PSYCHOMETRIKA—VOL. 87, NO. 1, 289–309
MARCH 2022
https://doi.org/10.1007/s11336-021-09799-6

SECOND-ORDER DISJOINT FACTOR ANALYSIS

CARLO CAVICCHIA
ERASMUS UNIVERSITY ROTTERDAM

MAURIZIO VICHI
UNIVERSITY OF ROME LA SAPIENZA

Hierarchical models are often considered to measure latent concepts defining nested sets of manifest variables. Therefore, by supposing a hierarchical relationship among manifest variables, the general latent concept can be represented by a tree structure where each internal node represents a specific order of abstraction for the latent concept measured. In this paper, we propose a new latent factor model called second-order disjoint factor analysis in order to model an unknown hierarchical structure of the manifest variables with two orders. This is a second-order factor analysis, which—respect to the second-order confirmatory factor analysis—is exploratory, nested and estimated simultaneously by maximum likelihood method. Each subset of manifest variables is modeled to be internally consistent and reliable, that is, manifest variables related to a factor measure "consistently" a unique theoretical construct. This feature implies that manifest variables are positively correlated with the related factor and, therefore, the associated factor loadings are constrained to be nonnegative. A cyclic block coordinate descent algorithm is proposed to maximize the likelihood. We present a simulation study that investigates the ability to get reliable factors. Furthermore, the new model is applied to identify the underlying factors of well-being showing the characteristics of the new methodology. A final discussion completes the paper.

Key words: factor analysis, hierarchical models, latent variable models, reflective models, second-order.

1. Introduction

Factor analysis (FA, Anderson & Rubin, 1956; Horst, 1965) is one of the most used models to reconstruct manifest variables (MVs) through a set of latent variables. However, when the studied latent concepts present a hierarchical structure, FA is not an appropriate method because it is not able to model the hierarchical structure of the concepts; therefore, a model with a hierarchical form is required. In psychometric studies, an epitome is the big five model, which is used to measure the big five personality traits of individuals. It is worth mentioning that several studies have stressed the hierarchical relations among such traits running from the more abstract to more specific (Cattell, 1947; de Raad & Mlačić, 1970; Digman, 2015; Eysenck, 1990). Most typically, hierarchies are studied in two ways: either following a bottom-up approach or a top-down approach (Goldberg, 2006). In this paper, we will be focusing on the bottom-up approach, starting from the MVs up to a general factor.

In multidimensional data analysis, two main classes of models were developed to analyze the hierarchical structures of a multidimensional phenomenon: the higher-order factor models (Cattell, 1948; Thompson, 1978) and the hierarchical factor models (Holzinger & Swineford, 1937). On the one hand, higher-order factor analysis provides a different hierarchical prospective of the data (Thompson, 2004), where factors are supposedly correlated. A frequent procedure to obtain higher-order factors consists of the *sequential* application of exploratory factor analysis

Correspondence should be made to Carlo Cavicchia, Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands. Email: cavicchia@ese.eur.nl

© 2021 The Author(s)

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

(EFA) followed by an oblique rotation method. After performing EFA, an oblique rotation method is applied in order to extract the simple structure (Thurstone, 1947) and the first-order factors. This *sequential* application continues to operate on the correlation matrix of the factors from the first-order upwards, until a single factor or an uncorrelated set of factors are identified (Gorsuch, 1983). Although the oblique rotation after EFA is used to identify the simple structure, Vichi (2017) showed that this sequential approach may fail to find the simple structure, hence underlying that a more specific methodology is necessary.

On the other hand, hierarchical factor models are characterized by a single order of orthogonal hierarchical factors usually obtained by applying the Schmid–Leiman transformation (Schmid & Leiman, 1957) to the corresponding higher-order solutions. Therefore, higher-order models identify the effect of the general factor on MVs only through the higher-order factors, whereas the hierarchical models also identify a direct effect of the general factor on the MVs. However, the link between the higher-order and hierarchical factor models was eventually established by Yung et al. (1999) determining the conditions for their equivalence.

It is worth remarking that many hierarchical extensions of FA were already proposed throughout the years (Le Dien & Pages, 2003; Schmid & Leiman, 1957; Thompson, 1951; Wherry, 1959; 1975; 1984). However, all of these hierarchical extensions were developed as sequential analysis, at times not even guaranteeing to obtain a simple structure, i.e., the partition in H classes of variables where common relations in each class are represented by a single factor.

A hierarchical model which may consider a simple structure of the MVs is the *Hierarchical Confirmatory Factor Analysis* (HCFA, Holzinger, 1944; Jöreskog, 1966; 1969; 1978; 1979). This latter is often used to measure general latent concepts when the number of factors and the most relevant relations between MVs and factors are known. However, such knowledge may also represent a limitation. Firstly, because the researcher might not have the a priori information concerning the relations between the different levels of factors and between variables and factors, or, secondly, because the theory might turn out to be erroneous in some parts or at least in its empirical application.

The researcher can overcome these issues accepting that each variable is related with a latent construct only, without imposing what the relations between MVs and factors are. The result is to obtain an exploratory simple structure model (SSM) where the relations between MVs and factors are determined by the data. It is worth recalling that many researchers have investigated factorial methods to obtain a simple structure. For instance, Hirose and Yamamoto (2014) developed a FA with non-convex sparse penalty which can provide a simple structure. Vichi (2017) proposed a model, named disjoint factor analysis (DFA), to identify the best SSM for the data, wherein the maximum likelihood estimation allows to make inference on the number of factors, on the relations between MVs and factors (i.e., loadings), and to assess the validity of the SSM for the observed data. Adachi and Trendafilov (2018) also proposed a matrix-based procedure for sparse FA such that each variable loads only on one common factor by obtaining a simple structure.

In this paper, we extend DFA, which is not appropriate to identify the hierarchical structure of factors since it assumes that factors are orthogonal, that is, factors are not mutually related and do not share common information that could be summarized by the general factor. Therefore, we release the orthogonal constraint and assume that the correlation structure in the data has an unknown second-order hierarchical form, where the first order indicates a reduced set of multidimensional concepts described by disjoint subsets of MVs, while the second order, denoted root or general level, represents the general factor. Gorsuch (1983) emphasized that the interpretation of higher-order factor models should be based on the MVs, in order to avoid interpretations of interpretations; this can be certainly achieved when disjoint classes of variables can be identified with each class represented by a factor.

Finally, it is crucial to remark that the *sequential* application of the EFA followed by an oblique rotation method cannot efficiently identify the block correlation structure, whereas a simultaneous

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

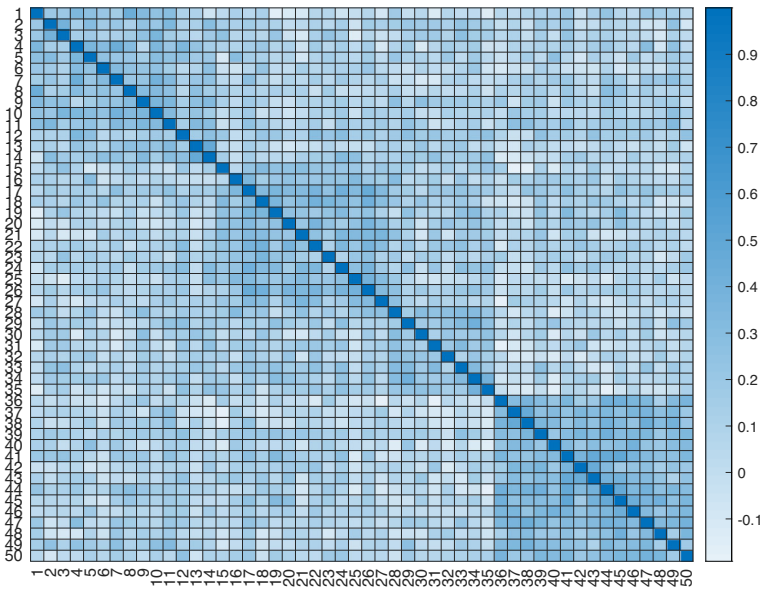


FIGURE 1.
(50 × 50) Correlation matrix with a block diagonal structure in four blocks

and exploratory model can. To verify this, we generated a dataset of 500 observations according to the block diagonal model 16 corresponding to second-order factor analysis. Correlations within blocks are on average around 0.85, while correlations between blocks are around 0.35 (e.g., the heatmap in Fig. 1). Four blocks of variables were considered: $(V_1 - V_{14})$, $(V_{15} - V_{27})$, $(V_{28} - V_{35})$ and $(V_{36} - V_{50})$. Each variable was assigned to the factor with the highest loading resulting from the *sequential* application of the EFA followed by an oblique rotation method. To assess the similarities between two partitions, induced by disjoint blocks of variables, we computed the adjusted Rand index (ARI, Hubert & Arabie, 1985) between the generated blocks and those identified by the methodology. The ARI is the corrected-for-chance version of the Rand index (Rand, 1971). This index has zero expected value in the case the identified partition is a random one, and it is bounded above by 1 in the case of a perfect agreement between the identified and the generated partitions. The ARI between the partition found by EFA followed by an oblique rotation method and the generated one resulted equal to 0.78; our methodology, applied to the same dataset, was able to perfectly detect the generated blocks (i.e., ARI $= 1$). To extend this result to a reasonable number of examples, we generated 200 samples with the same four-block structure. The ARI between the EFA followed by an oblique rotation method and the generated data resulted equal to 1 for 115 times (57, 5%), whereas our methodology perfectly detected the generated structure in 144 cases (72%). So far, our proposed model detected the true structure in sensibly more times than EFA followed by an oblique rotation method, and the averages of ARI resulted equal to 0.89 for our methodology and 0.45 for EFA followed by an oblique rotation method.

