Cost and performance: a composite indicator for separated waste collection in Italy

Carlo Cavicchia[1]; Pasquale Sarnacchiaro[2]; Maurizio Vichi[3]

[1] Department of Econometrics, Erasmus University Rotterdam, Rotterdam, Netherlands
[2] Department of Economics Management and Institution, University of Naples Federico II, Naples, Italy
[3] Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

Abstract:

In the last years, the production of information and statistics about waste management and separated waste collection has consistently increased. This paper builds a composite indicator for the separated waste collection in Italy taking into consideration both the performances and the costs via a hierarchical latent variable model.

In detail, we propose a composite indicator which complies good properties and detects the main dimensions of the phenomenon. Each dimension is measured as a specific composite indicator which reflects a subset of variables. This paper therefore provides a hierarchically aggregated model-based index that best describes the separated waste collection in Italy with its main features by identifying the most important second order (i.e., hierarchical) relationships among the subsets of manifest variables. All the parameters are estimated according to the maximum likelihood estimation method in order to make inference on the parameters and on the validity of the model.

Keywords:
Latent variable model; Hierarchical model; Model-based; Latent concept; Statistical estimation

1. Introduction:

According to the OECD Glossary of Statistical terms, a composite indicator is formed when individual variables (i.e., manifest indicators) are compiled into a single non-observable index, based on an underlying model for the multidimensional concept that is being measured. A composite indicator is therefore a mathematical combination, or weighted aggregation, of variables that generally have different units of measurement and can be differently combined (Nardo et al., 2005).

Composite indicators are non-observable latent variables which can summarize a big amount of information; for this specific feature they are very useful to measure multidimensional phenomena. On the other hand, composite indicators are frequently criticized because the methods for their construction are not always statistical and mathematically rigorous and they are often based on theories which do not seem to have a solid foundation (Mazziotta and Pareto, 2013). In detail, many researchers do not appreciate composite indicators determined by subjective weights on the variables because this approach can lead to the misinterpretation of the results (Nardo et al., 2005).

This research is focused on studying the separated waste collection in Italy taking into consideration both its performances and its costs via a model-based approach. This topic is increasingly important since, as the Environment European Union Commissioner J. Potočnik stated, many States are still land-filling huge amounts of municipal waste – the worst waste management option – despite the existence of better alternatives, and notwithstanding structural funds being available to finance better options. Valuable resources are being buried, potential economic benefits are being lost, jobs in the waste management sector are not being created, and human health and the environment suffer. Furthermore, the quantity of information and statistics about waste management are more and more consistent, yet only a few studies are actually available in this field. For instance, Cavicchia, Sarnacchiaro and Vichi (2021) detected which dimensions have an impact on the Waste Management in EU building a general composite indicator based on three specific composite indicators: recycling and circular economy performances, generation of recyclable waste, and private investments and innovation.

In this paper, we propose to measure the separated waste collection in Italy. In detail, we present a hierarchically composite indicator that best describes separated waste collection in Italy through a second-order factor analysis which highlights the most important dimensions of this multidimensional phenomenon (at the first order) and the general composite indicator (at the second order), by complying with some pivotal properties for a composite indicator. Moreover, this approach allows overcoming the main critiques about the construction of composite indicators, since it produces a model-based composite indicator based on reliable dimensions where all the parameters are statistically estimated. Another important feature of this approach is that it detects disjoint subsets of variables making the interpretation of the dimension easier.

2. Methodology:

In order to evaluate the study, the Second-order Disjoint Factor Analysis (2O-DFA) model was considered, which consists of two nested factor models. Formally, let $\boldsymbol{x}$ be the J-dimensional multivariate random variable with mean vector $\boldsymbol{\mu_x}$ and J-dimensional variance-covariance matrix $\boldsymbol{\Sigma_x}$. The following two simultaneous equations must therefore be considered:

$$\boldsymbol{x} = \boldsymbol{Ay} + \boldsymbol{e_x} \tag{1}$$
$$\boldsymbol{y} = \boldsymbol{c}g + \boldsymbol{e_y} \tag{2}$$

