

This is a “preproof” accepted article for *Psychometrika*.

This version may be subject to change during the production process.

DOI: 10.1017/psy.2025.6

ULTRAMETRIC FACTOR ANALYSIS FOR BUILDING HIERARCHIES OF RELIABLE AND UNIDIMENSIONAL LATENT CONCEPTS

MARIAELENA BOTTAZZI SCHENONE

DEPARTMENT OF STATISTICAL SCIENCES, SAPIENZA UNIVERSITY OF ROME, ROME, ITALY

CARLO CAVICCHIA

ECONOMETRIC INSTITUTE, ERASMUS UNIVERSITY ROTTERDAM, ROTTERDAM, THE
NETHERLANDS

MAURIZIO VICHI

DEPARTMENT OF STATISTICAL SCIENCES, SAPIENZA UNIVERSITY OF ROME, ROME, ITALY

GIORGIA ZACCARIA

DEPARTMENT OF STATISTICS AND QUANTITATIVE METHODS, UNIVERSITY OF
MILANO-BICOCCA, MILAN, ITALY

March 4, 2025

Correspondence should be sent to

E-Mail: mariaelena.bottazzisohenone@uniroma1.it

This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

ULTRAMETRIC FACTOR ANALYSIS FOR MODELING RELIABLE AND
UNIDIMENSIONAL HIERARCHICAL FACTORS

Abstract

This paper introduces a novel methodology to model the hierarchical dependence structure of manifest variables. This is done by reconstructing their correlation matrix considering a hierarchy of latent factors which forms an ultrametric correlation matrix. The proposed ultrametric factor analysis model will be shown to obtain reliable, unidimensional, and unique hierarchical factors. This approach addresses the limitations of traditional factor analysis methods that often oversimplify the multidimensional and complex relationships among manifest variables. The paper provides an in-depth mathematical description of the proposed model, as well as an algorithm for parameter estimation. An extensive simulation study with 3,000 generated samples validates the proposal across twelve different scenarios. Finally, we demonstrate the potential of the proposed model using a real data set that is a benchmark in psychological research.

Key words: hierarchical factor analysis, ultrametric correlation matrix, latent variables, higher-order models.

1. Introduction

Latent concepts represent aspects of reality that cannot be directly observed, such as quality of life, sustainable development, poverty, happiness, and personality traits. These concepts can be inferred through observable Manifest Variables (MVs) and are essentially measurable by means of a Latent Variable (LV) that is statistically related with two or more MVs. In general, latent concepts need statistical models to comprehensively and synthetically represent them through a mathematical formalization of the observed data and their relationships, i.e., the measurement model (Bollen and Bauldry, 2011, among others). However, the representation of a multidimensional latent concept through a single LV might result in a simplistic portrayal, by inadequately capturing the intricate nature of the complex phenomenon at hand. Exploratory Factor Analysis (EFA), pioneered by Spearman (1904), is one of the widely used statistical methods to identify LVs, often called *factors*, which are assumed to generate MVs and reflect into their linear relationships. EFA has been extensively applied to identify complex constructs like personality traits, intelligence, attitudes, and mental health in psychology; social stratification, cultural values, and interpersonal relationships in sociology; consumer preferences, brand perceptions, and product attributes in marketing; and quality of life, health behaviors, and disease risk factors in health sciences. Additionally, if a theoretical factorial structure can be hypothesized a priori, Confirmatory Factor Analysis (CFA), and more generally Structural Equation Model (SEM), can be used to test its fit to the observed data (Brown, 2015). Nonetheless, it may sometimes happen that a reduced set of factors, which are uncorrelated in EFA, is insufficient to analyze and reproduce the complexity of the phenomenon under study, since hierarchical relationships among its multiple factors exist.

Advanced approaches for exploring hierarchies of LVs underlying the data include hierarchical models, which organize various facets of a construct into hierarchical structures, extending from the original MVs to the ultimate General Factor (GF) through a refined group of Specific Factors (SFs). The GF represents the highest abstraction of the phenomenon under study, while the SFs denote dimensions that, in turn, measure specific concepts describing the primary components of the phenomenon under study. SFs are LVs less abstract than GFs and are more correlated with MVs, allowing a better understanding of the phenomenon. This hierarchical approach is necessary

to provide a more precise and comprehensive understanding, steering away from the pitfalls of oversimplification associated with a single factor or with a group of uncorrelated factors. Higher-Order Factor Analysis (HOFA, Gorsuch, 1983), Second-Order Disjoint Factor Analysis (Cavicchia and Vichi, 2022), both based on EFA, and Hierarchical Disjoint Principal Component Analysis (Cavicchia et al., 2023), a hierarchical extension of the model introduced by Ferrara et al. (2016), are examples of hierarchical models. Specifically, HOFA involves a sequential application of EFA followed by non-orthogonal rotation to detect higher-level factors representing broader constructs, while the other two models use a simultaneous approach, either limited to two orders or related to principal component analysis instead of EFA, respectively. In psychometric studies, hierarchical concepts are ubiquitous. For example, as described by Cattell (1971) and Carroll (1993) in the context of intelligence studies, a hierarchical model might include multiple levels of abstraction, with lower levels representing specific abilities or tasks (e.g., memory, verbal comprehension, spatial reasoning) and higher levels representing broader factors (e.g., fluid intelligence, crystallized intelligence, general intelligence or *g* factor). Another example is the Five-Factor Model of personality (Cronbach, 1990; Costa and McCrae, 1992), which organizes personality traits into five factors, namely openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism, each represented by specific aspects. The aggregation of neuroticism, agreeableness, and conscientiousness on the one hand, and extraversion and openness to experience on the other, give rise to two higher-order dimensions, called alpha and beta (or stability and plasticity), respectively (DeYoung et al., 2002). Hierarchical models are also employed in organizational psychology to understand the structure of job performance (Campbell, 1990).

Unlike the aforementioned models, the methodology introduced in this paper addresses the modeling of the unexplained correlation between higher-order factors through a simultaneous approach that goes beyond two levels and identifies higher-order LVs from the hierarchical structure of MVs. The proposal is called Ultrametric Factor Analysis (UFA), as it extends the EFA model to the case of factors with a hierarchical structure, which mathematically corresponds to an ultrametric correlation matrix. However, if UFA is applied to data without a true hierarchical structure, first-order factor correlations will be close to zero, resulting in near-zero loadings for higher-order factors. This indicates the absence of a hierarchical structure. The model requires that the identi-

fied factors possess certain relevant properties to define a sound methodology that adheres to the principles of high-quality statistics, as clearly described in the European Code of Practice (European Commission, 2017). Good properties for a hierarchy of factors include internal consistency, reliability, unidimensionality, and content validity of the factors. The paper proposes an adequate measurement model for constructing a hierarchy of LVs with these desirable properties starting from a group of MVs. The results of the simulation study underscore the superiority of the UFA method compared to well-established factor analysis techniques, particularly in scenarios where the data exhibit a nested or hierarchical structure within factors. Traditional methods often struggle to accurately capture the complexity of such hierarchical data, leading to sub-optimal factor solutions. In contrast, UFA demonstrated a remarkable ability to identify and model the nested relationships, providing more precise and interpretable factor structures. We also apply UFA to the Cattell's dataset, which focuses on the factorial study of intelligence. Given the inherent hierarchical nature of this construct, where specific cognitive abilities feed into broader dimensions like crystallized and fluid intelligence, UFA is perfectly suited to model this type of data. Indeed, by capturing both the unidimensional and hierarchical relationships by means of an ultrametric correlation structure, UFA provides deep insights into how different cognitive abilities interrelate, offering a more nuanced understanding of intelligence that traditional factor analysis methods may fail to capture.

The rest of the paper is organized as follows. In Section 2, we present the notation used across all sections, along with fundamental concepts regarding ultrametric matrices, EFA and its constrained version based upon a sparse structure of the loading matrix. A comprehensive illustration of the proposed methodology is presented in Section 3. Section 4 describes the algorithm built to implement the proposed procedure. Different measures for model selection and for the choice of the number of factors are examined in Section 5. An extended simulation study designed to evaluate the model is illustrated in Section 6, while a practical application is the core of Section 7. Conclusions and key findings are reported in Section 8.

2. Notation and background

For the convenience of the reader, the notation and terminology common to all sections is listed here.

- n, J, Q Number of units, MVs, and primary factors, respectively;
 n_q Number of MVs associated with the q -th factor;
 \mathbf{X} $(n \times J)$ data matrix of n multivariate units;
 $\mathbf{1}_a, \mathbf{0}_a$ $(a \times 1)$ a -dimensional vector of ones or zeros elements, respectively;
 \mathbf{I}_a $(a \times a)$ generic identity matrix of order a ;
 \mathbf{J} $(n \times n)$ centering matrix, $\mathbf{J} = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}'_n$;
 \mathbf{Y} $(n \times Q)$ factor score matrix of n multivariate units;
 \mathbf{A} $(J \times Q)$ factor loading matrix;
 \mathbf{E} $(n \times J)$ error matrix;
 \mathbf{B} $(J \times J)$ matrix of the weights of the MVs;
 \mathbf{V}_Q $(J \times Q)$ membership matrix defining a partition of the J MVs in Q groups;
 \mathbf{C}_Q $(Q \times Q)$ diagonal matrix fixing a unique non-null loading for a group of MVs;
 $\mathbf{R}_\mathbf{X}$ $(J \times J)$ correlation matrix;
 \mathbf{R} $(J \times J)$ observed correlation matrix;
 $\mathbf{E}_\mathbf{R}$ $(J \times J)$ error matrix related to \mathbf{R} ;
 \mathbf{R}_W $(Q \times Q)$ within-concept consistency diagonal matrix, the diagonal elements represent the consistency within the q -th group of MVs;
 \mathbf{R}_B $(Q \times Q)$ between-concept correlation matrix, with zero diagonal and the off-diagonal elements denote the correlation between two factors;
 ${}_Q\mathbf{R}_\mathbf{X}$ $(J \times J)$ correlation structure of the MVs partitioned into Q groups;
 Ψ $(J \times J)$ correlation matrix of the errors;
 \mathbf{M}^+ $(m \times r)$ Moore-Penrose inverse of an $(r \times m)$ matrix \mathbf{M} , $\mathbf{M}^+ = (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'$.