The paper is organized as follows. In Sect. 2, we propose the Second-Order Factor Analysis model. Section 3 includes an overview of disjoint models. Section 4 introduces the nonnegative constraints of the factors necessary to specify consistent latent variables. A simulation study is considered in Sect. 5. Section 6 shows an application about well-being. A final discussion completes the paper in Sect. 7.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

2. Second-order factor analysis

Let $\mathbf{x}$ be the $(J \times 1)$ multivariate random variable with mean vector $\boldsymbol{\mu}_{\mathbf{x}} = [\mu_1, \ldots, \mu_J]' = \mathbf{0}_J$ without loss of generality, and J-dimensional variance–covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$. Second-Order Factor Analysis (2O-FA) is modeled as a new factor model that considers two typologies of latent unknown constructs: $(H \leq J)$ first-order factors and a single (nested) general factor identified by the two simultaneous equations

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{w} \tag{1}$$

$$\mathbf{y} = \mathbf{c}g + \mathbf{u} \tag{2}$$

where $\mathbf{A}$ is the $(J \times H)$ matrix of unknown first-order factors loadings, $\mathbf{y}$ is the non-observable $(H \times 1)$ random vector denoting the first-order factor scores, and $\mathbf{w}$ is a non-observable $(J \times 1)$ random vector of errors. It is assumed that $\mathbf{y} \sim N_H(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}})$, where $\boldsymbol{\Sigma}_{\mathbf{y}}$ is the correlation matrix of the first-order factors, and $\mathbf{w} \sim N_J(\mathbf{0}, \boldsymbol{\Psi}_{\mathbf{x}})$, where $\text{cov}(\mathbf{w}) = \boldsymbol{\Psi}_{\mathbf{x}}$ is the J-dimensional diagonal positive definite variance–covariance matrix of the error of model (1) and $\text{cov}(\mathbf{w}, \mathbf{y}) = \mathbf{0}$.

Furthermore, g is a non-observable random variable normally distributed with mean 0 and variance $\text{cov}(g) = 1$, and $\mathbf{c}$ is the $(H \times 1)$ vector of unknown general factor loadings. In addition, $\mathbf{u}$ is a non-observable $(H \times 1)$ random vector of errors. It is assumed that $\mathbf{u} \sim N_H(\mathbf{0}, \boldsymbol{\Psi}_{\mathbf{y}})$, where $\text{cov}(\mathbf{u}) = \boldsymbol{\Psi}_{\mathbf{y}}$ is the H-dimensional diagonal positive definite variance–covariance matrix of the error of model (2). In addition, it is assumed that errors in the two models are uncorrelated $\text{cov}(\mathbf{w}, \mathbf{u}) = \mathbf{0}$; and errors and factors are uncorrelated, i.e., $\text{cov}(\mathbf{u}, g) = \mathbf{0}$.

Model (1) identifies H specific theoretical constructs by means of a common factor model that identifies common information with H factors related to the MVs, while model (2) detects the general latent construct by means of a one-factor model that identifies common information with one general factor related to the H first-order factors.

Given these assumptions and including (2) into (1), the 2O-FA model for centered data is defined

$$\mathbf{x} = \mathbf{A}(\mathbf{c}g + \mathbf{u}) + \mathbf{w}. \tag{3}$$

It can be derived that $\mathbf{x} \sim N_J(\mathbf{0}_J, \boldsymbol{\Sigma}_{\mathbf{x}})$, where the variance–covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ is

$$\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{x}} &= \text{cov}(\mathbf{A}\mathbf{c}g + \mathbf{A}\mathbf{u} + \mathbf{w}) \\
&= \mathbf{A}\mathbf{c}\ \text{cov}(g)\mathbf{c}'\mathbf{A}' + \mathbf{A}\ \text{cov}(\mathbf{u})\mathbf{A}' + \text{cov}(\mathbf{w}) \\
&= \mathbf{A}\mathbf{c}\mathbf{c}'\mathbf{A}' + \mathbf{A}\boldsymbol{\Psi}_{\mathbf{y}}\mathbf{A}' + \boldsymbol{\Psi}_{\mathbf{x}} \\
&= \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{y}}\mathbf{A}' + \boldsymbol{\Psi}_{\mathbf{x}}
\end{aligned} \tag{4}$$

with

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{c}\mathbf{c}' + \boldsymbol{\Psi}_{\mathbf{y}}. \tag{5}$$

Matrix $\boldsymbol{\Sigma}_{\mathbf{y}}$ is a correlation matrix since first-order factors are standardized. Similarly to exploratory factor analysis, 2O-FA (3) does not imply any a priori knowledge of relations between MVs and factors. Although many hierarchical extensions of FA are present into specialized literature (e.g., Le Dien & Pages, 2003; Schmid & Leiman, 1957; Thompson 1951; Wherry, 1959;

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

1975; 1984), the novelty of the proposal consists of the covariance structure given by the model (4) and (5) and the simultaneous estimation of the parameters.

Let us consider a random sample of $n > J$ multivariate observations $\mathbf{x}_i = [x_{i1}, \ldots, x_{iJ}]'$, $i = 1, \ldots, n$ drawn of $\mathbf{x}$, with mean vector $\bar{\mathbf{x}}$, and J-dimensional variance–covariance matrix $\mathbf{S}_\mathbf{x} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'$, the model (3) in matrix form corresponds to

$$\mathbf{X} = \mathbf{g}\mathbf{c}'\mathbf{A}' + \mathbf{E} \tag{6}$$

where $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]'$ is the $(n \times J)$ matrix containing the n multivariate observations, $\mathbf{g} = [g_1, \ldots, g_n]'$ is the non-observable $(n \times 1)$ vector denoting the second-order (general) factor scores and $\mathbf{E} = \mathbf{U}\mathbf{A}' + \mathbf{W}$. In detail, $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n]'$ with dimensions $(n \times J)$ and $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]'$ with dimensions $(n \times H)$ are matrices containing the non-observable errors related to (1) and (2), respectively. The reduced log-likelihood (i.e., conditional on $\boldsymbol{\mu}_\mathbf{x}$ equal to the sample mean) is

$$\begin{aligned} L(\mathbf{A}, \mathbf{c}, \boldsymbol{\Psi}_\mathbf{x}, \boldsymbol{\Psi}_\mathbf{y}) = -\frac{nJ}{2}\log(2\pi) - \frac{n}{2}\{&\log|\mathbf{A}(\mathbf{c}\mathbf{c}' + \boldsymbol{\Psi}_\mathbf{y})\mathbf{A}' + \boldsymbol{\Psi}_\mathbf{x}| \\ &+ \mathrm{Tr}[(\mathbf{A}(\mathbf{c}\mathbf{c}' + \boldsymbol{\Psi}_\mathbf{y})\mathbf{A}' + \boldsymbol{\Psi}_\mathbf{x})^{-1}\mathbf{S}_\mathbf{x}]\}. \end{aligned} \tag{7}$$

3. Disjoint Models

Disjoint orthogonal factor analysis (DFA, Vichi, 2017) assumes that observations can be reconstructed by a non-observable $(H \times 1)$ random vector $\mathbf{y}$ denoting a reduced set of $(H \leq J)$ common factors. DFA for centered data can be expressed via the following model

$$\mathbf{x} = \mathbf{B}\mathbf{V}\mathbf{y} + \mathbf{w} \tag{8}$$

with a covariance structure equal to (4), considering the loading matrix $\mathbf{A}$ restricted to the product

$$\mathbf{A} = \mathbf{B}\mathbf{V} \tag{9}$$

where $\mathbf{V} = [v_{jh}]$ is a $(J \times H)$ binary and row stochastic matrix identifying a partition of MVs into H subsets corresponding to H factors; the H subsets of MVs are denoted as C_h, with $h = 1, \ldots, H$. If the jth MV belongs to the hth subset then $v_{jh} = 1$, otherwise, $v_{jh} = 0$; whereas, $\mathbf{B} = \mathrm{diag}(b_1, \ldots, b_J)$ is a $(J \times J)$ weighting diagonal matrix such that $b_j^2 > 0$. If it is allowed $b_j^2 \geq 0$, when $b_j^2 = 0$ the DFA admits a model selection feature, that is, a MV is assigned to a subset with a loading equal to zero. In this case, the MV j is discarded from the model. Despite the fact that DFA assumes orthogonal factors, that is, $\boldsymbol{\Sigma}_\mathbf{y} = \mathbf{I}_H$, in this paper this condition is relaxed in order to allow a hierarchical structure of the data. Note that $\mathrm{diag}(\cdot)$ produces a diagonal matrix of a vector.