where $\boldsymbol{A}$ is the $(J \times H)$ matrix of unknown specific composite indicators loadings (H is the number of specific composite indicators included into the model), $\boldsymbol{y}$ is the non-observable $(H \times 1)$ vector of unknown specific composite indicators scores and $\boldsymbol{e_x}$ is a $(J \times 1)$ random vector of errors for the model (1). g is the realization of the general composite indicator $\boldsymbol{g}$ which is normally distributed with mean 0 and variance 1, $\boldsymbol{c}$ is the $(H \times 1)$ vector of unknown general composite indicator loadings and $\boldsymbol{e_y}$ is a $(H \times 1)$ random vector of errors for the model (2). The complete model might be written including equation (2) into equation (1) and considering the loading matrix $\boldsymbol{A}$ equal to the product $\boldsymbol{BV}$, where $\boldsymbol{B}$ is a diagonal matrix and $\boldsymbol{V}$ a row stochastic and binary matrix. 2O-DFA for n multivariate observations is therefore defined as follows

$$\boldsymbol{x} = \boldsymbol{BV}(\boldsymbol{c}g + \boldsymbol{e_y}) + \boldsymbol{e_x} \tag{3}$$

Under the assumption of normality for $\boldsymbol{y}$, $\boldsymbol{e_x}$ and $\boldsymbol{e_y}$, it can be easily derived that $\boldsymbol{x} \sim N_J(\boldsymbol{\mu_x}, \boldsymbol{\Sigma_x})$, with

$$\boldsymbol{\Sigma_x} = \boldsymbol{BV\Sigma_y V'B} + \boldsymbol{\Psi_x} \tag{4}$$

where $\boldsymbol{\Sigma_y} = \boldsymbol{cc'} + \boldsymbol{\Psi_y}$ represents the correlation matrix of the specific composite indicators, $\boldsymbol{\Psi_x}$ is the J-dimension diagonal positive definite variance-covariance matrix of the error of model (1) and $\boldsymbol{\Psi_y}$ is the H-dimension diagonal positive definite variance-covariance matrix of the error of model (2).
2O-DFA aims at reconstructing $\boldsymbol{\Sigma_x}$ in terms of $2J + H$ unknown free parameters in $\boldsymbol{B}$, $\boldsymbol{V}$, $\boldsymbol{\Psi_x}$, $\boldsymbol{c}$ and $\boldsymbol{\Psi_y}$. The discrepancy function to be minimized with respect to $\boldsymbol{B}$, $\boldsymbol{V}$, $\boldsymbol{\Psi_x}$, $\boldsymbol{c}$ and $\boldsymbol{\Psi_y}$ is

$$D(\boldsymbol{B}, \boldsymbol{V}, \boldsymbol{\Psi}_x, \boldsymbol{c}, \boldsymbol{\Psi}_y) = \log|\boldsymbol{\Sigma}_x| + \mathrm{tr}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{S}), \tag{5}$$

where $\boldsymbol{S}$ is the J-dimensional sample variance-covariance matrix. The minimization of the discrepancy function is a discrete and continuous problem that cannot be solved by a quasi-Newton type algorithm, reason why we developed a descent coordinates algorithm.

In order to assess the variable selection and the goodness of the estimations, we presented the standard errors for the estimation of specific composite indicators loadings for the model proposed in this application. Specifically, the standard errors (Std Err) were calculated according to the formulas presented by Lawley and Maxwell (1971, pp. 56-57) for the one-factor case. In order to test the significance of specific composite indicator loadings, we considered the Bonferroni correction which controls the family-wise Type I error. It is worth noticing that notwithstanding the stricter criterion of significant loadings under the Bonferroni correction, non-significant specific composite indicator loadings do not necessarily mean that the latter are zero in the population; this correction therefore does not allow detecting zero composite indicator loadings (Zhang, 2014). Furthermore, the Bonferroni critical point for simultaneous two-tailed tests is $\mathrm{Z}_{0.025/\mathrm{n_h}}$, which means that the critical point for the specific composite indicators loading corresponds to $\pm Z_{0.025/n_h}$ × Std Err, where Z_α represents the α level's z-score and n_h is the number of variables related to the specific composite indicator h. It is worth noticing that if the positive loading is not statistically significant, this means that a larger number of dimensions is required.
For assessing the reliability of the dimensions, we consider the widely used index Cronbach's alpha (Cronbach, 1951). A credited rule of thumb for describing reliability was given by George and Mallery (2003) as follows: if the index is larger than 0.9 the level of reliability is excellent, when it lies within $[0.8,0.9]$ the level is good, when the index is within $[0.7,0.8[$ we can consider the level of reliability acceptable, and, finally, indices under 0.7 are unacceptable. For assessing unidimensionality, we used the second largest eigenvalue of the variance-covariance sub-matrix related to the subset of variables, which must be smaller than 1 (Cavicchia and Vichi, 2020).