2.1. Ultrametric correlation matrix

To introduce our new factorial methodology, we first need to recall the definition of an ultrametric correlation matrix, that is

Definition 1. A matrix $\mathbf{R}_\mathbf{X}$ is an ultrametric correlation matrix if all its elements $r_{jl} \in \mathbb{R}$, for $j, l = 1, \dots, J$, satisfy the following conditions

- (i) symmetry: $r_{jl} = r_{lj}$ for $j, l = 1, \dots, J$;
- (ii) non-negativity and unitary of the diagonal: $r_{jl} \geq 0$ for all $j, l = 1, \dots, J$, $r_{jj} = 1$ for all $j = 1, \dots, J$;
- (iii) positive semi-definiteness (psd): $\mathbf{xR}_\mathbf{X}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^J$;

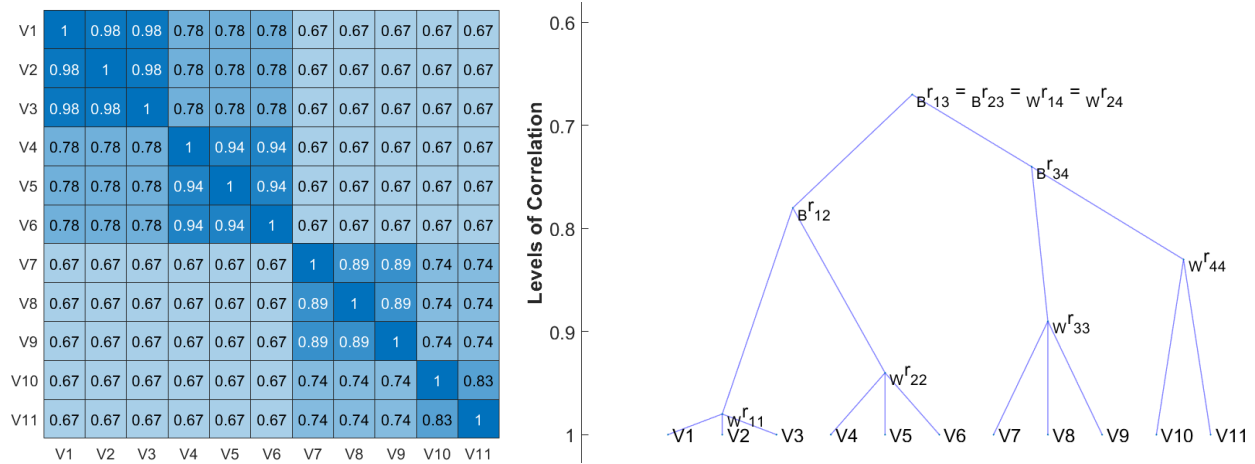
- (iv) ultrametricity: $r_{jl} \geq \min\{r_{jh}, r_{lh}\}$, for $j, l, h = 1, \dots, J$ or, equivalently, for each triplet $j, l, h = 1, \dots, J$, there exists a reordering $\{j, l, h\}$ of the elements s.t. $r_{jl} \geq r_{jh} = r_{lh}$.

In detail, $\mathbf{R}_{\mathbf{X}}$ is a $(2Q - 1)$ -ultrametric correlation matrix of order J if MVs are partitioned into Q groups and then they are agglomerated into $Q - 1, Q - 2, \dots, 1$ broader groups (Cavicchia et al., 2020). An example is provided in Fig. 1a. Each group $q = 1, \dots, Q$ in the initial partition of the MVs is characterized by a strong correlation among the MVs within the group. This correlation level, denoted by ${}_W r_{qq}$, can be stored as the q -th element of a diagonal matrix \mathbf{R}_W . The groups are also correlated with each other, though to a lesser extent. These correlations, denoted by ${}_B r_{qh}$, can be stored as off-diagonal elements of a hollow matrix \mathbf{R}_B . To arrange the MVs into groups, and those groups into larger groups (property (iii) for the whole matrix), the correlations within each group must be greater than the correlations between different groups, i.e., ${}_W r_{qq} \geq {}_B r_{qh}$, $q, h = 1, \dots, Q, h \neq q$. Additionally, the ultrametric property must hold for the correlations between groups. Hence, a $(2Q - 1)$ -ultrametric correlation matrix has a reduced number of parameters than a $(J \times J)$ correlation matrix since its off-diagonal elements assume one of the $(2Q - 1)$ different values, i.e., ${}_W r_{qq}$ $q = 1, \dots, Q$, and ${}_B r_{qh}$ ($q, h = 1, \dots, Q, h \neq q$). It is easy to show that a $(2Q - 1)$ -ultrametric correlation matrix is one-to-one associated with a hierarchy of Q groups of MVs (Cavicchia et al., 2020, Lemma 1). In fact, as shown in Fig. 1b, the first Q values ${}_W r_{qq}$ $q, q = 1, \dots, Q$, define the first-order aggregation levels of the hierarchy and represent the expected correlation within the first Q groups. The other $Q - 1$ values ${}_B r_{qh}$ $q, h = 1, \dots, Q, h \neq q$, identify the remaining $Q - 1$ levels and represent the expected correlation between groups of MVs.

2.2. Exploratory Factor and Disjoint Factor Analysis

Let \mathbf{x} be a J -dimensional random vector with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$. We assume that MVs are standardized; therefore, $\boldsymbol{\mu} = \mathbf{0}$ and the variance-covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}$ is the correlation of \mathbf{x} . The Exploratory Factor Analysis (EFA) assumes that the observed units are explained (reconstructed, generated) by the following model

$$\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{e}, \quad (1)$$



(a) $(2Q - 1)$ -ultrametric correlation matrix (b) Hierarchical tree representation of a $(2Q - 1)$ -ultrametric correlation matrix, with $Q=4$.

Figure 1: Example of a $(2Q - 1)$ -ultrametric correlation matrix, with $Q=4$, and its corresponding hierarchical tree. First-order groups in (b) are $G_1 = \{V_1, V_2, V_3\}$, $G_2 = \{V_4, V_5, V_6\}$, $G_3 = \{V_7, V_8, V_9\}$, $G_4 = \{V_{10}, V_{11}\}$. Higher-order groups are obtained as $G_5 = G_1 \cup G_2$, $G_6 = G_3 \cup G_4$, $G_7 = G_5 \cup G_6$.

where \mathbf{y} is a vector of non-observable Q random variables called factors, while \mathbf{A} is a $(J \times Q)$ matrix of unknown parameters called *factor loadings* that identifies the statistical relations between MVs and factors, and finally \mathbf{e} is a J -dimensional vector of non-observable random errors, whose elements are called the *specific* or *unique factors*. We shall assume that LVs and errors are centered, i.e., $\mathbb{E}(\mathbf{y}) = \mathbf{0}_Q$; $\mathbb{E}(\mathbf{e}) = \mathbf{0}_J$; factors are standardized, and uncorrelated for EFA, i.e., $\text{Cor}(\mathbf{y}) = \mathbf{I}_Q$ (we assume the orthogonal EFA hereinafter); and, finally, errors and factors are uncorrelated, i.e., $\text{Cor}(\mathbf{y}, \mathbf{e}) = \mathbf{0}_{Q \times J}$. For the Maximum Likelihood (ML) estimation, it is also assumed that $\mathbf{y} \sim \mathbf{N}_Q(\mathbf{0}_Q, \mathbf{I}_Q)$ and $\mathbf{e} \sim \mathbf{N}_J(\mathbf{0}_J, \mathbf{\Psi})$, where $\mathbf{\Psi}$ is a diagonal positive definite matrix. Given these assumptions, it can be found that $\mathbf{x} \sim \mathbf{N}_J(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\mathbf{x})$, where

$$\boldsymbol{\Sigma}_\mathbf{x} = \mathbf{A}\mathbf{A}' + \mathbf{\Psi}. \tag{2}$$

The EFA model clearly has a problem of identification of \mathbf{A} from $\mathbf{A}\mathbf{A}'$, since if $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}$ is

an arbitrary orthogonal transformation of \mathbf{A} , then $\tilde{\mathbf{A}}\tilde{\mathbf{A}}' = \mathbf{A}\mathbf{A}'$. This indeterminacy is usually solved by rotating the factor loadings to create maximum contrast between columns of \mathbf{A} , e.g., by using varimax rotation method (Kaiser, 1958). This frequently simplifies the loading matrix and facilitates interpretation. In this spirit, Vichi (2017) introduced the Disjoint Factor Analysis (DFA), in which the solution is constrained to have the *simplest* loading matrix \mathbf{A} that allows each MV to be uniquely correlated with one of the Q LVs. Thus, \mathbf{A} is parametrized by considering the product

$$\mathbf{A}_{DFA} = \mathbf{B}\mathbf{V}_Q, \tag{3}$$

where $\mathbf{B} = \text{diag}(b_1, \dots, b_J)$ is a $(J \times J)$ diagonal matrix and \mathbf{V}_Q is a $(J \times Q)$ binary and row stochastic matrix. The elements of \mathbf{B} identify the levels of correlation between MVs and LVs, while \mathbf{V}_Q is a membership matrix indicating each MV to which LV is assigned. This parameterization induces a partition of MVs into Q disjoint groups and the definition of a LV for each subgroup of MVs. Therefore, the model in Eq. (2) can be written as

$$\Sigma_{\mathbf{X}} = \mathbf{B}\mathbf{V}_Q\mathbf{V}'_Q\mathbf{B} + \Psi. \tag{4}$$

The ML estimation is obtained supposing that a random sample of $n > J$ multivariate units $\mathbf{x}_i = [x_{i1}, \dots, x_{iJ}]'$, $i = 1, \dots, n$, so that $\mathbf{X} = [\mathbf{x}'_1, \dots, \mathbf{x}'_n]'$ is observed; thus, the reduced log-likelihood is

$$L(\mathbf{B}, \mathbf{V}_Q, \Psi; \mathbf{X}) = -\frac{nJ}{2} \ln(2\pi) - \frac{n}{2} \ln |\mathbf{B}\mathbf{V}_Q\mathbf{V}'_Q\mathbf{B} + \Psi| + \frac{n}{2} \text{tr}((\mathbf{B}\mathbf{V}_Q\mathbf{V}'_Q\mathbf{B} + \Psi)^{-1}\mathbf{R}), \tag{5}$$

where \mathbf{R} is the sample correlation matrix (equal to the sample variance-covariance matrix \mathbf{S} since the observed sample is standardized). It is worth noting that the maximization of $L(\mathbf{B}, \mathbf{V}_Q, \Psi; \mathbf{X})$ is equivalent to the minimization of the following discrepancy function with respect to \mathbf{B} , \mathbf{V}_Q and Ψ

$$D(\mathbf{B}, \mathbf{V}_Q, \Psi) = \ln |\mathbf{B}\mathbf{V}_Q\mathbf{V}'_Q\mathbf{B} + \Psi| + \text{tr}((\mathbf{B}\mathbf{V}_Q\mathbf{V}'_Q\mathbf{B} + \Psi)^{-1}\mathbf{R}) \tag{6}$$

such that

$$\mathbf{V}_Q = [v_{jq} : \forall v_{jq} \in \{0, 1\}] \quad (\text{binary}), \quad (7)$$

$$\mathbf{V}_Q \mathbf{1}_Q = \mathbf{1}_J \quad \text{i.e.} \quad \sum_{q=1}^Q v_{jq} = 1 \quad j = 1, \dots, J \quad (\text{row stochastic}), \quad (8)$$

$$\mathbf{B} = \text{diag}(b_1, \dots, b_J) \quad \text{with } b_j^2 > 0 \quad (\text{diagonal, non-null}), \quad (9)$$

$$\mathbf{V}'_Q \mathbf{B} \mathbf{B} \mathbf{V}_Q = \text{diag}(b_{\cdot 1}^2, \dots, b_{\cdot Q}^2) \quad \text{with } b_{\cdot q}^2 = \sum_{j \in G_q} b_j^2 \quad (\text{orthogonal, non-empty}), \quad (10)$$

where G_q indicates the group of MVs related to the q -th factor. DFA involves a mixed binary (in \mathbf{V}_Q) and continuous (in \mathbf{B} and Ψ) optimization problem, thus, a quasi-Newton method cannot be properly applied. Two coordinate descent algorithms (Zangwill, 1969) have been developed to solve it in Vichi (2017).

