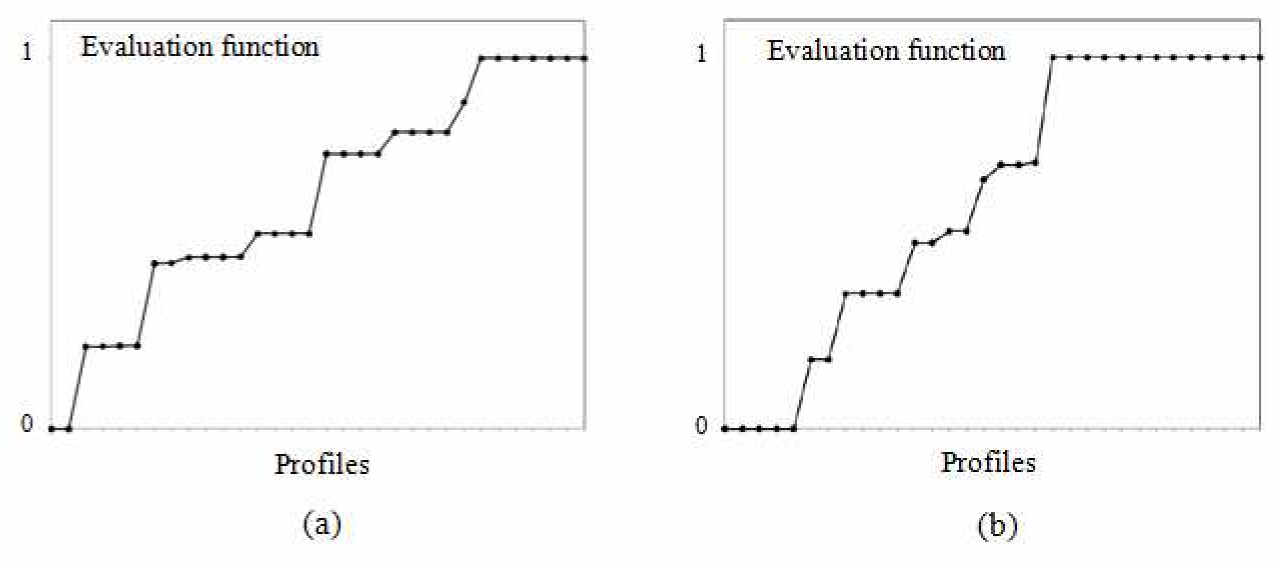

Figure 2. Evaluation functions for the original (a) and the extended (b) deprivation POSET, given the thresholds w = (01000) and d = (01110, 11001). Profiles are ordered on the x axis, according to increasing deprivation scores.

5. The "weighting" problem

The methodology introduced in the previous paragraph assumes the evaluation dimensions to share the same relevance. As a matter of fact, some asymmetry among the dimensions is only implicitly introduced when the thresholds are identified. As a legacy of the composite indicator methodology, the problem of accounting for the relevance of the evaluation dimensions is usually tackled using numerical weights, even in an ordinal setting (Cerioli and Zani, 1990; Lemmi & Betti, 2006). As a matter of fact, an alternative and more consistent solution comes from POSET theory. However, before introducing it, the weighting problem must be carefully reconsidered. Generally, weighting schemes are introduced in order to improve the informative content of the analysis and to reduce ranking ambiguities (often weights are computed through a principal component analysis, so as to maximize the variance, i.e. the informative power of the final index). Ambiguity reduction is the key to address the weighting problem also in an ordinal setting. The profile POSET P, built as described in Section 4, comprises only those comparabilities which are implied by the purely logical ordering criterion stated in Definition 1. Still, many ambiguities, i.e. incomparabilities, remain in P, since the ordering criterion is not enough informative to ''resolve" all of them. Adding information to the evaluation procedure should therefore yield a reduction of the set of incomparabilities in the profile POSET. This idea can be formally stated, through the concept of extension of a partial order.

Definition 2. Let $P_1 = (X, <_1)$ and $P_2 = (X, <_2)$ be two POSETs over the same ground set X. If a $<_1$ b implies a $<_2$ b, for any a, b in X, then P_2 is called an extension of P_1.

Therefore, "weighting" the profile POSET means extending it according to some exogenous criterion, suggesting which incomparabilities must be turned into comparabilities. Unfortunately, when new comparabilities are added to a POSET, the resulting set P' is in general not a POSET, since transitivity need not be fulfilled. To restore transitivity, all the comparabilities implied by those added to P must be also added. This way, the required extension of the profile POSET is obtained as the so-called transitive closure P* of P'. It is important to realize that the transitive closure of P' is the *smallest* POSET comprising the comparabilities of P *and* those added to it. In