3. Result:

The separated waste collection fits within the dimension called "generation of recyclable waste" by Cavicchia, Sarnacchiaro and Vichi (2021), which in turn contributes to define the more general concept of waste management. A good separated waste collection consists of two major aspects: the performances and the costs. The main goal of this analysis is the construction of a composite indicator which can consider these two aspects at the same time. It is worth underlining that 2O-DFA is a model which statistically detects the disjoint subsets of variables defining the dimensions. The analysis is therefore exploratory and not confirmatory, this means that all parameters are statistically estimated, including the membership matrix $\boldsymbol{V}$.

The data used in this application are from different sources: Eurostat, Joint Research Centre (JRC) and Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA). These variables are regularly updated and free. In detail, after several steps of pre-processing, we included in our analysis eight variables (Table 1) for the 40 largest Italian municipalities. Many variables about the characteristic of countries were considered in order to help the interpretation of the results, specifically the size population (i.e., number of inhabitants) was considered to normalize some of the variables included in the study. Two variables represent the costs of the separated waste while the other six variables express the performances of it.

The motivation of this study lies on the assumption that it is crucial to combine the information from the costs and the performances to provide a support for Italian municipalities' actions and policies. The information from these two aspects, if measured separately, might be either misleading or limited.

Table 1: List of variables

	Variable	*Code*
1	Cost of separated waste collection and transport (€ per capita)	CCT
2	Cost of separated waste treatment and recycle (€ per capita)	CTC
3	Organic waste collection (kg per capita)	OWC
4	Paper waste collection (kg per capita)	PaW
5	Glass waste collection (kg per capita)	GlW
6	Metal waste collection (kg per capita)	MeW
7	Plastic waste collection (kg per capita)	PlW
8	Percentage of separated waste over the total waste (%)	PSW

2O-DFA was therefore applied to our dataset – a few missing data were MCAR (Missing Completely at Random), and they were imputed by the K-nearest neighbors method by setting K$= 4$ and by using the Euclidean distance, the variables were then standardized – and the best model in terms of Bayesian Information Criterion (BIC, Schwarz, 1978) was found for $H = 2$. All variables resulted statistically significant, and the model outlines the two expected subsets of variables: costs (CCT and CTC) and performances (OWC, PaW, GlW, MeW, PlW and PSW). The latter resulted to be reliable (Cronbach's alpha was equal to 0.70 and 0.91, respectively) and unidimensional (the second largest eigenvalue of the variance-covariance sub-matrix related to the subset of variables resulted equal to 0.46 and 0.66, respectively).

The most interesting results obtained are that the 2 specific composite indicators, COS (costs) and PER (performances), are positively correlated (0.33) and that the general composite indicator (SWC) has a stronger relationship with PER than with COS. The first result is important because it shows that good performances are obtained also thanks to higher costs, and cities like Ferrara, which outperforms all the others in terms of PER, is one of the cities with the highest COS. The second result shows that, for instance, Ferrara is also first in the ranking given by the SWC.

At the same time, a composite indicator including these two aspects was needed because Spearman's correlation (Spearman, 1904) between the two specific composite indicators was only equal to 0.30, and PER and COS therefore explained different dimensions of the same general latent concept (the separated waste collection). In detail, it is interesting observing that the effect of the costs in the definition of the general composite indicator is significant (i.e., Spearman's correlation equal to 0.60). Although PER results crucial for measuring SWC and their Spearman's correlation is equal to 0.92, SWC highlights that the behavior of the largest Italian municipalities is quite different according to the two aspects detected by 2O-DFA and the costs also affect the general quality of the separated waste collection. Figure 1 displays the relationships between COS, PER and SWC.