EFA and DFA models are scale invariant under linear transformation of the data matrix \mathbf{X} , useful for data normalization or standardization. However, DFA, differently from EFA, has the additional property to be identified (unless for a label switching), thus, no rotation is needed.

3. The Ultrametric Factor Analysis model

The Ultrametric Factor Analysis models hierarchical structures by providing a specific factor loading matrix capable of identifying the ultrametric correlation matrix describing the set of the supposed nested latent concepts. Therefore, let us formulate the mathematical form of the ultrametric correlation matrix that UFA can derive. First suppose that Q factors are identified according to the DFA model in Eq. (4) with the following parsimonious formulation of the loading matrix (Parsimonious DFA, PDFA)

$$\mathbf{A}_{PDFA} = \mathbf{V}_Q \mathbf{C}_Q, \quad (11)$$

where $\mathbf{C}_Q = \text{diag}(c_1, \dots, c_Q)$ is a diagonal matrix of order Q . The elements of \mathbf{C}_Q represent equal levels of correlation between each LV and the corresponding MVs. Therefore, the correlation structure of PDFA is

$${}_Q \mathbf{R}_X = \mathbf{V}_Q \mathbf{C}_Q^2 \mathbf{V}'_Q + \Psi, \quad (12)$$

where the loadings of the MVs belonging to the same group, i.e., associated with the same LV, are equal. After a proper permutation of the rows of \mathbf{V}_Q so as to have all the ones in each column of \mathbf{V}_Q contiguous, the correlation structure in Eq. (12) has the following block diagonal form

$${}_Q\mathbf{R}_X = \text{blkdiag}({}_W\mathbf{R}_1, \dots, {}_W\mathbf{R}_q, \dots, {}_W\mathbf{R}_Q) = \begin{bmatrix} {}_W\mathbf{R}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & {}_W\mathbf{R}_Q \end{bmatrix}. \tag{13}$$

In (13), blocks represent the correlation matrices ${}_W\mathbf{R}_q$ of the q -th group of MVs related to the q -th factor, for $q = 1, \dots, Q$, that is

$${}_W\mathbf{R}_q = (\mathbf{1}_{n_q}\mathbf{1}'_{n_q})c_q^2 + \mathbf{\Psi}_q, \tag{14}$$

where c_q^2 is the correlation between the q -th factor and the n_q associated MVs, and $\mathbf{\Psi}_q$ is a diagonal matrix such that $\text{diag}(\mathbf{\Psi}_q) = [\psi_{1q}, \dots, \psi_{n_qq}]' = \mathbf{1}_{n_q}(1 - c_q^2)$. It is worth noting that ${}_Q\mathbf{R}_X$, due to its block diagonal form, automatically satisfies the ultrametric conditions of Definition 1. However, for the non-orthogonal EFA, correlation between factors is hypothesized to be different from zero, i.e., $\text{Cor}(\mathbf{y}) = \mathbf{R}_y$. To model the latter, an higher-order correlation structure is required, corresponding to additional “layers” that account for it. For the second-order correlation structure, we have

$${}_{Q+1}\mathbf{R}_X = \mathbf{v}_{Q+1}c_{Q+1}^2\mathbf{v}'_{Q+1}, \tag{15}$$

where \mathbf{v}_{Q+1} – hereinafter, we denote \mathbf{v}_q as the q -th column of \mathbf{V}_Q – identifies a group of variables which is obtained as the sum of two columns h and q of \mathbf{V}_Q , i.e., $\mathbf{v}_{Q+1} = \mathbf{v}_h + \mathbf{v}_q$, with $h, q \in \{1, \dots, Q\}$, $h \neq q$, while c_{Q+1} is the loading of the $(Q + 1)$ -th factor corresponding to the square root of the correlation between factors h and q . The matrix formulation of Eq. (15) is

$${}_{Q+1}\mathbf{R}_X = \begin{bmatrix} \mathbf{0} & {}_B\mathbf{R}_{hq} & \mathbf{0} \\ {}_B\mathbf{R}_{hq} & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \tag{16}$$

where ${}_B\mathbf{R}_{hq} = (\mathbf{1}_{n_h}\mathbf{1}'_{n_q})c_{Q+1}^2$. Note that the matrix obtained as ${}_Q\mathbf{R}_X + {}_{Q+1}\mathbf{R}_X = \begin{bmatrix} {}_W\mathbf{R}_h & {}_B\mathbf{R}_{hq} & \mathbf{0} \\ {}_B\mathbf{R}_{hq} & {}_W\mathbf{R}_q & \vdots \\ \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}$ is ultrametric, since if the h -th and q -th groups are aggregated it must hold that ${}_W r_{hh}, {}_W r_{qq} \geq {}_B r_{hq}$ (see Subection 2.1). For the additional orders of correlation structures

from $Q + 2$ to $2Q - 1$, it is supposed that the correlation not explained by the previous levels is partially explained by the successive one. Formally, the additional layers are

$${}_{Q+2}\mathbf{R}_X = \mathbf{v}_{Q+2}c_{Q+2}^2\mathbf{v}'_{Q+2}, \dots, {}_{2Q-1}\mathbf{R}_X = \mathbf{v}_{2Q-1}c_{2Q-1}^2\mathbf{v}'_{2Q-1}, \tag{17}$$

where vectors \mathbf{v}_{Q+q} , $q = 1, \dots, Q - 1$, represent the broader groups obtained by the aggregation in pairs of those from the previous levels.

The UFA model parameterizes a correlation matrix of order J as the sum of $2Q - 1$ different layers as follows

$$\mathbf{R}_X = \sum_{q=0}^{Q-1} {}_{Q+q}\mathbf{R}_X + \Psi = \mathbf{V}_Q\mathbf{C}_Q^2\mathbf{V}'_Q + \sum_{q=1}^{Q-1} \mathbf{v}_{Q+q}c_{Q+q}^2\mathbf{v}'_{Q+q} + \Psi = \sum_{q=1}^{2Q-1} \mathbf{v}_q c_q^2 \mathbf{v}'_q + \Psi, \tag{18}$$

where the values c_q , $q = 1, \dots, Q$, in the diagonal of $\mathbf{C}_Q = \text{diag}(c_1, \dots, c_Q)$ represent the loadings between the q -th factor and the MVs in the q -th group of the initial partition in Q groups, while the values c_{Q+q} , $q = 1, \dots, Q - 1$, are the loadings of the higher-order factors. It is easy to see that the right-hand side of Eq. (18), i.e., the second and third term, is, by construction, a $(2Q - 1)$ -ultrametric correlation matrix with values

$$r_{jl} = \begin{cases} \sum_{q=1}^{2Q-1} v_{jq}c_q^2v_{lq} & \text{if } j \neq l \\ 1 & \text{if } j = l \end{cases}, \tag{19}$$

such that $\sum_{q=1}^{2Q-1} v_{jq}c_q^2v_{lq} \leq 1$ and $0 < c_q^2 \leq 1$. Additionally, \mathbf{R}_X in Eq. (18) can be rewritten in a compact matrix form as follows

$$\mathbf{R}_X = \ddot{\mathbf{V}}\ddot{\mathbf{C}}^2\ddot{\mathbf{V}}' + \Psi, \tag{20}$$

where

$$\ddot{\mathbf{V}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_Q, \mathbf{v}_{Q+1}, \dots, \mathbf{v}_{2Q-2}, \mathbf{1}_J] = [\mathbf{V}_Q, \mathbf{v}_{Q+1}, \dots, \mathbf{v}_{2Q-2}, \mathbf{1}_J] \tag{21}$$

is the $(J \times 2Q - 1)$ binary matrix, with the first Q columns of $\ddot{\mathbf{V}}$ reported in \mathbf{V}_Q which identify the partition of the J MVs in Q groups, while the following $Q - 1$ columns identify the aggregations of groups of MVs into partitions formed by $Q - 1, \dots, 1$, nested groups. The last column of $\ddot{\mathbf{V}}$ corresponds to the total aggregation of the MVs into a single class. The UFA loadings in each of the $2Q - 1$ layers are stored into the matrix $\ddot{\mathbf{C}}$, which is given by

$$\ddot{\mathbf{C}} = \text{diag}([c_1, \dots, c_Q, c_{Q+1}, \dots, c_{2Q-1}]) = [\mathbf{C}_Q, \text{diag}(c_{Q+1}, \dots, c_{2Q-2}, c_{2Q-1})]. \tag{22}$$

It is worth noting that the loadings c_{Q+1}, \dots, c_{2Q-1} produce *cross-loadings* w.r.t. $\mathbf{V}_Q \mathbf{C}_Q$, i.e., MVs that load on two or more higher-order LVs and that induce their correlation. However, while the number of parameters in a generic non-negative correlation matrix increases in the order of $O(J^2)$, for \mathbf{R}_X in Eq. (18), the number of parameters increases linearly in the matrix as we will see in Section 5.

Remark 1. The elements r_{jl} of the correlation matrix \mathbf{R}_X in Eq. (18) can be decomposed in two parts as follows

$$r_{jl} = \sum_{q=1}^Q v_{jq} c_q^2 v_{lq} + \sum_{q=Q+1}^{2Q-1} v_{jq} c_q^2 v_{lq}. \quad (23)$$

In (23), the first sum is equal to zero if MVs j and l belong to two different groups of \mathbf{V}_Q . It is worth noting that, in the first sum, at most one of the Q terms is non-zero (i.e., the term corresponding to the case where j and l belong to the same group), while in the second sum, which has $Q - 1$ terms, more than one term may be non-zero.

Remark 2. When \mathbf{R}_X in Eq. (18) has a block diagonal form as in Eq. (13), that is, the Q LVs are uncorrelated, all the correlation between MVs is due to the correlation among MVs within the Q groups, i.e., ${}_B r_{qh} = 0, q, h = 1, \dots, Q, h \neq q$. In this case, PDFa best explains the correlation in \mathbf{R}_X , and therefore it can be considered a special case of UFA where only c_1, \dots, c_Q are different from zero.

3.1. Model properties

In this section, we illustrate the fundamental properties of the UFA model, which consist of uniqueness, internal consistency and unidimensionality. It is thus worth recalling that a group of MVs is *internally consistent* when the corresponding factor explains them coherently, where coherence means that given any triplet j, l, p of MVs, if j is positively correlated (concordant) with l and p , then also the correlation between l and p is positive. Moreover, a group of (at least two) MVs is *unidimensional* if the first largest eigenvalue of their correlation matrix is greater than 1, and all the others are < 1 . This means that there exists a unique factor that explains most of their total variance.

Property 1 (Uniqueness). The loading matrix \mathbf{C} is unique, up to a permutation of columns of \mathbf{V}_Q .

