A parsimonious parameterization of a nonnegative correlation matrix

Carlo Cavicchia, Maurizio Vichi and Giorgia Zaccaria

Abstract Hierarchical relationships among manifest variables can be detected by analyzing their correlation matrix. To pinpoint the hierarchy underlying a multidimensional phenomenon, the Ultrametric Correlation Model (UCM) has been proposed with the aim of reconstructing a nonnegative correlation matrix via an ultrametric one. In this paper, we illustrate the mathematical advantages that a simple structure induced by the ultrametric property entails for the estimation of the UCM parameters in a maximum likelihood framework.

Key words: Ultrametric correlation matrix, parameterization of a correlation matrix, nonnegative correlation matrix, partitioned matrix

1 Introduction

Correlation matrices can be analyzed to detect hierarchical relationships among p manifest variables (MVs). A general correlation matrix has $p(p-1)/2$ parameters, each one representing the level of correlation between pairs of MVs. The model proposed by [2], called Ultrametric Correlation Model (UCM), provides a parsimonious representation of a nonnegative correlation matrix via an ultrametric one [3, pp. 58-59], while maintaining the relevant relations among MVs. The model aims

Carlo Cavicchia
Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands, e-mail: cavicchia@ese.eur.nl

Maurizio Vichi
University of Rome La Sapienza, Piazzale Aldo Moro 5, Rome, Italy, e-mail: maurizio.vichi@uniroma1.it

Giorgia Zaccaria
University of Rome La Sapienza, Piazzale Aldo Moro 5, Rome, Italy, e-mail: giorgia.zaccaria@uniroma1.it

at identifying consistent disjoint groups of MVs, each one representing a latent concept, and the hierarchical relationships among them. The non-negativity assumption turns out to be realistic in real applications (e.g., the g factor [8], the mental ability tests [1]) since many multidimensional phenomena are described by a set of variables that are concordant each other. By assuming that the variable space is partitioned into Q groups ($Q \in \{1,...,p\}$), each one associated with a latent concept, a $(p \times p)$ nonnegative correlation matrix is approximated in the UCM by

$$\mathbf{R}_\mathrm{u} = \mathbf{V}(\mathbf{R}_\mathrm{B} - \mathbf{I}_Q)\mathbf{V}' + \mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}' - \mathrm{diag}\big(\mathrm{dg}(\mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}')\big) + \mathbf{I}_p, \tag{1}$$

where $\mathbf{V}, \mathbf{R}_\mathrm{W}, \mathbf{R}_\mathrm{B}$ are the $(p \times Q)$ binary and row stochastic membership matrix, the $(Q \times Q)$ within-concept consistency matrix and the $(Q \times Q)$ ultrametric between-concept correlation matrix, respectively. $\mathbf{R}_\mathrm{u}$ turns out to be a $(2Q-1)$-ultrametric matrix which induces a hierarchy [2, Lemma 1 and Theorem 1] and it is associated with a parsimonious correlation structure. The ultrametric parameterization allows a decrease of the number of parameters needed to reconstruct a nonnegative correlation matrix. Indeed, $\mathbf{R}_\mathrm{u}$ can have as few as 1 parameter if $Q = 1$, or as many as $p-1$ parameters if $Q = p \geq 2$. Thus, the lower the number of the variable groups, the simpler the structure of the ultrametric correlation matrix.

In this short paper, we start to inspect the mathematical advantages that a simplified structure induced by the ultrametric property entails in the maximum likelihood estimation of the UCM under the assumption of Gaussian distributed data. We derive the main elements of the likelihood function, i.e., the simplified determinant and inverse of $\mathbf{R}_\mathrm{u}$, for some specific structures of the ultrametric correlation matrix. The results presented herein will be used, generalized and integrated in the extended paper along with the estimates of the UCM parameters in a maximum likelihood framework.