Therefore, formally, DFA corresponds to (1) with covariance structure given by (4) once the loading matrix $\mathbf{A}$ is defined according to (9). The model is defined under the following constraints

$$\mathbf{V} = [v_{jh} \in \{0, 1\} : j = 1, \ldots, J, h = 1, \ldots, H] \tag{10}$$

$$\mathbf{V}\mathbf{1}_H = \mathbf{1}_J \quad \text{i.e.} \quad \sum_{h=1}^{H} v_{jh} = 1 \quad j = 1, \ldots, J \tag{11}$$

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

$$\mathbf{B} = \text{diag}(b_1, \ldots, b_J) \tag{12}$$

$$\mathbf{V}'\mathbf{B}\mathbf{B}\mathbf{V} = \text{diag}(b_{\cdot 1}^2, \ldots, b_{\cdot H}^2) \quad \text{with} \quad b_{\cdot h}^2 = \sum_{j \in C_h} b_j^2. \tag{13}$$

In detail, if $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{I}_H$ (disjoint orthogonal factor analysis) the variance–covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ is block diagonal:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \text{blkdiag}(\boldsymbol{\Sigma}_{11}, \ldots, \boldsymbol{\Sigma}_{hh}, \ldots, \boldsymbol{\Sigma}_{HH}) = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_{22} & 0 & \ldots & \ldots & 0 \\ \vdots & 0 & \ddots & 0 & \ldots & 0 \\ \vdots & \vdots & 0 & \boldsymbol{\Sigma}_{hh} & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{\Sigma}_{HH} \end{bmatrix}. \tag{14}$$

Each block is the variance–covariance matrix $\boldsymbol{\Sigma}_{hh}$ of the MVs relating to the factor h,

$$\boldsymbol{\Sigma}_{hh} = \mathbf{B}_h(\mathbf{1}_{n_h}\mathbf{1}'_{n_h})\mathbf{B}_h + \boldsymbol{\Psi}_h = \mathbf{b}_h\mathbf{b}'_h + \boldsymbol{\Psi}_h \tag{15}$$

where $\mathbf{B}_h = \text{diag}(\mathbf{b}_h)$, $\mathbf{b}_h = [b_{1h}, \ldots, b_{n_h h}]'$ and $\boldsymbol{\Psi}_h = \text{diag}(\boldsymbol{\Psi}_h)$, $\boldsymbol{\Psi}_h = [\boldsymbol{\Psi}_{1h}, \ldots, \boldsymbol{\Psi}_{n_h h}]'$. Note that n_h is the number of MVs related to the latent factor h. Therefore, DFA assumes that a relevant correlation among MVs related to the same latent factor is observed, while a negligible correlation between MVs related to different latent factors is detected.

If $\boldsymbol{\Sigma}_{\mathbf{y}}$ has non-diagonal elements different from zero (disjoint non-orthogonal factor analysis), the block diagonal variance–covariance disappears and $\boldsymbol{\Sigma}_{\mathbf{x}}$ has the form

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \ldots & \boldsymbol{\Sigma}_{1H} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \ldots & \boldsymbol{\Sigma}_{2H} \\ \vdots & \vdots & \boldsymbol{\Sigma}_{hh} & \vdots \\ \boldsymbol{\Sigma}_{H1} & \boldsymbol{\Sigma}_{H2} & \ldots & \boldsymbol{\Sigma}_{HH} \end{bmatrix} \tag{16}$$

where the generic matrix correlation between two subsets of MVs is constrained in

$$\boldsymbol{\Sigma}_{hk} = c_h c_k \mathbf{B}_h(\mathbf{1}_{n_h}\mathbf{1}'_{n_k})\mathbf{B}_k = c_h c_k \mathbf{b}_h\mathbf{b}'_k \tag{17}$$

and $c_h c_k$ expresses the correlation between factor h and factor k.

It is worth noticing that in order to identify two distinct factors h and k with two associated disjoint subsets of MVs defining matrices $\boldsymbol{\Sigma}_{hh}$ and $\boldsymbol{\Sigma}_{kk}$, both a high correlation within these matrices and a lower (or equal to, if and only if c_h and c_k are both equal to 1) correlation within $\boldsymbol{\Sigma}_{hk}$ need to be observed. In fact, if a high correlation is also observed in $\boldsymbol{\Sigma}_{hk}$, then a single factor in the data is present, since the two subsets of MVs are not distinct and actually form a single subset in which MVs are all highly correlated. Thus, (17) guarantees that correlations in $\boldsymbol{\Sigma}_{hk}$ are lower than correlations within $\boldsymbol{\Sigma}_{hh}$ and $\boldsymbol{\Sigma}_{kk}$.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

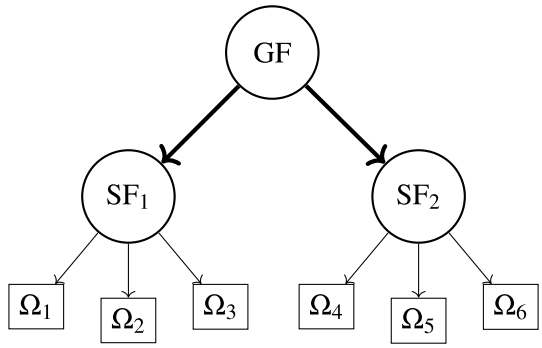


FIGURE 2.
Example of second-order disjoint factor model

3.1. *Second-Order Disjoint Factor Analysis*

When in the 2O-FA some a priori substantial knowledge is incorporated in the form of restrictions on the loading matrix, this usually improves the description of the latent factors and leads to a parsimonious model with a simple loading matrix structure. Therefore, if in 2O-FA a SSM is assumed observed for the data, this means that the factor loading matrix has the form $\mathbf{A} = \mathbf{BV}$ (Eq. 9) and 2O-FA becomes *Second-Order Disjoint Factor Analysis* (2O-DFA, Fig. 2).

Once the loading matrix $\mathbf{A}$ is defined according to (9), 2O-DFA is defined by (3), or alternatively in matrix form by (6). It is worth observing that the maximization of (7) corresponds to the minimization of

$$D(\mathbf{B}, \mathbf{V}, \mathbf{c}, \boldsymbol{\Psi}_{\mathbf{x}}, \boldsymbol{\Psi}_{\mathbf{y}}) = \log|\mathbf{BV}(\mathbf{cc}' + \boldsymbol{\Psi}_{\mathbf{y}})\mathbf{V}'\mathbf{B} + \boldsymbol{\Psi}_{\mathbf{x}}| + \mathrm{Tr}\{[\mathbf{BV}(\mathbf{cc}' + \boldsymbol{\Psi}_{\mathbf{y}})\mathbf{V}'\mathbf{B} + \boldsymbol{\Psi}_{\mathbf{x}}]^{-1}\mathbf{S}_{\mathbf{x}}\}. \tag{18}$$

3.2. *Second-Order Disjoint FA algorithm*

Given H, a cyclic block coordinate descent algorithm for the estimation of the model can be described by five steps which are sequentially repeated until a stopping rule is satisfied.

Step 0 [*Initialization*] A random partition $\widehat{\mathbf{V}}$ is generated from a multinomial distribution in H categories each with equal probability, where categories are not empty. Matrices $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}} = \mathrm{diag}(\mathbf{S}_{\mathbf{x}})$, $\widehat{\boldsymbol{\Psi}}_{\mathbf{y}} = \mathrm{diag}(\widehat{\psi}_{\mathbf{y}1}, \ldots, \widehat{\psi}_{\mathbf{y}H})$, where each value $\widehat{\psi}_{\mathbf{y}h}$, $h = 1, \ldots, H$ is generated from an uniform distribution $U(0, 1)$, and vector $\widehat{\mathbf{c}} = [\widehat{c}_1, \ldots, \widehat{c}_H]'$, where each value $\widehat{c}_h$, $h = 1, \ldots, H$ is generated from an uniform distribution $U(0, 1)$.

Step 1 Given $\widehat{\mathbf{V}} = [\widehat{\mathbf{v}}_{\cdot 1}, \ldots, \widehat{\mathbf{v}}_{\cdot H}]$, $\widehat{\mathbf{c}}$, $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}}$ and $\widehat{\boldsymbol{\Psi}}_{\mathbf{y}}$, the discrepancy function $D(\mathbf{B}, \widehat{\mathbf{V}}, \widehat{\mathbf{c}}, \widehat{\boldsymbol{\Psi}}_{\mathbf{x}}, \widehat{\boldsymbol{\Psi}}_{\mathbf{y}})$ (18) is minimized with respect to $\mathbf{B}_h = \mathrm{diag}(\mathbf{b}_h)$ by

$$\widehat{\mathbf{b}}_h = \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{\frac{1}{2}} \mathbf{m}_{1h} (\lambda_{1h} - 1)^{\frac{1}{2}} \quad h = 1, \ldots, H \tag{19}$$

where $\mathbf{S}_{\mathbf{x}h}$ and $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}$ are the variance–covariance matrix and the current estimation of the diagonal error variance–covariance matrix of MVs identified by $\widehat{\mathbf{v}}_{\cdot h}$, respectively. λ_{1h} is the largest eigenvalue of $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} \mathbf{S}_{\mathbf{x}h} \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}}$ and $\mathbf{m}_{1h}$ is the associated eigenvector.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