Proof. This is a direct consequence of DFA, which has a unique solution for the loading matrix (Vichi, 2017, Property 2). Given \mathbf{V}_Q , \mathbf{R}_W and \mathbf{R}_B are unique and therefore also the loading matrix $\ddot{\mathbf{V}}\ddot{\mathbf{C}}$. Any transformation of $\ddot{\mathbf{C}}$ with an arbitrary orthogonal matrix \mathbf{T} , i.e., $\ddot{\mathbf{C}}^* = \ddot{\mathbf{C}}\mathbf{T}$, produces a non-admissible loading matrix for UFA. ■

Property 2 (Reliability). The matrices ${}_W\mathbf{R}_q$ in Eq. (14) define each one a factor which is generally reliable.

Proof. The Cronbach’s alpha of ${}_W\mathbf{R}_q$ is $\alpha_{{}_W\mathbf{R}_q} = \frac{n_q c_q^2}{1+(n_q-1)c_q^2}$, where for $c_q^2 \rightarrow 1, \alpha_{{}_W\mathbf{R}_q} \rightarrow 1$ (Spearman, 1910; Brown, 1910). It easily follows that if $c_q^2 = 1$, then $\alpha_{{}_W\mathbf{R}_q} = 1$. ■

An example of this property is as follows. Supposing that ${}_W\mathbf{R}_q$ has size $n_q = 5$ and the loading is $c_q = 0.7$, $\alpha_{{}_W\mathbf{R}_q} = 0.83$, which corresponds to a *good* internal consistency (reliability, coherence). For $n_q = 5$, but this time $c_q = 0.6$, $\alpha_{{}_W\mathbf{R}_q} = 0.73$, which is still an *acceptable* reliability.

Property 3 (Unidimensionality). The Q first-order factors identified by UFA are unidimensional. Indeed, the matrices ${}_W\mathbf{R}_q = (\mathbf{1}_{n_q}\mathbf{1}'_{n_q})c_q^2 + \Psi_q$ have the first largest eigenvalue equal to $(n_q - 1)c_q^2 > 1$, while all the remaining $(n_q - 1)$ eigenvalues have value $1 - c_q^2$, which is smaller than 1, since $0 < c_q^2 \leq 1$.

Proof. Denote with \mathbf{M} the $(n_q \times n_q)$ matrix of unitary elements $(\mathbf{1}_{n_q}\mathbf{1}'_{n_q})$. We have that \mathbf{M} is such that $\mathbf{M}\mathbf{M} = n_q\mathbf{M}$, hence $\mathbf{M} = \frac{\mathbf{M}\mathbf{M}}{n_q}$. Let λ be an eigenvalue of \mathbf{M} and \mathbf{v} the corresponding eigenvector, then we have that

$$\lambda\mathbf{v} = \mathbf{M}\mathbf{v} = \frac{\mathbf{M}\mathbf{M}}{n_q}\mathbf{v} = \frac{\mathbf{M}\lambda\mathbf{v}}{n_q} = \frac{\lambda^2\mathbf{v}}{n_q}. \tag{24}$$

Considering that $\mathbf{v} \neq 0$, $\lambda - \frac{\lambda^2}{n_q} = \lambda\left(1 - \frac{\lambda}{n_q}\right) = 0$. Therefore, either $\lambda = 0$ or $\left(1 - \frac{\lambda}{n_q}\right) = 0$, i.e. $\lambda = n_q$. Since, by definition, the trace of a square matrix is the sum of its diagonal elements, we have that $\text{tr}(\mathbf{M}) = n_q = \sum_{i=1}^{n_q} \lambda_i$, hence it must hold that only one $\lambda = n_q$ and all the other $n_q - 1$ eigenvalues must be equal to 0. Similarly, consider $\Psi_q = \text{diag}(\mathbf{1}_{n_q}(1 - c_q^2)) = (1 - c_q^2)\mathbf{I}_{n_q}$, then $\Psi_q\Psi_q = (1 - c_q^2)\Psi_q$ and therefore $\Psi_q = \frac{\Psi_q\Psi_q}{(1 - c_q^2)}$. Since $\lambda\mathbf{v} = \Psi_q\mathbf{v} = \frac{\Psi_q\Psi_q}{(1 - c_q^2)}\mathbf{v} = \frac{\Psi_q\lambda\mathbf{v}}{(1 - c_q^2)} = \frac{\lambda^2\mathbf{v}}{(1 - c_q^2)}$, we have that either $\lambda = 0$ or $\left(1 - \frac{\lambda}{(1 - c_q^2)}\right) = 0$, i.e. $\lambda = 1 - c_q^2$. Given that $\text{tr}(\Psi_q) = n_q(1 - c_q^2) = \sum_{i=1}^{n_q} \lambda_i$,

it must hold that $\lambda_i = 1 - c_q^2$ for all $i = 1, \dots, n_q$. As a consequence, the matrices $W\mathbf{R}_q$, $q = 1, \dots, Q$, are each one associated with a unidimensional factor, as they have only the largest eigenvalue greater than 1. ■

It is worth stressing that UFA’s structural constraints serve to limit the solution space and substantially reduce the likelihood of encountering improper solutions, such as negative variances or non-convergent estimates. Unlike traditional EFA, UFA imposes strict structural constraints on the factor loading matrix, ensuring a well-defined hierarchical and nested factor structure. These constraints effectively limit the solution space, preventing issues like Heywood cases or inflated factor loadings. As a result, UFA is less prone to the common estimation problems associated with unconstrained factor models, offering a more stable and reliable solution for analyzing data with hierarchical structures.

4. UFA Algorithm and Maximum Likelihood estimation

The ML estimation of UFA is performed by minimizing the following discrepancy function with respect to $\ddot{\mathbf{C}}$, $\ddot{\mathbf{V}}$, and Ψ

$$\begin{aligned}
 D(\ddot{\mathbf{C}}, \ddot{\mathbf{V}}, \Psi) &= \ln \left| \sum_{q=1}^{2Q-1} \mathbf{v}_q c_q^2 \mathbf{v}_q' + \Psi \right| - \ln |\mathbf{R}| + \text{Tr} \left(\left(\sum_{q=1}^{2Q-1} \mathbf{v}_q c_q^2 \mathbf{v}_q' + \Psi \right)^{-1} \mathbf{R} \right) - J \\
 &= \ln |\ddot{\mathbf{V}} \ddot{\mathbf{C}}^2 \ddot{\mathbf{V}}' + \Psi| - \ln |\mathbf{R}| + \text{Tr} \left((\ddot{\mathbf{V}} \ddot{\mathbf{C}}^2 \ddot{\mathbf{V}}' + \Psi)^{-1} \mathbf{R} \right) - J,
 \end{aligned}
 \tag{25}$$

such that

$$\ddot{\mathbf{V}} = \{v_{jq} : \forall v_{jq} \in \{0, 1\}\} \quad (\text{binary}) \tag{26}$$

$$\mathbf{V}_Q \mathbf{1}_Q = \mathbf{1}_J \quad (\text{row stochastic}) \tag{27}$$

$$\mathbf{v}_{Q+q} = \mathbf{v}_s + \mathbf{v}_t, s, t \in \{1, \dots, Q + q - 1\}, s \neq t, q = 1, \dots, Q - 1 \quad (\text{nestedness}) \tag{28}$$

$$\ddot{\mathbf{C}} = \text{diag}([c_1, \dots, c_Q, c_{Q+1}, \dots, c_{2Q-1}]), c_q > 0, q = 1, \dots, 2Q - 1 \quad (\text{diagonal, non-null}) \tag{29}$$

$$r_{jl} \geq \min\{r_{jh}, r_{lh}\}, j, l, h = 1, \dots, J, \text{ where Eq. (23) holds (ultrametric)}. \tag{30}$$

Given Q , the estimation of the UFA parameters is performed via a cycling block coordinate descent algorithm composed of the following four steps, which are sequentially repeated until a stopping rule is satisfied.

Step 0 [*Initialization*] The matrix $\hat{\mathbf{V}}_Q = [\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_Q]$ is randomly generated from a multinomial distribution with equal probabilities $1/Q$ and under the constraint that the Q groups are not empty. The initial values for $\hat{\mathbf{C}}_Q$ are obtained taking the square root of $\hat{\mathbf{R}}_W = [(\hat{\mathbf{V}}_Q' \hat{\mathbf{V}}_Q)^2 - \hat{\mathbf{V}}_Q' \hat{\mathbf{V}}_Q]^+ \text{diag}(\hat{\mathbf{V}}_Q' (\mathbf{R} - \mathbf{I}_J) \hat{\mathbf{V}}_Q)$, as in Cavicchia et al. (2020). Thus, matrices $\hat{\mathbf{V}}$ and $\hat{\mathbf{C}}$ are initialized as

$$\begin{aligned} \hat{\mathbf{V}} &= [\hat{\mathbf{V}}_Q, \hat{\mathbf{v}}_{Q+1}, \dots, \hat{\mathbf{v}}_{2Q-2}, \mathbf{1}_J], \text{ where } \hat{\mathbf{v}}_{Q+1}, \dots, \hat{\mathbf{v}}_{2Q-2} \text{ are } (J \times 1) \text{ vectors of zeros,} \\ \hat{\mathbf{C}} &= [\hat{\mathbf{C}}_Q, \text{diag}(\hat{c}_{Q+1}, \dots, \hat{c}_{2Q-2}, \hat{c}_{2Q-1})], \text{ where } \hat{c}_{Q+1}, \dots, \hat{c}_{2Q-1} \text{ are set equal to zero.} \end{aligned}$$

The initial $(J \times 2Q - 1)$ UFA loading matrix $\hat{\mathbf{A}}$ is computed as $\hat{\mathbf{V}} \hat{\mathbf{C}}$. Matrix $\hat{\mathbf{\Psi}}$ is obtained as:

$$\hat{\mathbf{\Psi}} = \mathbf{I}_J - \text{diag}(\hat{\mathbf{A}} \hat{\mathbf{A}}'). \tag{31}$$

Step 1 [*Update $\hat{\mathbf{V}}$*]

Step 1.1 [*Update \mathbf{V}_Q*] Matrix $\mathbf{V}_Q = [\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_J]'$, where \mathbf{v}_j denotes the generic row j of \mathbf{V}_Q , is updated row-by-row by assigning each MV j to the group q that most decreases the discrepancy in Eq. (25). Formally,

$$\begin{cases} \hat{v}_{jq} = 1 & \text{if } q = \underset{m=1, \dots, Q}{\text{argmin}} D(\hat{\mathbf{C}}, [\mathbf{v}_1, \dots, \mathbf{i}_m, \dots, \mathbf{v}_J]', [\hat{\mathbf{v}}_{Q+1}, \dots, \hat{\mathbf{v}}_{2Q-2}, \mathbf{1}_J]', \hat{\mathbf{\Psi}}) \\ \hat{v}_{jq} = 0 & \text{otherwise} \end{cases}, \tag{32}$$

where \mathbf{i}_m is the m -th row of the identity matrix of order Q .