2 Multivariate normal distributions with the ultrametric correlation matrix

Let $\mathbf{X} = [X_1,...,X_p]'$ be a p-dimensional random vector with $\mathbf{X} \sim N_p(\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X)$ and $\mathbf{Y} = \mathrm{diag}(\mathrm{dg}(\boldsymbol{\Sigma}_X))^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}_X) \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_Y)$, where $\mathrm{dg}(\mathbf{A})$ is the vector including the elements of the diagonal of a square matrix $\mathbf{A}$ and $\boldsymbol{\Sigma}_Y = \mathbf{R}_\mathrm{u}$ is the $(p \times p)$ ultrametric correlation matrix in Eq. (1). The number of parameters of $\mathbf{R}_\mathrm{u}$ to be estimated depends on $Q \leq p$. Under the i.i.d. assumption, the log-likelihood function for the data $\mathbf{y} = [\mathbf{y}_1,...,\mathbf{y}_n]'$, obtained from the aforementioned transformation of $\mathbf{x} = [\mathbf{x}_1,...,\mathbf{x}_n]'$, is

$$\ell(\mathbf{R}_\mathrm{u};\mathbf{y}) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\mathbf{R}_\mathrm{u}| - \frac{n}{2}\mathrm{tr}(\mathbf{R}\mathbf{R}_\mathrm{u}^{-1}), \tag{2}$$

where $\mathbf{R}$ is the observed nonnegative correlation matrix.

Table 1 Ultrametric correlation structures.

Ultrametric Correlation Matrix	# parameters	Description
1-ultrametric correlation matrix	1	Constant correlation matrix[1]
3-ultrametric correlation matrix	$3+p$	2-block oblique correlation matrix
3-ultrametric correlation matrix with $\mathbf{R}_{\mathrm{B}} = \mathbf{I}_2$	$2+p$	2-block orthogonal correlation matrix
$(2Q-1)$-ultrametric correlation matrix	$2Q-1+p$	Q-block oblique correlation matrix[2]
$(2Q-1)$-ultrametric correlation matrix with $\mathbf{R}_{\mathrm{B}} = \mathbf{I_Q}$	$Q+p$	Q-block orthogonal correlation matrix[2]
$(2Q-1)$-ultrametric correlation matrix with $\mathbf{R}_{\mathrm{W}} = \lambda\mathbf{I_Q}$	$Q+p$	Q-block oblique correlation matrix with constant correlation within blocks[2]
$(2p-1)$-ultrametric correlation matrix	$p-1$	p-block correlation matrix[3]

[1] $\mathbf{V} = \mathbf{1}_p$.
[2] It is assumed $Q < p$.
[3] $\mathbf{V} = \mathbf{I}_p$.

Possible structures of the ultrametric correlation matrix $\mathbf{R}_{\mathrm{u}}$ are described in Table 1. They can be grouped in three main classes: 1-ultrametric correlation matrices, 3-ultrametric correlation matrices and $(2Q-1)$-ultrametric correlation matrices. The first one corresponds to an equicorrelation matrix in which a constant correlation occurs among MVs, i.e., $\mathbf{R}_{\mathrm{u}} = \lambda(\mathbf{1}_p\mathbf{1}_p' - \mathbf{I}_p) + \mathbf{I}_p$, where $\mathbf{1}_p$ is the p-dimensional vector of unitary elements and $\mathbf{I}_p$ is the identity matrix of order p. The second class contains two possible cases: (i) two-block oblique correlation matrix, where two groups of MVs have correlations within blocks equal to λ_1 and λ_2, respectively, and correlation between blocks equal to λ_3, i.e., Eq. (1) with $\mathbf{R}_{\mathrm{W}} = \mathrm{diag}([\lambda_1, \lambda_2]')$ and $\mathbf{R}_{\mathrm{B}} = \lambda_3(\mathbf{1}_2\mathbf{1}_2' - \mathbf{I}_2) + \mathbf{I}_2$; (ii) two-block orthogonal correlation matrix, where two groups of MVs have correlations within blocks equal to λ_1 and λ_2, respectively, and correlation among blocks equal to zero ($\lambda_3 = 0$), i.e., Eq. (1) with $\mathbf{R}_{\mathrm{W}} = \mathrm{diag}([\lambda_1, \lambda_2]')$ and $\mathbf{R}_{\mathrm{B}} = \mathbf{I}_2$. The third class contains four possible cases: (i) Q-block oblique correlation matrix, in which Q groups of MVs have correlations within blocks equal to the diagonal elements of $\mathbf{R}_{\mathrm{W}}$ and correlations between pairs of blocks equal to the off-diagonal elements of $\mathbf{R}_{\mathrm{B}}$, i.e. Eq. (1); (ii) Q-block orthogonal correlation matrix, in which Q groups of MVs have correlations within blocks equal to the diagonal elements of $\mathbf{R}_{\mathrm{W}}$ and zero correlation among them, i.e., Eq. (1) with $\mathbf{R}_{\mathrm{B}} = \mathbf{I}_Q$; (iii) Q-block oblique correlation matrix, with constant correlation λ within blocks and correlations between pairs of blocks equal to the off-diagonal elements of $\mathbf{R}_{\mathrm{B}}$, i.e., Eq. (1) with $\mathbf{R}_{\mathrm{W}} = \lambda\mathbf{I}_Q$; (iv) p-block correlation matrix, where $Q = p$, i.e., each group is composed of one MV, with correlations between pairs of MVs equal to the off-diagonal elements of $\mathbf{R}_{\mathrm{B}}$, i.e., Eq. (1) with $\mathbf{R}_{\mathrm{W}} = \mathbf{I}_p$.