Step 2 Given $\widehat{\mathbf{A}} = \widehat{\mathbf{B}}\widehat{\mathbf{V}}, \widehat{\mathbf{c}}$ and $\widehat{\boldsymbol{\Psi}}_{\mathbf{y}}$ the minimization of the discrepancy function (18) with respect to $\boldsymbol{\Psi}_{\mathbf{x}}$ is given by

$$\widehat{\boldsymbol{\Psi}}_{\mathbf{x}} = \operatorname{diag}(\mathbf{S}_{\mathbf{x}} - \widehat{\mathbf{A}}(\widehat{\mathbf{c}}\widehat{\mathbf{c}}' + \widehat{\boldsymbol{\Psi}}_{\mathbf{y}})\widehat{\mathbf{A}}'). \tag{20}$$

Step 3 Given $\widehat{\mathbf{A}} = \widehat{\mathbf{B}}\widehat{\mathbf{V}}$, $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}}$ and $\widehat{\boldsymbol{\Psi}}_{\mathbf{y}}$ the minimization of the discrepancy function (18) with respect to $\mathbf{c}$ is given by

$$\widehat{\mathbf{c}} = \widehat{\boldsymbol{\Psi}}_{\mathbf{y}}^{\frac{1}{2}} \mathbf{p}_1 (q_1 - 1)^{\frac{1}{2}} \tag{21}$$

where q_1 is the largest eigenvalue and $\mathbf{p}_1$ is the associated eigenvector of $\widehat{\boldsymbol{\Psi}}_{\mathbf{y}}^{-\frac{1}{2}}(\widehat{\mathbf{A}}^{+}(\mathbf{S}_{\mathbf{x}} - \widehat{\boldsymbol{\Psi}}_{\mathbf{x}})\widehat{\mathbf{A}}'^{+})\widehat{\boldsymbol{\Psi}}_{\mathbf{y}}^{-\frac{1}{2}}$, and $\mathbf{A}^{+}$ denotes the Moore–Penrose inverse of a matrix $\mathbf{A}$ (i.e., $\mathbf{A}^{+} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$).

Step 4 Given $\widehat{\mathbf{A}} = \widehat{\mathbf{B}}\widehat{\mathbf{V}}, \widehat{\mathbf{c}}$ and $\widehat{\boldsymbol{\Psi}}_{\mathbf{x}}$ the minimization of the discrepancy function (18) with respect to $\boldsymbol{\Psi}_{\mathbf{y}}$ is given by

$$\widehat{\boldsymbol{\Psi}}_{\mathbf{y}} = \operatorname{diag}(\mathbf{I}_H - \widehat{\mathbf{c}}\widehat{\mathbf{c}}') \tag{22}$$

where $\mathbf{I}_H$ is the identity matrix of order H.

Step 5 Partition $\widehat{\mathbf{V}} = [\widehat{\mathbf{v}}_{\cdot 1}, \ldots, \widehat{\mathbf{v}}_{\cdot H}]$ is obtained row by row by assigning each jth MV to the hth subset that most decreases the discrepancy function (18). Thus, formally

$$\begin{cases} \widehat{v}_{jh} = 1 \quad \text{if} \quad h = \underset{h'=1,\ldots,H}{\operatorname{argmin}} \; D(\widehat{\mathbf{B}}, [\widehat{\mathbf{v}}_{1\cdot}, \ldots, \widehat{\mathbf{v}}_{j\cdot} = \mathbf{i}_{h'\cdot}, \ldots, \widehat{\mathbf{v}}_{J\cdot}]', \widehat{\mathbf{c}}, \widehat{\boldsymbol{\Psi}}_{\mathbf{x}}, \widehat{\boldsymbol{\Psi}}_{\mathbf{y}}) \\ \widehat{v}_{jh} = 0 \quad \text{otherwise} \end{cases} \tag{23}$$

where $\mathbf{i}_{h\cdot}$ is the hth column of the identity matrix of order H. Note that the update of each row of $\mathbf{V}$ induces the update of the loadings of the two subsets of MVs that are eventually changed. This means that the jth MV is assigned to the hth subset if Eq. (18) is maximized by $\widehat{v}_{jh} = 1$. This step considers a constraint of non-emptiness for each column: thus, the constraints guarantee that each column of $\widehat{\mathbf{V}}$ is not empty and the solution is feasible. In practice, the assignment of a MV is not allowed if this makes a column of $\widehat{\mathbf{V}}$ empty.

In order to better follow the different iterations, the algorithm is presented as follows:

Step 0 $\mathbf{V}^{(0)}$, $\boldsymbol{\Psi}_{\mathbf{x}}^{(0)}$ and $\boldsymbol{\Psi}_{\mathbf{y}}^{(0)}$ randomly generated as described in the initialization step.

For $t = 0, \ldots, T-1$, where T is the maximum number of iterations:

Step 1 $\mathbf{B}^{(t+1)}$ as in the maximum likelihood FA, according to Eq. (19); and $\mathbf{A}^{(t+1)} = \mathbf{B}^{(t+1)}\mathbf{V}^{(t)}$.
Step 2 $\boldsymbol{\Psi}_{\mathbf{x}}^{(t+1)}$ as we obtain the unique variances, according to Eq. (20).
Step 3 $\mathbf{c}^{(t+1)}$ as in the maximum likelihood FA, according to (21).
Step 4 $\boldsymbol{\Psi}_{\mathbf{y}}^{(t+1)}$ as we obtain the unique variances, according to Eq. (22).
Step 5 $\mathbf{V}^{(t+1)}$ according to Eq. (23).

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

Step 1, **2**, **3**, **4** and **5** are therefore alternated repeatedly, and they form a coordinate descent algorithm where, at each iteration (t), the discrepancy function decreases or, at least, it does not increase (Zangwill, 1969). The algorithm continues to iterate until the reduction of the discrepancy is larger than an arbitrary small positive constant and $t < T$ (in our experiments T was set equal to 100), otherwise the algorithm stops and is considered to have converged to a solution that is not guaranteed to be a global minimum. To avoid the well-known sensitivity of the coordinate descent algorithms to the starting values and to increase the chance of finding the global minimum, the algorithm should be run several times starting from different initial estimates of $\mathbf{V}$ and retaining the best solution. The algorithm generally stops after a few iterations (in our experiments and simulation study less than 15). In order to assess the complexity of 2O-DFA, it is worth recalling that the variance–covariance matrix $\mathbf{\Sigma_x}$ has $\frac{J(J+1)}{2}$ elements to be estimated. 2O-DFA reconstructs this matrix in terms of $2J + H$ unknown free parameters in $\mathbf{A} = \mathbf{BV}$, $\mathbf{\Psi_x}$ and $\mathbf{c}$. In detail, we have to consider JH parameters in $\mathbf{A}$, J parameters in $\mathbf{\Psi_x}$ and $\frac{H(H-1)}{2}$ parameters in $\mathbf{\Sigma_y}$, which is in turn estimated in terms of H parameters in $\mathbf{c}$. Yet, in order to obtain a SSM, it is necessary to include $J(H-1)$ constraints in $\mathbf{A}$ (i.e., only one loading per row different from zero). Finally, the effective number of unknown free parameters in 2O-DFA is obtained, and the degrees of freedom are

$$\nu = \frac{J(J+1)}{2} - JH - H - J + J(H-1) = \frac{J(J+1)}{2} - (2J + H). \tag{24}$$

There are many methods to estimate the factor scores that were proposed throughout the years. For instance, we consider the weighted least square estimation of $\mathbb{E}(\mathbf{X}|\mathbf{Y})$

$$\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\mathbf{\Psi}}_{\mathbf{x}}^{-1}\widehat{\mathbf{A}}(\widehat{\mathbf{A}}'\widehat{\mathbf{\Psi}}_{\mathbf{x}}^{-1}\widehat{\mathbf{A}})^{-1} \tag{25}$$

as proposed by Bartlett (1937), or simply the Thompson (1934) regression estimator

$$\widehat{\mathbf{Y}} = \mathbf{X}(\widehat{\mathbf{\Sigma}}_{\mathbf{x}}^{-1}\widehat{\mathbf{A}}). \tag{26}$$

4. Second-Order Nonnegative Disjoint Factor Analysis

2O-DFA can be considered to build statistically estimated (i.e., nonnormative) composite indicators (CIs, OECD, 2008) for multidimensional phenomena, when the first-order factors represent specific CIs (one per subset of MVs) and the second-order factor is the general CI. This approach guarantees to comply with certain good properties on which a CI should be based (e.g., model-based with a two-order structure, scale-invariance, uni-dimensionality and reliability). However, the presence of negative loadings can be a limitation since positive relationships among MVs might be compensated by negative ones. First-order loadings show the importance of each MV into the definition of the related first-order factor, while second-order loadings represent the importance of the first-order factors into the definition of the second-order factor; thus, the researcher may constrain the loadings to be positive and consistent. If some or all MVs increase, then also the corresponding first-order factor consistently increases. The same holds for the relationship between the first-order factors and the second-order factor. In order to ensure that loadings remain a measure of importance, the aggregation method should not allow compensability (Greco et al., 2019; Munda & Nardo, 2009; OECD, 2008). We therefore propose a constrained version of 2O-DFA called Second-Order Nonnegative Disjoint Factor Analysis (2ON-DFA).