Then, correlations are updated as in Cavicchia et al. (2020), i.e.,

$$\hat{\mathbf{R}}_B = (\hat{\mathbf{V}}_Q^+ \mathbf{R} \hat{\mathbf{V}}_Q'^+) \odot (\mathbf{1}_Q \mathbf{1}_Q' - \mathbf{I}_Q) \tag{33}$$

$$\hat{\mathbf{R}}_W = [(\hat{\mathbf{V}}_Q' \hat{\mathbf{V}}_Q)^2 - \hat{\mathbf{V}}_Q' \hat{\mathbf{V}}_Q]^+ \text{diag}(\hat{\mathbf{V}}_Q' (\mathbf{R} - \mathbf{I}_J) \hat{\mathbf{V}}_Q) \tag{34}$$

Step 1.2 [*Update last $Q - 2$ columns of $\hat{\mathbf{V}}$*] Given $\hat{\mathbf{R}}_B$, we update the $(Q + q)$ -th vector \mathbf{v}_{Q+q} , $q = 1, \dots, Q - 2$, by aggregating the two vectors \mathbf{v}_q and \mathbf{v}_h corresponding to the two factors with maximum between correlation ${}_{Br}r_{qh}$, $q, h = 1, \dots, Q, h \neq q$. This is the agglomerative part of the algorithm that considers the correlations between groups. Note that, if after

the q -th iteration columns q and h are aggregated, the labels of $\hat{\mathbf{R}}_B$ in positions q and h will be both replaced by $Q + q$. Thus, after each iteration, we move from Q groups of MVs to $Q - 1, \dots, 1$.

Step 2 [Update $\check{\mathbf{C}}$]

Step 2.1 [Update \mathbf{C}_Q] Given $\hat{\mathbf{V}}$ and $\hat{\mathbf{R}}_W$, the first Q values of matrix $\check{\mathbf{C}}$ are updated as

$$\hat{c}_q = \sqrt{W\hat{r}_{qq} - \sum_{s=1}^{Q-1} \frac{1}{n_q} \sum_{\substack{j,l=1 \\ j,l \in G_q}}^J \hat{v}_{js}\hat{c}_{Q+s}^2 \hat{v}_{ls}}, q = 1, \dots, Q.$$

Step 2.2 [Update the last $Q - 1$ columns of $\check{\mathbf{C}}$] Given $\hat{\mathbf{V}}$ and $\hat{\mathbf{R}}_B$, the $Q - 1$ higher order

levels are updated as $\hat{c}_{Q+q} = \sqrt{B\hat{r}_{qh} - \sum_{s=1}^{Q-1} \frac{1}{(n_q+n_h)} \sum_{\substack{j,l=1 \\ j \in G_q \\ l \in G_h}}^J \hat{v}_{js}\hat{c}_{Q+s}^2 \hat{v}_{ls}}, q, h = 1, \dots, Q, h \neq q.$

Step 3 [Update Ψ] Given $\hat{\mathbf{A}} = \hat{\mathbf{V}}\hat{\mathbf{C}}$, the discrepancy function $D(\hat{\mathbf{A}}, \Psi)$ in Eq. (25) is minimized with respect to Ψ by $\hat{\Psi} = \mathbf{I}_J - \text{diag}(\hat{\mathbf{A}}\hat{\mathbf{A}}')$.

The last three steps are alternated repeatedly, and with each iteration, the discrepancy function decreases, or at least does not increase. If the decrease of the discrepancy is larger than an arbitrarily small positive constant, the algorithm continues to iterate; otherwise, it stops and it is considered to have converged to a solution which is at least a local 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}_Q , retaining the best solution. The algorithm generally stops after a few iterations. In our simulation study reported in Section 6, this number is always less than 15, and it can be observed that the algorithm accurately identifies the hierarchical structure in factors of MVs generated according to the UFA model.

Remark 3. $\check{\mathbf{V}}$ defines a set of subgroups of MVs, described by their binary vectors, such that \mathbf{v}_l is the binary vector identifying the non-empty l -th subgroup of $V = \{1, 2, \dots, J\}$, and if $\mathbf{v}_j, \mathbf{v}_l \in \check{\mathbf{V}} \Rightarrow (\mathbf{v}_j \cap \mathbf{v}_l) \in \{\mathbf{v}_j, \mathbf{v}_l, \emptyset\}$. This structure is a J -tree formed by J leaves corresponding the MVs $\{j\}$, with $j \in V$. Then, the first Q groups of MVs, corresponding to the first Q internal nodes of the tree, define a partition of V in Q disjoint groups. The successive $Q - 1$ steps specify the agglomeration of the Q groups in $Q - 1$ steps of group-fusions up to the root of the tree.

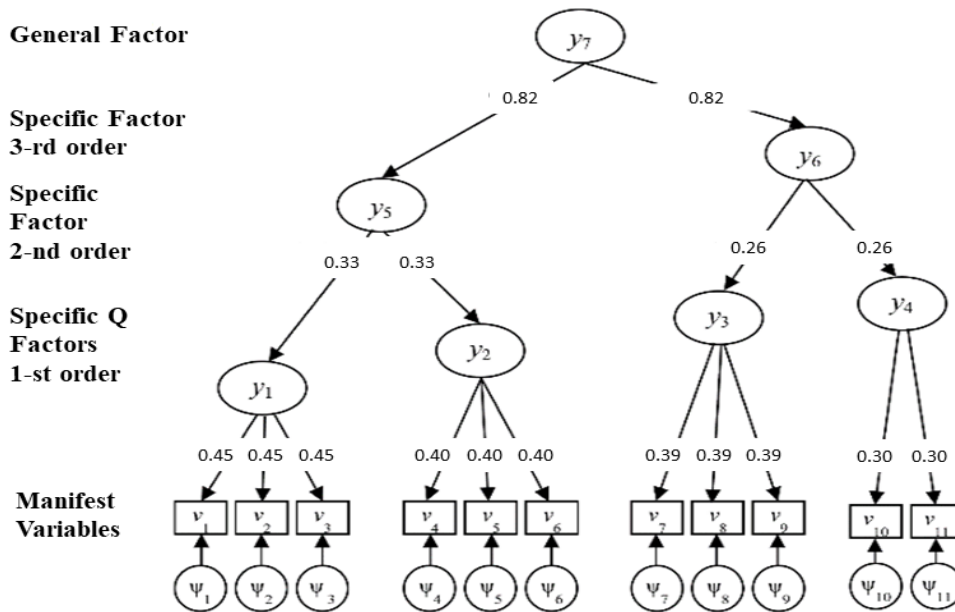


Figure 2: Path diagram for the given example.

Remark 4. In general, there is no reason why the optimal estimation of UFA should include the PDFFA solution at the Q -th level. However, if one wants to improve the algorithm’s computational efficiency, the estimates in **Step 1.1** and **Step 2.1** can be replaced by the PDFFA solution at convergence. This avoids computing Eqs. (32), (33) and (34) at every iteration. Therefore, we define a new constrained algorithm for UFA that is forced to include the optimal solution of PDFFA for the given Q .

The following example is provided to help the reader better follow the UFA modeling.

Example. We apply UFA to the set of eleven MVs in the example of Fig. 1. In the correlation matrix in Fig. 1a, $Q = 4$ groups of highly correlated variables are clearly visible: $G_1 = \{V_1, V_2, V_3\}$, $G_2 = \{V_4, V_5, V_6\}$, $G_3 = \{V_7, V_8, V_9\}$, $G_4 = \{V_{10}, V_{11}\}$. The specific Q first-order factors are characterized by loadings $\hat{c}_1 = 0.45$, $\hat{c}_2 = 0.40$, $\hat{c}_3 = 0.39$ and $\hat{c}_4 = 0.30$, as reported in Fig 2. The loadings are computed considering the within correlations in Fig. 1a as reported in **Step 2.1** of the algorithm and detailed in *Remark 1*. The membership matrix for this partition is identified by the first four columns $\hat{v}_1, \hat{v}_2, \hat{v}_3$ and \hat{v}_4 of matrix $\hat{\mathbf{V}}$. The within and between correlations of groups/factors are given by

$$\hat{\mathbf{R}}_W = \begin{bmatrix} 0.98 & 0 & 0 & 0 \\ 0 & 0.94 & 0 & 0 \\ 0 & 0 & 0.89 & 0 \\ 0 & 0 & 0 & 0.83 \end{bmatrix} \quad \hat{\mathbf{R}}_B = \begin{bmatrix} 0 & 0.78 & 0.67 & 0.67 \\ 0.78 & 0 & 0.67 & 0.67 \\ 0.67 & 0.67 & 0 & 0.74 \\ 0.67 & 0.67 & 0.74 & 0 \end{bmatrix}$$

Let us now consider the hierarchical structure of factors – from the most specific ones (first-order) to the most general, through second- and third-order SFs – as shown in the higher layers of the path diagram in Fig. 2. The second-order SF (2nd layer) agglomerates the first two groups of MVs since their correlation is the highest in $\hat{\mathbf{R}}_B$ (equal to 0.78). Therefore, $G_1 \cup G_2 = G_5 = \{V_1, V_2, V_3, V_4, V_5, V_6\}$, where this aggregation is identified by the vector $\hat{\mathbf{v}}'_5 = \hat{\mathbf{v}}'_1 + \hat{\mathbf{v}}'_2 = [1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0]$ representing the fifth column of matrix $\hat{\mathbf{V}}$. The corresponding loading is $\hat{c}_5 = 0.33$. Having merged these two groups, the third-order SF (3-rd layer) lumps together groups G_3 and G_4 (the correlation equals 0.74), i.e., $G_3 \cup G_4 = G_6 = \{V_7, V_8, V_9, V_{10}, V_{11}\}$, obtaining the additional vector $\hat{\mathbf{v}}'_6 = \hat{\mathbf{v}}'_3 + \hat{\mathbf{v}}'_4 = [0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1]$ and $\hat{c}_6 = 0.26$. The last aggregation corresponds to $G_5 \cup G_6 = G_7 = \{V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}, V_{11}\}$, where $\hat{\mathbf{v}}_7$ is the unitary vector of dimension eleven and $\hat{c}_7 = 0.82$. All higher-order loadings can be found by considering the between correlations in $\hat{\mathbf{R}}_B$, as reported in **Step 2.2**. As a result, the UFA membership matrix and the corresponding loading matrix are given by

$$\hat{\mathbf{V}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad \hat{\mathbf{A}} = \hat{\mathbf{V}}\hat{\mathbf{C}} = \begin{bmatrix} 0.45 & 0 & 0 & 0 & 0.33 & 0 & 0.82 \\ 0.45 & 0 & 0 & 0 & 0.33 & 0 & 0.82 \\ 0.45 & 0 & 0 & 0 & 0.33 & 0 & 0.82 \\ 0 & 0.40 & 0 & 0 & 0.33 & 0 & 0.82 \\ 0 & 0.40 & 0 & 0 & 0.33 & 0 & 0.82 \\ 0 & 0.40 & 0 & 0 & 0 & 0.26 & 0.82 \\ 0 & 0 & 0.39 & 0 & 0 & 0.26 & 0.82 \\ 0 & 0 & 0.39 & 0 & 0 & 0.26 & 0.82 \\ 0 & 0 & 0.39 & 0 & 0 & 0.26 & 0.82 \\ 0 & 0 & 0 & 0.30 & 0 & 0.26 & 0.82 \\ 0 & 0 & 0 & 0.30 & 0 & 0.26 & 0.82 \end{bmatrix},$$

where the first Q columns of $\hat{\mathbf{V}}$ correspond to $\hat{\mathbf{V}}_Q$. Finally, $\hat{\mathbf{\Psi}} = \text{diag}([0.02, 0.02, 0.02, 0.06, 0.06, 0.06, 0.11, 0.11, 0.11, 0.17, 0.17])$.