In this section, we focus on three structures of $\mathbf{R}_{\mathrm{u}}$ shown in Table 1 - the 1-, 3- and $(2Q-1)$-ultrametric correlation matrix - illustrating the simplification of the main elements of Eq. (2) under the aforementioned parameterization of a nonnegative correlation matrix. For further details on the partitioned matrices which the following results are based on, see [4, 5].

2.1 Case 1: 1-ultrametric correlation matrix

If we assume that $Q = 1$, the 1-ultrametric correlation matrix can be written as $\mathbf{R}_\mathrm{u} = (1-\lambda)\mathbf{I}_p + \lambda \mathbf{1}_p\mathbf{1}'_p$, with $0 \leq \lambda < 1$. Thus, the determinant of $\mathbf{R}_\mathrm{u}$ is

$$\det(\mathbf{R}_\mathrm{u}) = [1+\lambda(p-1)](1-\lambda)^{p-1} \tag{3}$$

and its inverse - [3, p. 61] and [7] - is

$$\mathbf{R}_\mathrm{u}^{-1} = \frac{1}{1-\lambda}\left(\mathbf{I}_p - \frac{\lambda}{1+\lambda(p-1)}\mathbf{1}_p\mathbf{1}'_p\right). \tag{4}$$

2.2 Case 2: 3-ultrametric correlation matrix

If we assume that $Q = 2$, the 3-ultrametric correlation matrix can be written as $\mathbf{R}_\mathrm{u} = \lambda_3\mathbf{V}(\mathbf{1}_2\mathbf{1}'_2 - \mathbf{I}_2)\mathbf{V}' + \mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}' - \mathrm{diag}\big(\mathrm{dg}(\mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}')\big) + \mathbf{I}_p$, where $\mathbf{R}_\mathrm{W} = \mathrm{diag}([\lambda_1,\lambda_2]')$, $\lambda_1,\lambda_2,\lambda_3$ are the correlations within the first, the second group and between groups, respectively, with $0 \leq \lambda_3 \leq \lambda_s < 1$, $s = 1,2$. $\mathbf{V}$ is assumed to have contiguous rows representing MVs which belong to the same group after an appropriate row permutation. The 3-ultrametric correlation matrix can be rewritten as

$$\mathbf{R}_\mathrm{u} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix},$$

where $\mathbf{A} = (1-\lambda_1)\mathbf{I}_{p_1} + \lambda_1\mathbf{1}_{p_1}\mathbf{1}'_{p_1}$, $\mathbf{D} = (1-\lambda_2)\mathbf{I}_{p_2} + \lambda_2\mathbf{1}_{p_2}\mathbf{1}'_{p_2}$, $\mathbf{B} = \lambda_3(\mathbf{1}_{p_1}\mathbf{1}'_{p_2})$ and p_1, p_2 represent the number of MVs in the first and the second group, respectively, s.t. $p_1 + p_2 = p$. It is worth noticing that the matrices $\mathbf{A}$ and $\mathbf{D}$ are 1-ultrametric (see Section 2.1). It follows that the determinant of $\mathbf{R}_\mathrm{u}$ is