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

Let us recall that the discrepancy function (18) is minimized with respect to $\mathbf{B}_h = \text{diag}(\mathbf{b}_h)$ by (19), and that λ_{1h} and $\mathbf{m}_{1h}$ minimize $||\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}}\mathbf{S}_{\mathbf{x}}\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} - \lambda_{1h}\mathbf{m}_{1h}\mathbf{m}_{1h}'||^2$, or equivalently

$$||\mathbf{X}_h\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}} - \sqrt{\lambda_{1h}}\mathbf{y}_h\mathbf{m}_{1h}'||^2 \tag{27}$$

where $\mathbf{X}_h$ is the data matrix formed by MVs identified by $\mathbf{v}_{\cdot h}$, and $\mathbf{y}_h$ is the factor score vector. The problem in (27) can be solved by an alternating least-squares algorithm that alternates the solution of two regression problems. Given $\widehat{\mathbf{m}}_{1h}$, compute $\widehat{\mathbf{y}}_h$ by

$$\widehat{\mathbf{y}}_h = \mathbf{X}_h\widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}}\widehat{\mathbf{m}}_{1h}(\widehat{\mathbf{m}}_{1h}'\widehat{\mathbf{m}}_{1h})^{-1}. \tag{28}$$

Given $\widehat{\mathbf{y}}_h$, compute $\widehat{\mathbf{m}}_{1h}$ by

$$\widehat{\mathbf{m}}_{1h} = \widehat{\boldsymbol{\Psi}}_{\mathbf{x}h}^{-\frac{1}{2}}\mathbf{x}_h'\widehat{\mathbf{y}}_h(\widehat{\mathbf{y}}_h'\widehat{\mathbf{y}}_h)^{-1}. \tag{29}$$

At each iteration of **Step 1**, **2**, **3**, **4** and **5**, the discrepancy function (18) decreases or at least does not increase. The algorithm stops when the discrepancy function (18) decreases less than a positive arbitrary constant. It is now required that the vector $\mathbf{m}_{1h}$ gets filled by nonnegative elements, and, consequently, the algorithm based on (18), (19), (20), (21), (22) and (23) has to be modified to include nonnegative constraints on $\mathbf{m}_{1h}$. The solution can be found by the nonnegative least-squares algorithm (Lawson & Hanson, 1974, pp. 168–169), which is an active set algorithm. The H inequality constraints are active if the regression coefficients $\mathbf{m}_{1h}$ in the loss function (27) are negative (or zero) when unconstrained estimated, otherwise constraints are passive. The nonnegative solution of (18) with respect to $\mathbf{m}_{1h}$ will simply be the unconstrained least-squares solution using only the MVs corresponding to the passive set, imposing the regression coefficients of the active set to zero. A similar step is used to constrain $\widehat{\mathbf{c}}$ to be nonnegative.

5. Simulation study

In this section, a simulation study was implemented in order to assess the classification of MVs and evaluate the reliability of the first-order factors. Moreover, 2ON-DFA was compared with respect to EFA followed by rotations. Each simulated random sample of $n > J$ multivariate observations $\mathbf{x}_i$ $(i = 1, \ldots, n)$ is generated according to $\mathbf{X} = \mathbf{X}_t + \mathbf{E}$, where $\mathbf{X}_t \sim N_J(\mathbf{0}, \mathbf{R}_{\mathbf{x}})$ and $\mathbf{E} \sim N_J(\mathbf{0}, a\mathbf{I}_J)$, where a allows to set an error. The correlation structure $\mathbf{R}_{\mathbf{x}}$ is modeled as follows:

$$\mathbf{R}_{\mathbf{x}} = \beta(\mathbf{V}\mathbf{V}' - \mathbf{I}_J) + \zeta(\mathbf{1}_J\mathbf{1}_J' - \mathbf{V}\mathbf{V}') + \mathbf{I}_J \tag{30}$$

where $\beta = 0.8 + 0.05\delta$, $\zeta = 0.35 + 0.05\delta$, with $\delta \sim N(0, 1)$ and $\beta > \zeta$. In detail, β represents the correlation among MVs related to the same first-order factor (i.e., the elements in Eq. 15) and ζ represents the correlation among MVs relate to different first-order factors (i.e., the elements in Eq. 17).

In particular, the following scenarios were considered in the simulation study:

- number of MVs, $J = 20, 50$;
- number of the first-order factors, $H = 5, 10$;

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

- different error levels, $a = 0.33, 0.66, 0.85, 1, 1.33$.

500 datasets containing 200 observations were generated for each setting and the data were standardized. In addition, a different random membership matrix $\mathbf{V}$, representing a random partition of J MVs in H nonempty subsets, was generated for each dataset. In detail, the matrix $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_J]'$ was generated according to $\mathbf{v}_j \sim \text{Multinomial}(H : p_h = \frac{1}{H},\ h = 1, \ldots, H),\ \forall j = 1, \ldots, H$. The algorithm was run from 30 random starting points to increment the possibility of obtaining the optimal solution, although it was observed that a few random starting points were generally enough. The models' performance was analyzed using ARI (Hubert & Arabie, 1985), comparing the simple structure of the MVs generated by the true matrix $\mathbf{V}$ with the $\widehat{\mathbf{V}}$ estimated by the 2ON-DFA (with H first-order factors). In fact, ARI ranges between 0 and 1, where 1 implies that the estimated partition $\widehat{\mathbf{V}}$ is identical to the true $\mathbf{V}$.

We therefore counted the percentage of times that ARI is equal to 1 and the average of ARI in the 500 experiments. The model was also evaluated with Bayesian Information Criterion (BIC, Schwarz, 1978), Akaike's Information Criterion (AIC, Akaike, 1973) and Root Mean Square of Residuals (RMSR, Hooper et al., 2008) for the correlation matrix.

In Fig. 3, we showed examples of generated correlation matrices with different level of errors in order to allow the reader to visually appreciate the meaning of low, medium and high errors. As we can see, in Fig. 3a, b, f, g the correlation matrices were generated with low error and the block diagonal structure is clearly visible, whereas in Fig. 3c, d, h, i the correlation matrices were generated with medium error and the block diagonal structure is less clear, but still visible. Finally, in Fig. 3e, j the structure is entirely invisible.

Tables 1 and 2 report the performances of the model values of the following indices: the average of ARI, BIC, AIC and RMSR on the 500 generated datasets and the percentage of ARI equal to 1. On average, in Table 1, ARI decreases from 1 to 0.85 when the error increases, and the model detects the real correlation structure most of the times for a equal to 0.33, 0.66, 0.85, 1. Inferior performances in terms of % of ARI equal to 1 for $a = 1.33$ are recorded. Nevertheless, the performances of the model in scenario $n = 200$, $J = 20$, $H = 5$ are appreciable even when the block diagonal structure tends to be less visible. RMSR always indicates a good fit of the model (i.e., values always lower than or equal to 0.08). In Table 2, ARI's mean values are always greater than 0.78. Thus, the performances of the model in scenario $n = 200$, $J = 50$, $H = 10$ are noticeable considering the average of ARI as figure of merit, even when the model identifies the true partition only in 20% of all cases. RMSR never assumes a value higher than 0.09. Even if BIC and AIC, as expected, show a decreasing performance of 2O-DFA as the error increases, the slope of this decrement tends to reduce when the error increases. Additionally, BIC and AIC of the optimal solution of 2O-DFA (BIC_{opt} and AIC_{opt}, respectively) were compared with BIC and AIC of a random block diagonal structure of model (6) (BIC_{rand} and AIC_{rand}, respectively), i.e., with an initial random solution of model (6). This comparison was included because individual AIC and BIC values are not interpretable as they contain arbitrary constants and are much affected by sample size. A comparison was therefore imperative (Burnham & Anderson, 2002). Tables 1 and 2 show that BIC_{rand} and AIC_{rand} are approximately between 5.4 and 1.21 times worse than BIC_{opt} and AIC_{opt} in all scenarios, showing how the optimal solution of 2O-DFA is different from a random solution as the error increases.

The simulation study additionally examined the ability of the model to build reliable first-order factors via Cronbach's α (Cronbach, 1951), which can be seen as the expected correlation between all pairs of MVs used to specify the concept (Chaouachi & Rached, 2012; Fornell & Larcker, 1981; Kline, 2000). Cronbach's α ranges between negative infinity and 1. Tables 3 and 4 show evaluations of internal consistency of the first-order factors estimated by the model.