5. Model selection and choice of Q

In this section, we discuss the choice of the optimal UFA model, among a set of candidates that consider different values of Q . The optimal model is the one that strikes the best balance between complexity and ability to accurately fit the observed data. Through this careful selection, we determine the appropriate number of factors Q to retain, ensuring both efficiency and effectiveness in our analysis. To assess the complexity of the model, we start from noticing that the population correlation matrix has $\frac{J(J-1)}{2}$ parameters to be estimated (the correlations on the diagonal are always equal to 1) in terms of $J(2Q-1) + J$ unknown parameters in $\ddot{\mathbf{A}}$ and Ψ . However, note that the first Q columns of $\ddot{\mathbf{A}}$ have only one loading different from 0 for each row. Thus, we have $J(Q-1)$ constraints in $\ddot{\mathbf{A}}$. Moreover, we have other Q constraints on $\ddot{\mathbf{A}}$ to ensure that each of the first Q columns of $\ddot{\mathbf{A}}$ is not empty. Finally, we suppose that all MVs loading on the same factor have the same coefficient ($\ddot{\mathbf{A}} = \ddot{\mathbf{V}}\ddot{\mathbf{C}}$), therefore we do not have to estimate J coefficients but only Q ; this corresponds to have $J-Q$ additional constraints on $\ddot{\mathbf{A}}$. This latter condition also holds for the last $Q-1$ columns of $\ddot{\mathbf{A}}$: we do not have to estimate $J(Q-1)$ coefficients but only $Q-1$ and thus we have additional $J(Q-1) - (Q-1)$ constraints on $\ddot{\mathbf{A}}$. Since we have the same coefficient for all MVs loading onto the same factor, also for the diagonal of Ψ we do not have to estimate J parameters but Q , and therefore there are $J-Q$ constraints on Ψ . Hence, the effective number of unknown parameters in UFA is $p_{UFA} = J(2Q-1) + J - J(Q-1) - (J-Q) - Q - [J(Q-1) - (Q-1)] - (J-Q) = 2Q-1$. Note that, due to the factorization of $\ddot{\mathbf{A}}$, p_{UFA} does not depend on J . The degrees of freedom of the model are therefore $df = \frac{J(J-1)}{2} - 2Q + 1$.

Two popular model selection criteria are the Akaike Information Criterion (AIC, Akaike, 1998) and the Bayesian Information Criterion (BIC, Schwarz, 1978), which can be easily estimated for the UFA model, given Eq. (25) and p_{UFA} . Among a set of candidate models (i.e., different values of Q), the preferred is the one for which AIC and BIC are minima. Alternatively, a goodness of fit index frequently used to address this task is the Adjusted Goodness of Fit Index (AGFI, Mulaik et al., 1989), which is computed as follows

$$\text{AGFI} = 1 - \frac{J(J-1)}{2df}(1 - \text{GFI}), \quad (35)$$

where GFI is the Goodness of Fit Index (GFI, Jöreskog, 1981) obtained as $\text{GFI} = 1 - \frac{\text{tr}[(\mathbf{R}_X^{-1}\mathbf{R} - \mathbf{I}_J)^2]}{\text{tr}[(\mathbf{R}_X^{-1}\mathbf{R})^2]}$. This selection criterion chooses Q such that AGFI is maximized.

It is worth stressing that even if these criteria are well-recognized and commonly used, AIC and BIC can be more helpful for model comparison (i.e., selecting the optimal Q), while AGFI is more appropriate for evaluating model fit once the number of first-order factors Q has been chosen. Furthermore, in practice, combining multiple methods – both statistical and conceptual – typically leads to the most robust factor solution.

Remark 5. Note that when performing non-parsimonious DFA, the number of parameters to be estimated is $p_{DFA} = 2J - Q$ (Vichi, 2017), whereas for PDFA, this number is $p_{PDFA} = 2Q - Q = Q$. It can be immediately seen that p_{PDFA} corresponds to UFA without considering the $Q - 1$ parameters of the factorial hierarchy. In fact, $p_{PDFA} = p_{UFA} - (Q - 1) = 2Q - 1 - (Q - 1) = Q$.

6. Simulation study

The proposed model is evaluated through an extended simulation study that considers twelve different scenarios and involves 3,000 data samples, generated according to the following procedure. First, the matrix \mathbf{V}_Q is randomly generated. Then, a random binary J -tree is generated starting from \mathbf{V}_Q and including $Q - 1$ random agglomerations of subsets (couples). The matrix $\ddot{\mathbf{V}}$ is identified by this J -tree. As a second step, a diagonal matrix $\ddot{\mathbf{C}}$ with non-increasing values in $[0, 1]$ is generated, and the loading matrix $\ddot{\mathbf{A}}$ is obtained as $\ddot{\mathbf{A}} = \ddot{\mathbf{V}}\ddot{\mathbf{C}}$. The matrix Ψ is then computed accordingly using Eq. (31). The correlation matrix is obtained as $\mathbf{R}_X = \ddot{\mathbf{A}}\ddot{\mathbf{A}}' + \Psi$, and it is $(2Q - 1)$ -ultrametric by construction. Finally, the observed correlation matrix \mathbf{R} is obtained adding to \mathbf{R}_X a uniform error \mathbf{E}_R in $[0, \sigma_e]$, which is symmetrized. The psd of \mathbf{R} has been verified: if this property is not satisfied, we add the absolute value of the smallest eigenvalue of \mathbf{R} to its main diagonal, so that the resulting matrix is positive semi-definite (Cailliez, 1983).

Twelve different scenarios have been taken into account to test the proposed model. We considered $J = 30, 70$ MVs, $Q = 4, 7$ factors and three error levels $\sigma_e = 0.01$, $\sigma_e = 0.5$ and $\sigma_e = 1$. In Fig. 3, examples of simulated correlation matrices for each scenario are illustrated. The groups and their hierarchical structure are clearly visible in the first row of Fig. 3. In the four plots of the

second row, corresponding to the medium error level, the hierarchical structure of the Q first-order factors is still quite visible while the one for higher-order aggregations tends to disappear. Finally, for the high error case (third row of Fig. 3), the whole hierarchical structure among both MVs and LVs almost disappears.

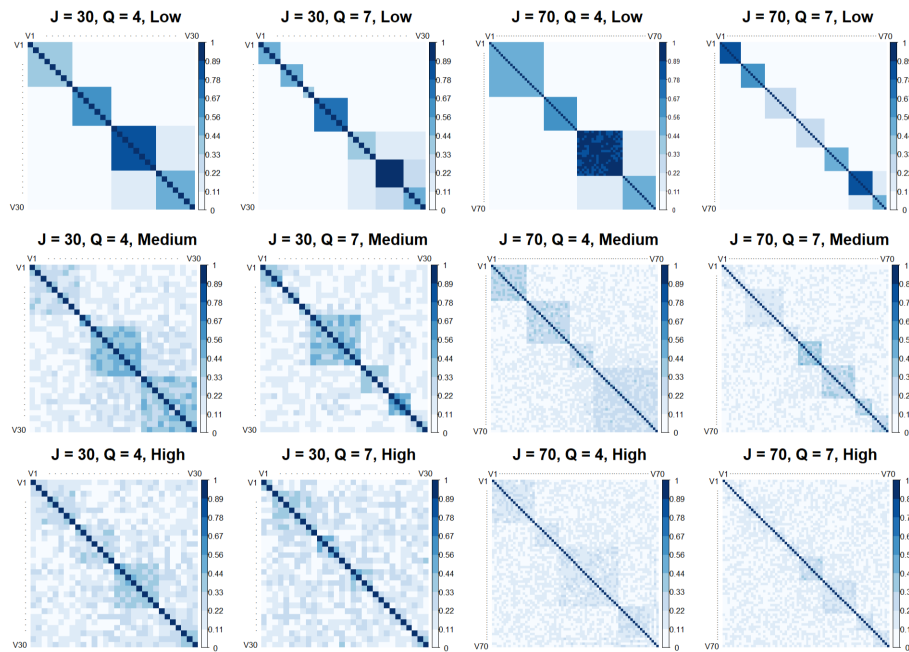


Figure 3: Examples of simulated correlation matrices with different levels of error, J and Q .

6.1. Random starts assessment

UFA does not guarantee the identification of the global optimal solution, which is expected since the partitioning problem is known to be NP-hard (Křivánek and Morávek, 1986). To increase the chance of find the global optimum, the *multistart* technique is used and in this subsection it is discussed how the number of random starts is chosen. To address this, the UFA algorithm is run 250 times with a high error level and $J = 70, Q = 7$, using different numbers of *randstarts* values: [1, 5, 10, 20, 30, 50, 100]. For each case, the percentage of local minima is recorded.

From the results in Tab. 1, the number of random starts for the simulation study is set equal to 50.

Table 1: Local minima occurrences (%).

<i>randstarts</i>	1	5	10	20	30	50	100
% Local minima	80.40	36.80	16.80	3.60	1.60	0.00	0.00

6.2. Model performance and comparison with competing procedures

The proposed model is evaluated according to the Adjusted Rand Index (ARI, Hubert and Arabie, 1985), which compares the generated J -tree with the estimated one. First, the ARI between the true and estimated matrices \mathbf{V}_Q for the first Q aggregations is computed. Then, an ARI is computed for each of the remaining $Q - 1$ aggregations levels. For each $q = 1, \dots, Q - 1$, the ARI is computed between the two factors aggregated by the algorithm at level $Q + q$ and the two generated ones. As a result, the generated J -tree is optimally recovered if the overall ARI equals to Q , since it is obtained as the sum of one ARI for the first Q levels and $Q - 1$ ARIs for the factors' hierarchy. Finally, this value is rescaled, dividing it by Q , to range in $[0, 1]$. We will denote with ARI_Q the ARI referring to the first Q levels and with ARI_J the overall ARI for the J -tree. Furthermore, the fit of the $(2Q - 1)$ -ultrametric correlation matrix resulting from UFA to the generated \mathbf{R} is evaluated through the AGFI. The discrepancy between the generated loading matrix, denoted with $\mathbf{A}^* = \mathbf{\hat{V}}^* \mathbf{\hat{C}}^*$ and the estimated one $\hat{\mathbf{A}} = \hat{\mathbf{V}} \hat{\mathbf{C}}$ is evaluated considering the Root Mean Squared Error (RMSE) defined as $\frac{\|\hat{\mathbf{A}} \hat{\mathbf{A}}' - \mathbf{A}^* \mathbf{A}^{*'}\|^2}{JQ}$ since the loading matrix is not unique.

UFA performance is compared with the ones of several well-established factor rotation methods, such as geomin (geo), promax (pro), and quartimax (qua), as well as with the *prenet* penalization method for factor analysis proposed by Hirose and Terada (2023). Geomin, introduced by (Yates, 1987), minimizes the complexity of factor loadings by reducing the number of near-zero values. Promax, a widely used oblique rotation method proposed by Hendrickson and White (1964), allows factors to correlate, thereby enhancing the interpretability of the factor solution. Quartimax, introduced by Neuhaus and Wrigley (1954), seeks to simplify the rows of the factor loading matrix by maximizing the variance of squared loadings across variables. The Sparse and Simple Structure (SSS) estimation method developed by Hirose and Terada (2023) employs the *prenet* penalization to achieve a factor solution that is both sparse and interpretable. This estimation requires tuning two

Table 2: Simulation study results.