PSW is the most important variable - i.e., the one which contributes more in terms of weight - in the definition of PER, and it reflects the performances expressed by the other variables within the same subset. In order to assess the sensibility of our study, we performed the same analysis after discarding PSW from the model and we obtained the same results. Our model therefore is robust, and its framework is consistent with the latent concept to be measured.

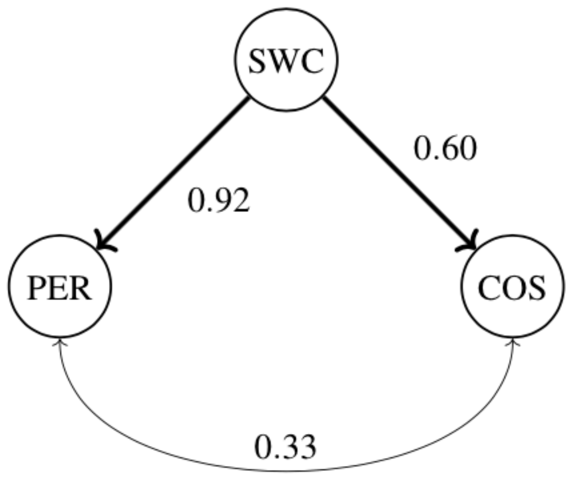

Figure 1: Path diagram of two-order hierarchy

4. Discussion and Conclusion:

In this paper, we proposed a hierarchically aggregated composite indicator for the separated waste collection in Italy, specifically, we considered the 40 largest Italian municipalities. The composite indicator was built through a two-order hierarchy, and it detected two specific composite indicators at the first order and the general one at the second.

The specific composite indicators detected by the model represent the costs related to the separated waste collection and the performances of the municipalities in terms of separated waste generation. Our index results helpful as tool for policy-makers and institutions due to its statistical properties. In detail, it best reconstructs the variables preserving the information contained in the dataset. This model-based approach therefore limits the choices of the researcher which are unfrequently grounded on a verified theory. The most important specific composite indicator in the definition of SWC is PER, while COS contributes less. This model is crucial to compare the municipalities' behaviors in terms of SWC, but also to measure in greater detail and more specifically PER and COS.

In conclusion, this study provides a useful tool to measure the "goodness" of SWC in Italy together with its main aspects, by identifying the most important relationships among variables. The goal is to provide a support for Italian municipalities' actions and policies.

References:

1. Cavicchia, C., Vichi, M. (2020). Statistical Model-based Composite Indicators for tracking coherent policy conclusions. Social Indicators Research. https://doi.org/10.1007/s11205-020-02318-7
2. Cavicchia, C., Sarnacchiaro, P., Vichi, M. (2021). A composite indicator for the waste management in the EU via Hierarchical Disjoint Non-Negative Factor Analysis. Socio-Economic Planning Sciences. https:// 10.1016/j.seps.2020.100832
3. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika. 16 (3): 297–334
4. George, D., Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update. Boston, MA: Allyn and Bacon.
5. Lawley, D.N., Maxwell, A.E. (1971). Factor Analysis as a Statistical Method. London, UK: Butterworths.
6. Mazziotta, M., Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one? Rivista Italiana di Economia Demografia e Statistica 67 (2): 67-80
7. Nardo, M., Saisana, M., Saltelli, A., Tarantola, S. (2015). Tools for Composite Indicators Building. European Commission (Join Research Centre, Ispra, Italy). Report EUR 21682.

8. Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., Giovannini, E. (2005). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing. OECD Statistics Working Papers 2005/3.
9. Schwarz, G. Estimating the dimension of a model. Annals of Statistics 6 (2) (1978) 461-464
10. Spearman C. (1904). "The proof and measurement of association between two things". American Journal of Psychology. 15 (1): 72–101
11. Zhang, G. (2014). Estimating Standard Errors in Exploratory Factor Analysis. Multivariate Behavioral Research. 49 (4): 339-353