The results in Table 3 could lead the reader to think that the proposed model does not provide reliable first-order factors when a is equal to 1 and 1.33; but looking at Table 4, we can state

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

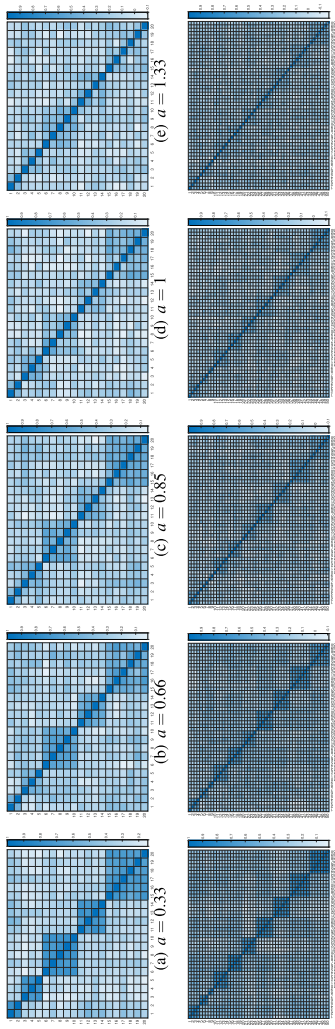


FIGURE 3.
Heatmaps of examples of correlation matrix produced by the simulation study with different levels of error. First row: scenario $n = 200$, $J = 20$, $H = 5$; second row: scenario $n = 200$, $J = 50$, $H = 10$

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

TABLE 1.
Simulated datasets with $n = 200$, $J = 20$, $H = 5$ and different levels of error

Criterion	$a = 0.33$	$a = 0.66$	$a = 0.85$	$a = 1$	$a = 1.33$
ARI	1	1	0.99	0.97	0.85
% ARI	100	100	92	72	26
BIC	1131	2155	2402	2523	2660
AIC	1314	2038	2285	2406	2543
BIC_{rand}/BIC_{opt}	5.1	2.11	1.64	1.44	1.21
AIC_{rand}/AIC_{opt}	5.4	2.17	1.67	1.45	1.22
RMSR	0.02	0.03	0.04	0.05	0.08

a sets the level of error; rand and opt represent the solutions of a random block diagonal structure model and 2O-DFA, respectively.

TABLE 2.
Simulated datasets with $n = 200$, $J = 50$, $H = 10$ and different levels of error

Criterion	$a = 0.33$	$a = 0.66$	$a = 0.85$	$a = 1$	$a = 1.33$
ARI	1	0.99	0.98	0.96	0.78
% ARI	100	84	55	38	20
BIC	7140	8990	9621	10010	10327
AIC	6854	8703	9334	9723	10040
BIC_{rand}/BIC_{opt}	4.6	2.22	1.70	1.49	1.26
AIC_{rand}/AIC_{opt}	4.7	2.27	1.73	1.50	1.27
RMSR	0.02	0.04	0.04	0.06	0.09

a sets the level of error; rand and opt represent the solutions of a random block diagonal structure model and 2O-DFA, respectively.

TABLE 3.
Percentage of times Cronbach's alpha (computed for each subset of MVs) > 0.9 with different scenario and different levels of error

Scenario	$a = 0.33$	$a = 0.66$	$a = 0.85$	$a = 1$	$a = 1.33$
$n = 200$, $J = 20$, $H = 5$	90.5	70.9	14.4	7.3	6.8
$n = 200$, $J = 50$, $H = 10$	85.3	84.2	33	28.2	21.7

TABLE 4.
Percentage of times Cronbach's alpha (computed for each subset of MVs) < 0.7 with different scenario and different levels of error

Scenario	$a = 0.33$	$a = 0.66$	$a = 0.85$	$a = 1$	$a = 1.33$
$n = 200$, $J = 20$, $H = 5$	0	0	9.2	11.1	14.2
$n = 200$, $J = 50$, $H = 10$	0.4	0.5	6.2	8.3	10.8

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

TABLE 5.
Comparison among methods to detect the SSM on 500 datasets

Methods	$a = 0.33$	$a = 0.66$	$a = 0.85$	$a = 1$	$a = 1.33$
Scenario $n = 200$, $J = 20$, $H = 5$					
Second-Order DFA	1 (100)	1 (100)	0.99 (95)	0.96 (69)	0.88 (31)
EFA + Varimax	1 (100)	0.99 (98)	0.95 (73)	0.91 (51)	0.69 (10)
EFA + Promax	1 (100)	0.99 (98)	0.95 (75)	0.90 (48)	0.67 (8)
EFA + Quartimin	1 (100)	0.99 (98)	0.95 (77)	0.92 (53)	0.69 (11)
EFA + Geomin	1 (100)	0.99 (98)	0.95 (76)	0.91 (53)	0.70 (10)
SSFA	1 (100)	0.95 (59)	0.76 (5)	0.59 (0)	0.35 (0)
FANC	1 (100)	1 (100)	0.96 (75)	0.92 (50)	0.85 (28)
Scenario $n = 200$, $J = 50$, $H = 10$					
Second-Order DFA	1 (100)	0.99 (89)	0.99 (61)	0.95 (32)	0.79 (25)
EFA + Varimax	0.99 (99)	0.98 (65)	0.95 (24)	0.89 (3)	0.64 (0)
EFA + Promax	0.99 (99)	0.98 (70)	0.95 (29)	0.88 (4)	0.61 (0)
EFA + Quartimin	0.99 (99)	0.98 (72)	0.97 (53)	0.89 (21)	0.65 (6)
EFA + Geomin	0.99 (99)	0.99 (73)	0.97 (54)	0.90 (23)	0.66 (6)
SSFA	0.43 (0)	0.40 (0)	0.39 (0)	0.33 (0)	0.18 (0)
FANC	0.22 (0)	0.22 (0)	0.21 (0)	0.21 (0)	0.18 (0)

The results indicates the average of ARI (% of ARI equal to 1).

that Cronbach's α, computed for each subset of MVs, is always higher than 0.70 in settings with $a = 0.33$ and $a = 0.66$ for the scenario $n = 200$, $J = 20$, $H = 5$ and almost always for the scenario $n = 200$, $J = 50$, $H = 10$. Overall, most of the time the method is able to get reliable first-order factors, at least 85.8%.

We also conducted an additional study to assess the performance of the 2ON-DFA in detecting a SSM, with respect to EFA followed by rotation, where each variable was assigned to the factor with the highest loading (in absolute value) resulting from EFA followed by a rotation method. For the rotation, we considered the four widely used methods: Varimax, Promax, Quartimin and Geomin (Abdi, 2003; Browne, 2001). Moreover, 2O-DFA's performance was compared with the results obtained by SSFA (Adachi & Trendafilov, 2018) and FANC (Hirose & Yamamoto, 2014). SSFA was performed imposing the number of common factors equal to H, a convergence tolerance equal to 10^5 and 50 different initial solutions for the loading matrix. Let us recall that FANC is formulated as follows:

$$\min_{\mathbf{A},\boldsymbol{\Sigma}_{\mathbf{y}},\boldsymbol{\Psi}_{\mathbf{x}}} l(\mathbf{A}, \boldsymbol{\Sigma}_{\mathbf{y}}, \boldsymbol{\Psi}_{\mathbf{x}}) + \rho P_{\gamma}(\mathbf{A}) \tag{31}$$

where $P_{\gamma}(\mathbf{A})$ penalizes $\mathbf{A}$ to have nonzero elements, with γ and ρ being tuning parameters. Since finding the tuning parameters that provide sparsest loading matrices was exceedingly onerous, FANC was performed imposing $\gamma = 1.01$ and with ρ selected according to BIC in order to have a SSM to compare with the other methods.

The parameters considered in this Section were reused also for this comparison: 500 datasets, $n = 200$, $J = 20, 50$, $H = 5, 10$ and $a = 0.33, 0.66, 0.85, 1, 1.33$. In Table 5, it is possible to see that 2ON-DFA outperforms the other methods considering both the average of ARI between the generated partition of variables and the partition predicted by the methodologies and the percentage of ARI equal to 1 as figures of merit.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

It is evident that our proposal is able to detect the SSM better than both the sequential application of EFA followed by rotation, SSFA and FANC, while also simultaneously estimating the correlation between the first-order factors.

6. Application

Our proposal was applied on the OECD data on *Better Life Index 2015*, [1] an initiative started in 2011 which focused on statistics that capture important aspects of life and that characterize the quality of people's lives. The dataset is formed by 34 countries and 24 variables which reflect two pillars that OECD identified as essential to well-being:

Material Living Conditions (MLC) is specified by three dimensions (characterized by their own sub-dimensions): *Housing* (*Dwellings without basic facilities*, *Housing expenditure* and *Rooms per person*), *Income* (*Household net adjusted disposable income* and *Household net financial wealth*) and *Jobs* (*Employment rate*, *Job security*, *Long-term unemployment rate* and *Personal earnings*).

Quality of Life (QL) is identified by eight dimensions (characterized by their own sub-dimensions): *Community* (*Quality of social support network*), *Education* (*Educational attainment*, *Student skills* and *Years in education*), *Environment* (*Air pollution* and *Water quality*), *Civic Engagement* (*Consultation on rule-making* and *Voter turnout*), *Health* (*Life expectancy* and *Self-reported health*), *Life Satisfaction* (*Life Satisfaction*), *Safety* (*Assault rate* and *Homicide rate*) and *Work-Life Balance* (*Employees working very long hours* and *Time devoted to leisure and personal care*).

2ON-DFA is scale invariant, thus the model results equivariant under an affine transformation of the data. In this application, MVs were standardized, yet the re-scaling normalization (also known as Min-Max normalization, which converts MVs to the same scale and makes all the values fall within the interval [0, 1]) could also be used obtaining exactly the same results.

Since 2ON-DFA identified a system of nonnegative loadings, the first application was used to detect eventual MVs that could measure a negative component of the general latent construct, in our case the well-being. The number of specific constructs H was fixed equal to the minimum 2. When MVs had positive loadings, it meant that they were positive dimensions of the general latent construct.