$$\begin{aligned}\det(\mathbf{R}_\mathrm{u}) = \; & \det(\mathbf{D})\det(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}') = [1+\lambda_2(p_2-1)]\,(1-\lambda_2)^{p_2-1} \\ & \cdot\left\{\left[\lambda_1 - \tfrac{p_2\lambda_3^2}{1-\lambda_2}\left(1-\tfrac{p_2\lambda_2}{1+\lambda_2(p_2-1)}\right)\right]p_1 + (1-\lambda_1)\right\}(1-\lambda_1)^{p_1-1}\end{aligned} \tag{5}$$

and the inverse of $\mathbf{R}_\mathrm{u}$ is

$$\mathbf{R}_\mathrm{u}^{-1} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{K} & \mathbf{N} \\ \mathbf{N}' & \mathbf{M} \end{bmatrix}, \tag{6}$$

where $\mathbf{K} = (\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}')^{-1} = \left[(1-\lambda_1)\mathbf{I}_{p_1} + \left[\lambda_1 - \frac{p_2\lambda_3^2}{1-\lambda_2}\left(1-\frac{p_2\lambda_2}{1+\lambda_2(p_2-1)}\right)\right]\mathbf{1}_{p_1}\mathbf{1}'_{p_1}\right]^{-1}$, $\mathbf{N} = -\mathbf{K}\mathbf{B}\mathbf{D}^{-1}$ and $\mathbf{M} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{B}'\mathbf{K}\mathbf{B}\mathbf{D}^{-1}$, $\mathbf{D}$ and $(\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{B}')$ nonsingular.

2.3 Case 3: (2Q-1)-ultrametric correlation matrix with zero correlation among blocks of variables

If we assume that $Q = 2$ and $\lambda_3 = 0$, i.e., the correlation between the variable groups is equal to zero, $\mathbf{R}_\mathrm{u} = \mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}' - \mathrm{diag}\big(\mathrm{dg}(\mathbf{V}\mathbf{R}_\mathrm{W}\mathbf{V}')\big) + \mathbf{I}_p$, where $\mathbf{R}_\mathrm{W} = \mathrm{diag}([\lambda_1, \lambda_2]')$. Then, the determinant of $\mathbf{R}_\mathrm{u}$ is

$$\det(\mathbf{R}_\mathrm{u}) = [1+\lambda_1(p_1-1)]\,[1+\lambda_2(p_2-1)]\,(1-\lambda_1)^{p_1-1}(1-\lambda_2)^{p_2-1} \qquad (7)$$

and its inverse is

$$\begin{aligned}\mathbf{R}_\mathrm{u}^{-1} &= \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0}_{p_1,p_2} \\ \mathbf{0}_{p_2,p_1} & \mathbf{D}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{1-\lambda_1}\left(\mathbf{I}_{p_1} - \frac{\lambda_1}{1+\lambda_1(p_1-1)}\mathbf{1}_{p_1}\mathbf{1}'_{p_1}\right) & \mathbf{0}_{p_1,p_2} \\ \mathbf{0}_{p_2,p_1} & \frac{1}{1-\lambda_2}\left(\mathbf{I}_{p_2} - \frac{\lambda_2}{1+\lambda_2(p_2-1)}\mathbf{1}_{p_2}\mathbf{1}'_{p_2}\right) \end{bmatrix}. \qquad (8)\end{aligned}$$

In order to generalize the latter case to Q groups with no correlation among them, $\mathbf{R}_\mathrm{u}$ can be rewritten as a block diagonal matrix

$$\mathbf{R}_\mathrm{u} = \begin{bmatrix} (1-\lambda_1)\mathbf{I}_{p_1} + \lambda_1\mathbf{1}_{p_1}\mathbf{1}'_{p_1} & \mathbf{0}_{p_1,p_2} & \dots & \mathbf{0}_{p_1,p_Q} \\ \mathbf{0}_{p_2,p_1} & (1-\lambda_2)\mathbf{I}_{p_2} + \lambda_2\mathbf{1}_{p_2}\mathbf{1}'_{p_2} & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \mathbf{0}_{p_Q,p_1} & \dots & \dots & (1-\lambda_Q)\mathbf{I}_{p_Q} + \lambda_Q\mathbf{1}_{p_Q}\mathbf{1}'_{p_Q} \end{bmatrix},$$