general, other extensions of P exist comprising P', but all of them would contain comparabilities not implied by P'. In this sense, choosing one of them would be arbitrary.
With reference to the deprivation example, let us suppose that goods 2, 3 and 4 are considered as more important than goods 1 and 5, so that possessing (say) good 1 does not compensate for not possessing good 4. Based on this assumption, some incomparabilities of the original deprivation POSET P can indeed be turned into comparabilities and added to P itself. Through the transitive closure device, the minimal extension P* of P can be readily[3] obtained; its Hasse diagram is depicted in Figure 3, together with sets D and W. As it can be seen, the extended POSET comprises much less incomparabilities that the original profile POSET, that is, it is much more informative on the relative deprivation degree of the profiles. The evaluation function is then computed using the same procedure described previously. The rank distributions of the profiles over the set of linear extensions is affected by the addition of information to the profile POSET, resulting into a much more polarized evaluation function, as it can be seen in Figure 2(b). As expected, the added information has reduced the ambiguity of the original partial order, resulting in a much steeper evaluation function.

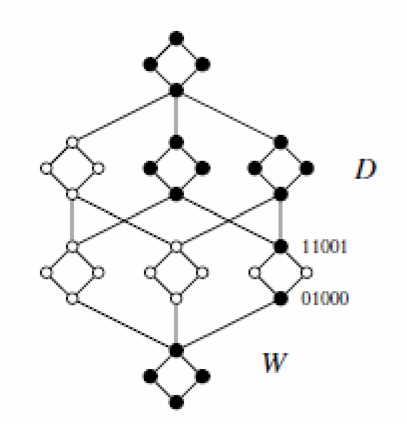

Figure 3. Hasse diagram of the extended deprivation POSET built on 5 binary variables.

6. Conclusion

In this paper, we have outlined a new methodology for evaluation purposes in multidimensional systems of ordinal data. The methodology draws upon POSET theory, so as to overcome the conceptual and computational drawbacks of the standard aggregative procedures, which involve composite indicators. Poset tools allow to describe and to exploit the relational structure of the data, so as to compute evaluation scores in purely ordinal terms, avoiding any aggregation of variables. The effectiveness of the partial order approach is particularly evident in the way the ''weighting" problem is addressed and solved, avoiding the introduction of numerical weights in the computations. Although simplified, the examples discussed in the paper show how the methodology can be applied in practice and to real datasets. The software routines needed for the computations can in fact be easily implemented through standard programming languages. As any novel proposal, our methodology can be improved in many respects and extended in many directions, both at theoretical and applied level. These are interesting perspectives for future research.

[3] Computing the transitive closure is a trivial task, using the matrix representation of POSETs (Patil and Taillie, 2004).

References

Cerioli A., Zani S. (1990) "A Fuzzy Approach To The Measurement Of Poverty" in: Dagum C., Zenga M. (eds.) *Income and Wealth distribution, Inequality and Poverty*, Springer-Verlag, Berlino Heidelberg, 272-284

Davey B. A., Priestley B. H. (2002) *Introduction to Lattices and Order*, CUP, Cambridge.

Freudenberg M., Nardo M. (2003) *Composite Indicators of Country Performance: A Critical Assessment*, OECD Science, Technology and Industry Working Papers, 2003/16, OECD.

Fattore M., Brueggemann R., Owsinski J. (2011, in press) "Using POSET theory to compare fuzzy multidimensional material deprivation across regions" in: Ingrassia S., Rocci R., Vichi M. (eds.) *New Perspectives in Statistical Modeling and Data Analysis*, Springer-Verlag.

Hagerty, M.R., Land, K.C. (2007) "Constructing Summary Indices of Quality of Life: A Model for the Effect of Heterogeneous Importance Weights", *Sociological Methods and Research*, Vol. 35, n. 4, 455-496.

Howell R. D., Breivik E., Wilcox J. B.,(2007) Reconsidering formative measurement, *Psychological Methods*, 12 n. 2, pp. 205-218.

Lemmi A., Betti G. (2006) *Fuzzy Set Approach to Multidimensional Poverty Measurement*, Springer, New York.

Munda G., Nardo M. (2008) Noncompensatory/non linear composite indicators for ranking countries: a defensible setting, *Applied Economics*, Vol. 99999, No. 1, pp. 1-11.

Nardo M., M. Saisana, A. Saltelli and S. Tarantola (EC/JRC), A. Hoffman and E. Giovannini (OECD) (2005) *Handbook on Constructing Composite Indicators: Methodology and Userguide*, OECD, Statistics Working Paper.Neggers J., Kim H. S. (1988) *Basic POSETs*, World Scientific, Singapore.

Noll H.-H. (2009) *Measuring and Monitoring the Quality of Life*, Lecture at the Università degli Studi di Firenze, April 23-24, 2009, available at http://www.starse.it/joomla/it/seminari-e-workshop

Ray, A. K. (2008) "Measurement of social development: an international comparison", *Social Indicators Research*, 86, n.1, 1-46.

Schroeder B. (2003) *Ordered sets*, Birkhäuser, Boston.

Sharpe A., J. Salzman (2004) *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*, Center for the Study of Living Standards, Ottawa, CAN.