	J	Low error				Medium error				High error			
		30		70		30		70		30		70	
		4	7	4	7	4	7	4	7	4	7	4	7
Mean ARI_Q	Q	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00	0.97	0.72	1.00	0.93
	SSS	0.66	0.78	0.65	0.77	0.63	0.68	0.65	0.72	0.41	0.19	0.65	0.72
	pro	0.66	0.78	0.65	0.77	0.66	0.78	0.65	0.77	0.66	0.77	0.66	0.77
	geo	0.66	0.78	0.65	0.77	0.66	0.77	0.65	0.77	0.65	0.76	0.66	0.76
	qua	0.66	0.78	0.65	0.77	0.66	0.78	0.65	0.77	0.65	0.77	0.66	0.76
% $ARI_Q = 1$	UFA	100.00	100.00	100.00	100.00	99.60	86.80	100.00	100.00	79.60	3.20	99.60	34.00
	SSS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	pro	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	geo	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	qua	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mean ARI_J	UFA	1.00	0.98	1.00	0.98	1.00	0.85	1.00	0.93	0.96	0.70	1.00	0.83
% $ARI_J = 1$	UFA	100.00	86.00	100.00	86.40	97.20	4.80	100.00	38.40	62.40	0.00	98.80	3.60
AGFI	UFA	1.00	1.00	0.99	1.00	0.78	0.78	0.79	0.79	0.77	0.77	0.78	0.79
	SSS	0.86	0.91	0.83	0.89	0.51	0.54	0.69	0.69	0.43	0.46	0.67	0.62
	pro	0.98	0.97	0.99	0.98	0.64	0.45	0.74	0.71	0.63	0.44	0.74	0.70
	geo	0.98	0.97	0.99	0.98	0.65	0.51	0.75	0.72	0.65	0.52	0.74	0.73
	qua	0.98	0.97	0.99	0.98	0.65	0.52	0.75	0.72	0.65	0.53	0.74	0.73
RMSE \mathbf{AA}'	UFA	0.00	0.00	0.00	0.00	0.16	0.08	0.60	0.19	0.28	0.14	0.88	0.30
	SSS	1.22	0.44	2.72	0.97	1.08	0.39	2.31	0.81	1.48	0.40	7.86	0.82
	pro	1.26	0.44	2.95	1.03	1.03	0.40	2.23	0.79	1.03	0.37	2.24	0.79
	geo	1.25	0.44	2.93	1.02	0.98	0.36	2.18	0.77	0.98	0.34	2.16	0.77
	qua	1.26	0.44	2.95	1.03	0.98	0.36	2.20	0.77	0.98	0.33	2.17	0.77

parameters, denoted by ρ and γ , and has been carried out to maximize AGFI. By comparing these techniques with UFA, which offers an innovative approach to factor analysis, we aim to highlight its strengths and potential advantages in achieving more interpretable and accurate factor solutions when data exhibit a hierarchical structure.

Simulation results, obtained considering 50 random starts for all cases, are shown in Tab. 2. For each scenario, $nsim = 250$ data matrices have been generated and the reported performance measures correspond to the average over the 250 samples.

Under the low error condition, our method demonstrates best performance across all evaluated metrics. Specifically, the generated J -tree is always perfectly identified for $Q = 4$, while the percentage of times this happens is around 86% for $Q = 7$ (see Tab. 2). Moreover, the AGFI is always greater than or equal to 0.99 and the RMSE for the loadings reaches 0. In contrast, the competing procedures poorly perform in recovering the variables structure in Q groups: the ARI_Q always ranges between 0.65 and 0.78 and it is never equal to 1. Additionally, the high values for the RMSE of the loading matrix indicate a sub-optimal recovery of the generated loadings across all

scenarios. This first part of the simulation study indicates that the proposed methodology and the given algorithm do not fail to identify the correct generated ultrametric correlation matrix when its ultrametricity is clearly visible, as shown in the first row plots of Fig. 3. In comparison, the competing procedures fail to achieve an accurate identification under the same conditions.

For the medium error level scenarios, the UFA algorithm achieves best performance in terms of ARI_Q , while the recovery of the higher-order hierarchies starts worsening. Indeed, in the plots of Fig. 3 corresponding to the medium error level, the structure in Q factors is still visible while the higher-order correlations start deviating from ultrametricity. For instance, for $J = 70$ and $Q = 7$, the $\% ARI_Q = 1$ and the $\% ARI_J = 1$ are 100.00 and 38.40, respectively. Also the model-data fit in terms of RMSE and AGFI worsens, but the fit is still better than the competing procedures.

Under the high error condition, as expected, the UFA performance also in terms of ARI_J worsens because the ultrametricity tends to be masked by the error, as can be seen in the last row of Fig. 3. The true J -tree is never identified for $J = 30$ and $Q = 7$ (under the same scenario, the $\% ARI_Q = 1$ is 3.20). However, also under these scenarios, UFA outperforms the competing procedures, especially in terms of ARI_Q and RMSE.

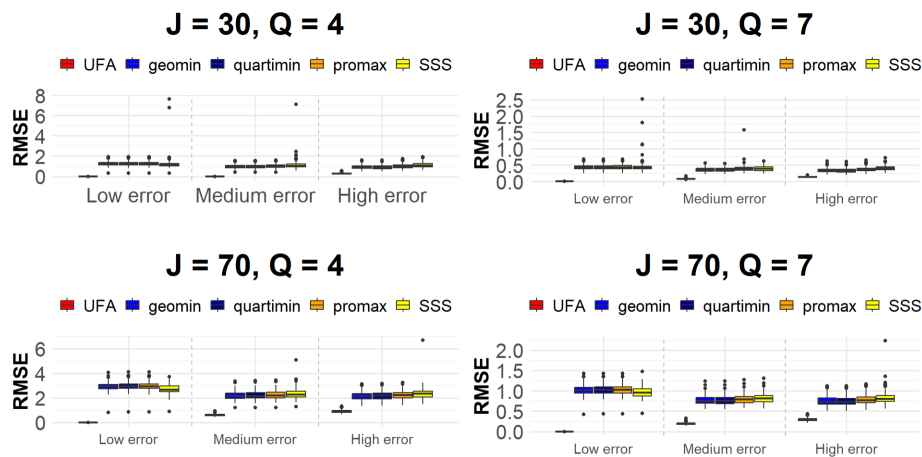


Figure 4: RMSE of factor loadings for UFA and the competing procedures.

As illustrated in Fig. 4, UFA shows the lowest RMSE across all scenarios, with its interquartile range being very tight, indicating both a low error and consistent performance across simulations. In contrast, the competing methods exhibit higher median RMSE values and wider interquartile

ranges, indicating poorer performance and higher variability. These results underscore the efficacy and accuracy of UFA in capturing hierarchical relationships in data, making it a reliable choice for this kind of scenarios, compared to traditional methods.

To summarize, simulation results collectively suggest that the proposed algorithm not only achieves high accuracy in identifying the hierarchical structure of the generated data and finds its best ultrametric reconstruction, but also exhibits stability and reliability in its performance. The effectiveness of the algorithm against the benchmark criteria and competing procedures suggests its potential utility in applications in which phenomena that can be modeled by a hierarchical latent structures are studied.

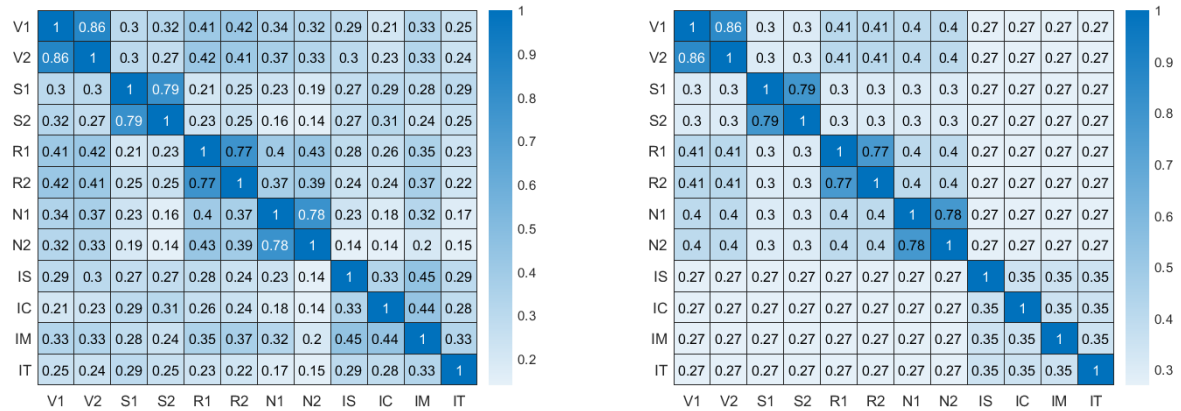
7. Application

The proposed model is applied to the data employed by Cattell (1971) for its factorial study on intelligence, in which the author investigates his theory of crystallized and fluid intelligence factors. Crystallized intelligence involves the ability to use learned knowledge and experience. It involves understanding facts, solving problems based on past experiences, and possessing a rich vocabulary, and it tends to improve with age. On the other hand, fluid intelligence represents the ability to solve novel problems, independent of any knowledge from the past. It includes skills like pattern recognition, abstract reasoning, and problem-solving in new situations. Fluid intelligence is thought to peak in early adulthood and then gradually decline. Cattell's study used data from 277 8-th grade children, focusing on their abilities across different dimensions. These dimensions were measured through various tests:

- **Thurstone Primaries:** four areas (Verbal, Spatial, Reasoning, Number) were each measured by two different tests, resulting in eight measures (V1, V2, S1, S2, R1, R2, N1, N2). These areas represent different cognitive abilities as identified by Thurstone (1938), another key figure in the study of human intelligence.
- **Culture Fair Intelligence Tests:** developed by the Institute of Personality and Ability Testing (IPAT), these tests aim to measure intelligence in a way that minimizes cultural and educational biases. The tests include Series, Classification, Matrices, and Topology, each providing a score

(IS, IC, IM, IT).

The non-negative correlations among the 12 considered MVs are shown in Fig. 5a. These high correlations suggest the need to explore how the various aspects of intelligence interrelate and potentially validate or refine Cattell’s theory of crystallized and fluid intelligence.



(a) Observed correlation matrix.

(b) Ultrametric correlation matrix estimated by UFA.

Figure 5: Observed and estimated correlation matrices of Cattell’s data.