However, if H is optimally chosen and there still is a MV with no significant loading, this suggests discarding the MV from the analysis because it is irrelevant within the model; the MV can actually be confounding for the analysis and it could be removed without incurring into a significant loss of information. Finally, if some MVs have exactly zero loadings on the two factors of the model, this means that they measure a negative component of the general latent construct and therefore that they must be reversed.

In our analysis, MVs (1), (2), (7), (8), (14), (21), (22) and (23) loaded zero on the two factors (i.e., a row with zero loadings for each of the above MVs was observed). In fact, these MVs measured a negative component of well-being and therefore had no positive loadings to show; thus, they were reversed by changing the sign (i.e., by multiplying each MV by -1). When reapplying the model with $H = 2$, all MVs loaded on factors with a positive loading except for MVs (2) (*Housing expenditure*) and (16) (*Consultation on rule-making*). Under the hypothesis $H = 2$, the MVs (2) and (16) are, therefore, irrelevant in the model and were consequently removed. The subsequent analyses were repeated without these two MVs.

In the first application of the model, the 22 MVs were constrained to define the MLC and QL factors as indicated by the OECD. The results of the analysis are reported in Table 6 columns

[1] Available at https://stats.oecd.org/Index.aspx?DataSetCode=BLI2015.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

1 and 2. The methodology resulted in a second-order nonnegative disjoint confirmative factor analysis because all MVs were forced a priori to load on the specified MLC and QL as indicated by OECD, and all loadings were hence constrained to be nonnegative. BIC and AIC resulted 661.76 and 591.55, respectively. The two identified factors were not unidimensional constructs, since the second largest eigenvalue of the variance–covariance sub-matrices related to the two subsets of MVs were clearly larger than 1 and equal to 1.88 and 2.16. However, the reliability or internal consistency of each factor measured by Cronbach's α was good (Kline, 2000) and equal to 0.87 for both subsets. The internal consistency is the extent to which MVs, associated with a factor, measure a unique theoretical construct. The observed results suggested verifying if the selected model could be statistically confirmed by the data of 2015. A common approach for model selection is to use theory to specify an initial model, in our case the MLC and QL dimensions defined by OECD, and then to use the likelihood ratio χ^2_ν test to decide whether the model is confirmed or should be simplified or expanded (Bollen, 1989). However, this test is problematic in practice due to its sensitivity to the sample size and to its tendency to reject good models. This situation is well-known in literature and for this reason Marsh, Balla and McDonald (1988) and several other authors proposed more than 30 alternative measures. Among these, we considered the two information criteria BIC and AIC that proved to work well on SEM and confirmatory factor analysis problems (Raftery, 1995).

In order to examine how much the latent constructs MLC and QL change by considering a less constrained model, the 2ON-DFA was repeated by forcing only a single MV for each of the 11 dimensions to load on MLC or QL as indicated by OECD. In practice, the MV that in the first analysis had the maximal loading within its dimension was forced to load as indicated by OECD on MLC or QL. The dimension *Civic engagement* (7) was characterized only by the MV *Voter turnout* which had a low loading (0.21) in the OECD proposal. This MV was thus left free to load on one of the two latent constructs, either MLC or QL. The 10 constrained MVs are reported in bold in columns 3 and 4 of Table 6. The construct MLC has: *Housing* represented by *Rooms per person* (3); *Income by Household net adjusted disposable income* (4); and *Jobs* by *Personal earnings* (9). The construct QL has: *Community* represented by *Quality of support network* (10); *Education* by *Student skills* (12); *Environment* by *Water quality* (15); *Health* by *Life expectancy* (18); *Life Satisfaction* by *Life Satisfaction* (20); *Safety* by *Homicide rate* (22); *Work-Life Balance* by *Employees working very long hours* (23). The rest of MVs was left free to load on one of the two latent constructs.

After constraining 10 out of the 22 MVs to define MLC and QL as indicated by OECD, it was expected that the new analysis confirmed that also the rest of the MVs followed the two hypothesized latent constructs. This proved false for 7 of the remaining 11 MVs. In Table 6, columns 3 and 4 show the optimal solution. In this second analysis, the values of BIC and AIC were considerably lower, and equal to 642.22 and 572.01, respectively. According to the study of Raftery (1995), there is a "very strong" evidence that this model is better than the one proposed by OECD, since there is a difference in BIC between models larger than 10 ($\text{BIC}_{\text{OECD}} - \text{BIC}_{\text{Con}} = 20.54$). In this new application of our model, MVs *Employment rate*, *Job security* (reversed), *Long-term unemployment rate* (reversed) showed to be better related to the *Quality of Life*, while *Voter turnout*, *Self-reported health* and *Assault rate* (reversed) were better related to *Material Living Conditions*. Therefore, the *Quality of Life* defined by this constrained model also included the need for a secure and a long-term job, while the *Material Living Conditions* included the need for a good personal health condition, personal safety and interest to become involved in the political process.

It remained to understand if these two latent constructs defined could be further improved by leaving all the MVs free to load on one of the two latent constructs. 2ON-DFA was therefore applied for a third time, yet without constraints on the MVs. The best solution over 100 runs was retained and the results were reported in Table 6, columns 5 and 6. Although the values of

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

TABLE 6.
Analysis of different second-order factor analysis models for defining two dimensions of wellbeing: material living conditions and quality of life

Column		1	2	3	4	5	6
		OECD		Constrained		Unconstrained	
		MLC	QL	MLC	QL	MLC	QL
First-order factors							
1. Housing	1	0.61			0.62	0.64	
	2						
	3	0.80		**0.81**		0.82	
2. Income	4	0.95		**0.95**		0.92	
	5	0.75		0.74		0.71	
3. Jobs	6	0.63			0.85		0.85
	7	0.27			0.42		0.45
	8	0.33			0.39		0.41
	9	0.94		**0.94**		0.96	
4. Community	10		0.69		**0.60**	0.56	
5. Education	11		0.62		0.58		0.54
	12		0.68		**0.55**	0.44	
	13		0.62		0.53		0.54
6. Environment	14		0.56		0.52		0.48
	15		0.78		**0.95**		0.99
7. Civic engagement	16						
	17		0.21	0.43		0.43	
8. Health	18		0.59		**0.55**	0.70	
	19		0.29	0.52		0.53	
9. Life satisfaction	20		0.44		**0.62**	0.71	
10. Safety	21		0.63	0.38		0.39	
	22		0.67		**0.42**		0.38
11. Work-life balance	23		0.67		**0.55**		0.51
	24		0.54		0.40	0.37	
Second-order factor		0.88	0.89	0.91	0.88	0.88	0.88
Communality		3.95	4.87	3.59	5.28	5.61	3.29
Cronbach's α		0.87	0.87	0.88	0.85	0.90	0.82
Unidimensionality		1.87	2.16	2.88	1.13	2.24	1.80
BIC		661.76		642.22		630.73	
AIC		591.55		572.01		560.52	
Discrepancy		192.50		140.54		134.16	
Total communality		10.40		10.42		10.46	

Constrained MVs are reported in bold in columns 3 and 4.

BIC and AIC decreased to 630.73 and 560.52, the reduction was not as strong as in the second application of our proposal. However, the difference of BIC ($BIC_{Con} - BIC_{Unc} = 12.65$) was still extremely relevant. Differently from the OECD model, *Life Satisfaction*, *Community* (*Quality of support network*), *Students Skills* and *Time devoted to leisure and personal care* loaded on MLC in the unconstrained model. *Life expectancy*—the other MV of the *Health* dimension—loaded on MLC. Therefore, the MLC construct of the unconstrained model also included *Life Satisfaction*, *Community* and part of *Work-Life Balance*. It is worth observing that the subsets of specific MVs proposed by OECD to define the 11 dimensions were frequently "incoherent", in the sense that they loaded on different latent constructs. For example, the two MVs *Employees working very*

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

long hours and *Time devoted to leisure and personal care* that defined the dimension *Work-Life Balance*, actually loaded on MLC and QL, respectively. The same happens for *Safety*, *Education* and *Jobs*. This thereby suggested that the proposed MVs, when considered all together, were not coherent with the hypothesized theoretical construct.

6.1. The optimal number of dimensions for well-being

The analysis was repeated for different values of H in order to find the solution with the lowest BIC and AIC and to understand if there were more than two pillar dimensions (MLC and QL). For $H = 5$, BIC and AIC resulted as the lowest, with 590.07 and 515.27, respectively. Note that MVs (2) and (16) were re-inserted in the analysis; however, these increased drastically BIC and AIC and were discarded again.

The five factors were almost all unidimensional since the second largest eigenvalue of the variance–covariance sub-matrices related to the five subsets was not strongly larger than 1 and, in four subsets, it was actually lower (1.024, 0.619, 0.770, 0.795, 0.261). Cronbach's α was equal to 0.88, 0.86, 0.82, 0.77, 0.85, respectively, thus confirming the good internal consistency of factors. The results are reported in Table 7.