with $p_1 + p_2 + \dots + p_Q = p$. Thus,

$$\begin{aligned}\det(\mathbf{R}_\mathrm{u}) &= [1+\lambda_1(p_1-1)]\cdot[1+\lambda_2(p_2-1)]\cdot\ldots\cdot[1+\lambda_Q(p_Q-1)]\cdot(1-\lambda_1)^{p_1-1} \\ &\cdot\ (1-\lambda_2)^{p_2-1}\cdot\ldots\cdot(1-\lambda_Q)^{p_Q-1} \qquad (9)\end{aligned}$$

and

$$\mathbf{R}_\mathrm{u}^{-1} = \begin{bmatrix} \frac{1}{1-\lambda_1}\left(\mathbf{I}_{p_1} - \frac{\lambda_1}{1+\lambda_1(p_1-1)}\mathbf{1}_{p_1}\mathbf{1}'_{p_1}\right) & \dots & \mathbf{0}_{p_1,p_Q} \\ \dots & \dots & \dots \\ \mathbf{0}_{p_Q,p_1} & \dots & \frac{1}{1-\lambda_Q}\left(\mathbf{I}_{p_Q} - \frac{\lambda_Q}{1+\lambda_Q(p_Q-1)}\mathbf{1}_{p_Q}\mathbf{1}'_{p_Q}\right) \end{bmatrix} \qquad (10)$$

which is a block diagonal matrix, where each block is the inverse of a 1-ultrametric correlation matrix (see Section 2.1).

3 Conclusions and Further Developments

In this paper, a parsimonious parameterization of a nonnegative correlation matrix via an ultrametric correlation one is proposed. Moreover, we inspect the advantages that a simple structure, induced by an ultrametric correlation matrix, entails in the maximum likelihood estimation of the Ultrametric Correlation Model parameters, assuming the normality of the data. The parameterization is studied to derive, in closed form, the equation of the determinant and inverse of an ultrametric correlation matrix in three cases, i.e., 1-ultrametric correlation matrix, 3-ultrametric correlation matrix and $(2Q-1)$-ultrametric correlation matrix with no correlation among groups of MVs. These elements are crucial in the maximum likelihood estimation of the Ultrametric Correlation Model parameters. The ultrametric correlation matrix allows a decrease of the number of parameters to be estimated compared to a general correlation matrix with $p(p-1)/2$ parameters. The generalization of the results herein to a $(2Q-1)$-ultrametric correlation matrix for estimating the Ultrametric Correlation Model in a maximum likelihood framework will be illustrated in an extended paper.

Our goal for future studies is also to introduce a test for correlation in order to pinpoint non-significant correlations in the ultrametric matrix; this can further reduce the number of parameters in the model. Furthermore, the ultrametric correlation matrix in Eq. (1) can be used to parameterize a nonnegative correlation matrix in Gaussian mixture models [6] when a multidimensional phenomenon is studied on observations coming from $G < +\infty$ sub-populations with a Gaussian distribution.

References

1. Bechtoldt, H.: An Empirical Study of the Factor Analysis Stability Hypothesis. Psychometrika **26**(4), 405–432 (1961)
2. Cavicchia, C., Vichi, M., Zaccaria, G.: The Ultrametric Correlation Matrix for Modelling Hierarchical Latent Concepts. Advances in Data Analysis and Classification **14**(4), 837–853 (2020)
3. Dellacherie, C., Martinez, S., San Martin, J.: Inverse M-Matrices and Ultrametric Matrices. Springer International Publishing, Lecture Notes in Mathematics (2014)
4. Lütkepohl, H.: Handbook of Matrices. John Wiley & Sons, England (1996)
5. Magnus, J.R., Neudecker, H.: Matrix Differential Calculus with Applications in Statistics and Econometrics. John Wiley & Sons, Chicester (1988)
6. McLachlan, G., Peel, D.: Finite Mixture Models. John Wiley & Sons, New York (2000)
7. Miller, K.S.: On the Inverse of the Sum of Matrices. Math. Mag. **54**(2), 67–72 (1981)
8. Spearman, C.E.: The Abilities of Man: Their Nature and Measurement. Macmillan, New York (1927)