The UFA algorithm has been run considering 50 random starts and it has found the expected partition in 5 first-order factors (Verbal, Spatial, Reasoning, Number and IPAT). It is worth noting that $Q = 5$ is the best choice for the number of parameters also according to the AGFI, AIC, and BIC. The unidimensionality of the identified latent dimensions is ensured by the variance of the second factor equal to 0.14, 0.21, 0.23, 0.22 and 0.64, respectively. The Cronbach’s alpha values computed for each factor align with Cattell’s hypothesis: they are greater than 0.95 for F1, F2, F3, F4, while equal to 0.46 for F5. As it is possible to observe comparing the two matrices in Fig. 5, the proposed model is able to well reconstruct the correlation matrix of the MVs, simultaneously identifying its hierarchical structure. The latter is depicted in Fig. 6a: it can be noticed that the second last aggregation (F5-F8) characterizes the fluid ability and crystallized ability factors hypothesized by Cattell.

Moreover, the last factor aggregated to build F8 is the one related to the Spatial ability, the

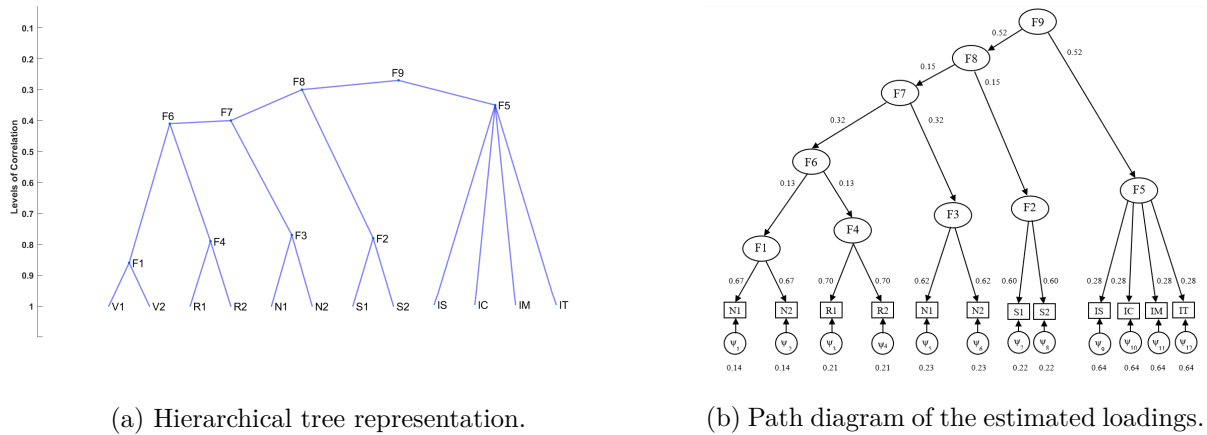


Figure 6: Hierarchical tree and loading structure estimated by UFA on Cattell's data.

one for which Cattell hypothesized the strongest relationship with the IPAT MVs (F5). This is confirmed by the lowest Cronbach's alpha value for F8 (0.37) compared to the values for F6 and F7 (0.77 and 0.75, respectively); indeed, Cattell hypothesized an high correlation between Verbal, Reason and Number. A general concept defining Mental Abilities is also defined (F9), its reliability is ensured by a Cronbach's alpha value equal to 0.85. The estimated loadings are graphically shown in Fig. 6b.

Note that the estimated matrix represented in Fig. 5b is ultrametric even if the within deviance of the IPAT group (0.353) is higher than, for instance, the between deviance of the Verbal and Spatial groups (0.415). In fact, the ultrametric condition defined in Subsection 2.1 imposes that the within deviance of the IPAT group must be greater than or equal to the deviance between IPAT and the other four groups (i.e. $0.353 \geq 0.275$) and not greater than or equal to any of the between deviances in the correlation matrix.

UFA results on Cattell data represent a great contribution to the state of the art in Cattell's factorial study on intelligence. UFA does a step forward with respect to both First Order and Second Order Confirmatory Factor Analyses (CFA) that Rindskopf and Rose (1988) fitted to these data. It recognizes the five factors despite the fully exploratory approach and it also accounts for intercorrelations among them, going beyond the construction of just two second-order factors (with a correlation of 0.78). UFA model addresses the unexplained correlation among factors building

Table 3: AIC and BIC values for UFA model and for the two CFAs applied to Cattell data.

Models	AIC	BIC
First Order CFA	7890.62	8013.84
Second Order CFA	7887.01	7995.73
Ultrametric Factor Analysis	2289.31	2340.04

a hierarchy of Specific Latent Variables. UFA model has been compared with the two CFAs in terms of AIC and BIC (Tab. 3). As it is possible to observe, the lowest values correspond to UFA, reflecting the better modeling of the underlying factors' correlation that characterize Cattell's data. The goodness of the fit of the estimated UFA model is further confirmed by an AGFI equal to 0.95.

The results obtained in this application suggest that an UFA can be the proper choice in all psychometric applications in which the researcher is interested in evaluating an hypothesized hierarchy of latent concepts and seeks to define and characterize broader dimensions of the phenomenon under study.

8. Conclusions

In the realm of psychometric studies, hierarchical concepts are pivotal for uncovering and comprehending fundamental phenomena, such as intelligence. In the literature, hierarchical models based on sequential procedures of EFA followed by oblique rotations have been proposed to arrange various facets or subcomponents of a construct into a hierarchical structure, showcasing their interconnectedness and offering an in-depth comprehension of complex phenomena. Additional models widely employed in psychometric studies to assess the compatibility of theoretical hierarchical factorial structures with observed data are CFA and SEM. These statistical tools help confirm that the used models accurately reflect underlying constructs, offering a reliable framework for analyzing intricate data. These models are widely used across various psychometric domains, with the Five-Factor Model of personality serving as just one example. This model categorizes personality traits into five overarching factors, each subdivided into lower-level facets that depict specific aspects of these traits.

In this paper, we introduce a novel methodology called Ultrametric Factor Analysis. This approach expands upon the traditional EFA model with the aim of identifying hierarchical structures

within factors that represent latent concepts through an exploratory and simultaneous approach, as opposed to confirmatory or sequential procedures based on EFA and oblique rotations. UFA proves particularly valuable in studying complex phenomena that exhibit multiple levels of abstraction and interconnected dimensions. The proposal can be fully exploratory, meaning that no prior assumptions are made regarding the relationships among MVs, first-order factors, and higher-order factors. However, it also offers the flexibility to fix some or all of these relationships in a partially or fully confirmatory approach. UFA involves a mathematical formalization of the hierarchical relationships among variables, which enables a synthesized representation of latent concepts associated with complex and multidimensional phenomena. By extending EFA to hierarchical structures, this methodology empowers researchers to delve into the underlying structure of complex phenomena with several facets by uncovering factors and their hierarchical relationships in a simultaneous approach.

Furthermore, the new methodology addresses several crucial properties essential for constructing a hierarchy of factors. These properties include internal consistency, reliability, unidimensionality, uniqueness, and content validity. Ensuring that these properties are met guarantees that the constructed hierarchical model accurately represents the underlying constructs and provide a comprehensive framework for analyzing complex data. Simulation study results showcased the superiority of the UFA method over traditional factor analysis techniques, especially when dealing with data that have a nested or hierarchical factor structure. While conventional methods often fail to effectively capture the complexity of such data, leading to sub-optimal factor solutions, UFA excels in identifying and modeling these intricate relationships, resulting in more accurate and interpretable factor structures. In summary, the UFA methodology presents a novel approach to investigate hierarchical concepts in psychometric studies and other disciplines where latent factors need to be constructed. This methodology offers insights into the multidimensional nature of latent constructs and facilitates a comprehensive understanding of complex phenomena. It enhances our ability to analyze data and extract meaningful information from hierarchical structures, contributing to advancements in various research fields.

Further developments of UFA include relaxing the non-negativity assumption on the correlation matrix, which, while realistic in many applications – especially in psychometric studies – could be

too restrictive in other fields. Additionally, when the phenomenon under study does not show a hierarchical structure, the ultrametricity assumption could also be restrictive, and an hypothesis test to determine how far the data are from this condition could be useful for discriminating the applicability of the proposal.

Declarations

Conflict of interest. The authors did not receive support from any organization for the submitted work, and the authors have no financial or non-financial conflict of interest to disclose.

Code availability. The MATLAB source code of UFA is openly available at <https://github.com/AuthorsSubmission/UFA>.

References

- Akaike, H. (1998). *Information Theory and an Extension of the Maximum Likelihood Principle*. Springer.
- Bollen, K. A. and Bauldry, S. (2011). Three cs in measurement models: causal indicators, composite indicators, and covariates. *Psychological methods*, 16(3):265–284.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3(3):296–322.
- Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, 48(2):305–308.
- Campbell, J. P. (1990). *Modeling the performance prediction problem in industrial and organizational psychology*. Consulting Psychologists Press.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Houghton Mifflin.
- Cavicchia, C. and Vichi, M. (2022). Second-order disjoint factor analysis. *Psychometrika*, 87(1):289–309.

- Cavicchia, C., Vichi, M., and Zaccaria, G. (2020). The ultrametric correlation matrix for modelling hierarchical latent concepts. *Advances in Data Analysis and Classification*, 14:837–853.
- Cavicchia, C., Vichi, M., and Zaccaria, G. (2023). Hierarchical disjoint principal component analysis. *AStA Advances in Statistical Analysis*, 107:537–574.
- Costa, P. T. and McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Psychological Assessment Resources.
- Cronbach, L. J. (1990). An alternative "description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, 59(6):1216–1229.
- DeYoung, C. G., Peterson, J. B., and Higgins, D. M. (2002). Higher-order factors of the big five predict conformity: Are there neuroses of health? *Personality and Individual Differences*, 33(4):533–552.
- European Commission (2017). *European statistics code of practice*. Luxembourg: Publications Office of the European Union.
- Ferrara, C., Martella, F., and Vichi, M. (2016). Dimensions of well-being and their statistical measurements. In Alleva, G. and Giommi, A., editors, *Topics in Theoretical and Applied Statistics*, pages 85–99. Springer International Publishing.
- Gorsuch, R. L. (1983). *Factor Analysis*. Psychology Press.
- Hendrickson, A. E. and White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology*, 17(1):65–70.
- Hirose, K. and Terada, Y. (2023). Sparse and simple structure estimation via prenet penalization. *Psychometrika*, 88(4):1381–1406.
- Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218.
- Jöreskog, K. G. (1981). *LISREL V: analysis of linear structural relationships by maximum likelihood and least squares methods*. Uppsala, Chicago: University of Uppsala, Dept. of Statistics; distributed by International Educational Services.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23:187–200.
- Křivánek, M. and Morávek, J. (1986). Np-hard problems in hierarchical-tree clustering. *Acta*

- informatica*, 23(3):311–323.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., and Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, 105(3):430–445.
- Neuhaus, J. O. and Wrigley, C. (1954). The quartimax method: An analytic approach to orthogonal simple structure. *British Journal of Statistical Psychology*, 7(2):81–91.
- Rindskopf, D. and Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23:51–67.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464.
- Spearman, C. (1904). “general intelligence,” objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.
- Spearman, C. E. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3):271–295.
- Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.
- Vichi, M. (2017). Disjoint factor analysis with cross-loadings. *Advances in Data Analysis and Classification*, 11:563–591.
- Yates, A. (1987). *Multivariate Exploratory Data Analysis: A Perspective on Exploratory Factor Analysis*. SUNY Press.
- Zangwill, W. I. (1969). *Nonlinear Programming; a Unified Approach*. Prentice-Hall.