The first latent construct was mostly MLC including *Civic Engagement* (*Voter turnout*), *Health* (*Life expectancy*), *Life Satisfaction* and *Work-Life Balance*. This factor was mainly formed by a subset of MVs defining MLC in the unconstrained solution of the 2-factor model. It is interesting to observe that there was another factor for MLC, the fifth, dedicated to the dimension *Jobs* with *Job insecurity* (reversed) and *Long-term unemployment rate* (reversed). The other three factors eventually split the QL into first-order factors. The second factor was formed by *Education* (*Educational attainment*, *Student skills*) and *Safety* (*Assault rate*, *Homicide rate* both reversed) and therefore measured the level of safety and education in *Society* as a relevant aspect of the QL. The third factor measured the *Quality of Society* in terms of *Community* (*Quality of social support network*), *Environment* (*Water quality*), *Years in education* and *Employment rate*. Therefore, the latter measured fundamental aspects for *Society* such as the quality of the social network support, the quality of the water, the quality of education and the quality of working life. The fourth factor measured the *Quality of Habitat for Humanity* based on *Dwellings without basic facilities* (reversed), *Air pollution* (reversed), *Self-reported health* and *Employees working very long hours* (reversed). The fourth factor, thus, focused on the quality of the habitat for the humanity based on decent and affordable housing, good air quality, the perceived health in the habitat and the working stability of the habitat.

7. Conclusion

Our proposal, *Second-Order Disjoint Factor Analysis*, allows modeling an unknown hierarchical structure of the MVs with two orders. It is a second-order factor analysis with a disjoint structure for the MVs. Each subset of MVs is detected to be reliable, that is, MVs related to a first-order factor consistently measure a unique theoretical construct. The second-order represents the general factor which summarizes the common information related to the H specific first-order factors. A nonnegative version of the model is presented which entails that MVs are concordant with the related factor and, therefore, the associated factor loadings are constrained to be nonnegative. According to the distributional assumptions, the maximum likelihood estimation of the models allows us to make inference on the parameters. Since the maximization of the likelihood represents a discrete and continuous problem that cannot be solved by a quasi-Newton type algorithm, the estimation of all parameters is simultaneously obtained following a cyclic block coordinate descendent algorithm. An application about well-being presents the characteristics of

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

TABLE 7.
Analysis of second-order factor analysis model for defining five dimensions of well-being

Column		1 MLC	2 ES	3 QS	4 QSHH	5 J
First-order factors						
1. Housing	1				0.94	
	2					
	3	0.81				
2. Income	4	0.94				
	5	0.72				
3. Jobs	6			0.84		
	7					0.86
	8					0.86
	9	0.97				
4. Community	10			0.56		
5. Education	11		0.66			
	12		0.76			
	13			0.55		
6. Environment	14				0.58	
	15			0.99		
7. Civic engagement	16					
	17	0.44				
8. Health	18	0.70				
	19				0.49	
9. Life satisfaction	20	0.71				
10. Safety	21		0.92			
	22		0.76			
11. Work-life balance	23				0.70	
	24	0.35				
Second-order factor		0.92	0.35	0.65	0.82	0.77
Communality		4.26	2.44	2.32	1.95	1.48
Cronbach's α		0.88	0.86	0.82	0.77	0.85
Unidimensionality		1.02	0.62	0.77	0.80	0.26
BIC				590.07		
AIC				515.27		
Discrepancy				104.17		
Total communality				15.11		

the new methodology, which shows in detail the limitation of the confirmatory analysis when the framework proposed is not respected by the data. Four models are proposed: a fully confirmatory analysis (framework and number of factors are given), a semi-confirmatory analysis (main relations and number of factors are given), an exploratory analysis (only the number of factors is given) and a full exploratory analysis (the number of factors is selected by an information criterion). This latter shows the real potential of our proposal: the detection of the best partition of MVs, and thus of the latent concepts and the related factors; and, moreover, their aggregation into a second-order factor.

The paper, thus, provides a generalization of both hierarchical factor analysis and non-orthogonal disjoint factor analysis, and proposes a constrained version with the additional request of nonnegativity of loadings.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

Nevertheless, some further challenges are foreseen within our research. Our future goal is to develop the inclusion of cross-loadings in the SSM in order to increase the fit of the model using the work by Vichi (2017). Indeed, by applying 2O-DFA it may occur that sample covariance matrix is block diagonal, but that there are relevant cross-loadings between blocks, i.e., some variables belonging to different blocks are highly correlated. Here, the 2O-DFA should include the estimation of cross-loadings and therefore define factors that are not disjoint. Moreover, another important development consists of the introduction of a time dependence in order to provide a model able to fit the phenomenon of study throughout the years. Two ideas are currently being developed: the first is that of encompassing a time dependence in the loading matrices, as already done by Maruotti and Vichi (2014) on the centroids matrices in their time-varying extension of k-means. The second idea consists of the inclusion of a hidden Markov model (HMM) in order to follow the evolution of MVs depending on both the first-order factors and on the general one (i.e., the second-order factor).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Abdi, H. (2003). Factor rotations in factor analyses. In *Encyclopedia of social sciences research methods*, pp. 792–795.

Adachi, K., & Trendafilov, N. T. (2018). Sparsest factor analysis for clustering variables: A matrix decomposition approach. *Advances in Data Analysis and Classification, 12*, 559–585.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199–213). Springer New York.

Anderson, T. W., & Rubin, H. (1956). Statistical inferences in factor analysis. *Proceedings of the Third Symposium on Mathematical Statistics and Probability, 5*, 111–150.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology. General Section, 28*(1), 97–104.

Bollen, K. A. (1989). *Structural equations with latent variables*.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111–150.

Burnham, K., & Anderson, D. (2002). *Model selection and inference: A practical information-theoretic approach* (2nd ed.). Springer.

Cattell, R. B. (1947). Confirmation and clarification of primary personality factors. *Psychometrika, 12*, 197–220.

Cattell, R. B. (1978). Higher-order factors: Models and formulas. In *The scientific use of factor analysis in behavioral and life sciences* (pp. 192–228). Springer US.

Chaouachi, S. G., & Rached, K. S. B. (2012). Perceived deception in advertising: Proposition of a measurement scale. *Journal of Marketing Research and Case Studies, 2012,* 1.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

de Raad, B., & Mlačić, B. (2015). Big five factor model, theory and structure. In J. D. Wright (Ed.), *International encyclopedia of the social behavioral sciences* (2nd ed., pp. 559–566). Elsevier.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41*(1), 417–440.

Eysenck, H. J. (1970). *The structure of human personality* (3rd ed.). Methuen.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Goldberg, L. (2006). Doing it all bass-ackwards: The development of hierarchical factor structures from the top down. *Journal of Research in Personality, 40*(4), 347–358.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Erlbaum.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicator Research, 141*, 61–94.

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press

Hirose, K., & Yamamoto, M. (2014). Estimation of an oblique structure via penalized likelihood factor analysis. *Computational Statistics and Data Analysis, 79*, 120–132.
Holzinger, K. J. (1944). A simple method of factor analysis. *Psychometrika, 9*, 257–262.
Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41–54.
Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53–60.
Horst, P. (1965). Factor analysis of data matrices.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193–218.
Jöreskog, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika, 31*(2), 165–178.
Jöreskog, K. G. (1969). A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika, 34*(2), 183–202.
Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika, 43*(4), 443–477.
Jöreskog, K. G. (1979). A general approach to confirmatory maximum likelihood factor analysis with addendum. In K. Jöreskog, D. Sörbom, & J. Magidson (Eds.), *Advances in factor analysis and structural equation models* (pp. 21–43). Abt Books.
Kline, P. (2000). *The handbook of psychological testing* (2 ed.). Routledge.
Lawson, C. L., & Hanson, R. J. (1974). *Solving least squares problems*. Prentice-Hall.
Le Dien, S., & Pages, J. (2003). Analyse factorielle multiple hierarchique. *Revue de statistique applique, 51*(2), 47–73.
Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*(3), 391–410.
Maruotti, A., & Vichi, M. (2014). Time-varying clustering of multivariate longitudinal observations. *Communications in Statistics - Theory and Methods, 45*, 430–443.
Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics, 41*(12), 1513–1523.
OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.
Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association, 66*(336), 846–850.
Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464.
Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association.
Thompson, G. H. (1934). The meaning of "i" in the estimate of "g." *British Journal of Psychology, 25*(1), 92–99.
Thompson, G. H. (1948). *The factorial analysis of human ability*. Houghton Mifflin.
Thompson, G. H. (1951). *The factorial analysis of human ability* (5th ed.). Houghton Mifflin.
Thurstone, L. L. (1947). *Multiple factor analysis*. The University of Chicago press.
Vichi, M. (2017). Disjoint factor analysis with cross-loadings. *Advances in Data Analysis and Classification, 11*(3), 563–591.
Wherry, R. J. (1959). Hierarchical factor solutions without rotation. *Psychometrika, 24*(1), 45–51.
Wherry, R. J. (1975). Underprediction from overfitting: 45 years of shrinkage. *Personnel Psychology, 28*(1), 1–18.
Wherry, R. J. (1984). *Contributions to correlational analysis*. Academic Press.
Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113–128.
Zangwill, W. I. (1969). *Nonlinear programming: A unified approach*. Prentice-Hall.

Manuscript Received: 23 JUL 2020
Final Version Received: 16 JUN 2021
Published Online Date: 17 AUG 2021

https://doi.org/10.1007/s11336-021-09799-6 Published online by Cambridge University Press