Italian Welfare Expenditures: a compositional analysis in a three-way perspective

V. Simonacci[1], M.A. Di Palma[2], and M. Gallo[3]
[1]University of Naples L'Orientale - DISUS, Naples, Italy; *vsimonacci@unior.it*
[2]University of Naples L'Orientale - DISUS, Naples, Italy
[3]University of Naples L'Orientale - DISUS, Naples, Italy

Abstract

A considerable part of the welfare system in Italy is represented by the social assistance expenditures. The intervention categories of service users are supplied by municipalities, following the regions guidelines, and concerning seven areas: 'families and children', 'disabled', 'addictions', 'elderly', 'poverty and social inclusion', 'immigrants and nomads', 'multi-purpose activities'. Expenditures are expressed as absolute values and as a percentage of national GDP. The case study presented relates to the investigation of welfare expenditures, collected by survey, aggregated by 21 regions in 7 social services along a timeline of 10 years (2003 - 2012). The compound of the three aspects, observations, variables and occasions can be arranged in a tridimensional array. In order to model these data without losing the multi-fold structure of their variability, each mode of the analysis should be kept separate. This can be achieved by employing multilinear techniques such as the CANDECOMP/PARAFAC (CP) model (Carroll and Chang, 1970; Harshman, 1970). When multidimensional data are compositional, additional difficulties arise. Compositional parts do not vary independently of each other and are characterized by a biased covariance structure (Aitchison, 1999; Pawlowsky-Glahn et al., 2015). Like most statistical procedures, the CP model was not designed for dealing with bounded values and cannot be used to model three-way compositional data directly.

Thus far, the problem of fitting the CP Model to CoDa has been formalized in a core methodology only in reference to the most widely used CP algorithm, PARAFAC-ALS, by focusing on key problems such as preprocessing data as log-ratio coordinates and interpretation of results (Gallo, 2013; Engle et al., 2014; Gallo and Simonacci, 2013). However, other estimating methods have been proposed in statistical literature for the CP model, all characterized by different merits and shortcomings. Several comparative studies (Faber et al., 2003; Tomasi and Bro, 2006) have shown that depending on their objective functions these procedures respond differently to specific features of the analyzed data.

In this perspective the main focus of this contribution is to investigate how the expenditure for each component emerges in the overall sources accounted by single regions and how it modifies along a time span. Furthermore, the current research on multilinear models for CoDa is expanded by investigating and comparing the performance of expressing the welfare expenditures as a compositional problem or as a full size problem reconstructing the magnitude of the data (Pawlowsky-Glahn et al., 2015) and thus applying different fitting algorithms. Specifically, the accuracy of the solution, occurrence of degeneracies, sensitivity to over-factoring, data multicollinearity, factor collinearity and noise level will be evaluated.

References

Aitchison, J. (1999). Logratios and natural laws in compositional data analysis. *Mathematical Geology 31*(5), 563–580.

Carroll, J. D. and J.-J. Chang (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of eckart-young decomposition. *Psychometrika 35*(3), 283–319.

Engle, M. A., M. Gallo, K. T. Schroeder, N. J. Geboy, and J. W. Zupancic (2014). Three-

way compositional analysis of water quality monitoring data. *Environmental and Ecological Statistics 21*(3), 565–581.

Faber, N. K. M., R. Bro, and P. K. Hopke (2003). Recent developments in CANDE-COMP/PARAFAC algorithms: a critical review. *Chemometrics and Intelligent Laboratory Systems 65*(1), 119–137.

Gallo, M. (2013). Log-ratio and parallel factor analysis: an approach to analyze three-way compositional data. In *Advanced dynamic modeling of economic and social systems*, pp. 209–221. Springer.

Gallo, M. and V. Simonacci (2013). A procedure for the three-mode analysis of compositions. *Electronic Journal of Applied Statistical Analysis 6*(2), 202–210.

Harshman, R. A. (1970). Foundations of the parafac procedure: Models and conditions for an" explanatory" multi-modal factor analysis.

Pawlowsky-Glahn, V., J. J. Egozcue, and D. Lovell (2015). Tools for compositional data with a total. *Statistical Modelling 15*(2), 175–190.

Pawlowsky-Glahn, V., J. J. Egozcue, and R. Tolosana-Delgado (2015). *Modeling and analysis of compositional data*. John Wiley & Sons.

Tomasi, G. and R. Bro (2006). A comparison of algorithms for fitting the parafac model. *Computational Statistics & Data Analysis 50*(7), 1700–1734.