The ultrametric covariance model for modelling teachers' job satisfaction

Il modello di covarianza ultrametrica per lo studio della soddisfazione lavorativa dei professori

Carlo Cavicchia, Maurizio Vichi and Giorgia Zaccaria

Abstract Multidimensional phenomena are often characterised by nested latent concepts ordered in a hierarchical structure, from the most specific to the most general ones. In this paper, we model a nonnegative data covariance matrix by extending the Ultrametric Correlation Model to covariance matrices. The proposal is a parsimonious model which identifies a partition of variables in a reduced number of groups, and the relationships among them via the ultrametric property. The proposed model is applied to investigate the relationships among the dimensions of the Teachers' Job Satisfaction in Italian secondary schools.

Abstract *I fenomeni multidimensionali sono spesso caratterizzati da concetti latenti ordinati in una struttura gerarchica, dai più specifici al più generale. In questo articolo ci proponiamo di modellare una matrice di coviarianza nonnegativa, estendendo il modello chiamato Ultrametric Correlation Model alle matrici di covarianza. La proposta metodologica si esplica in un modello parsimonioso che identifica una partizione di variabili in un numero ridotto di gruppi e le loro relazioni mediante la proprietà di ultrametricità. Il modello proposto è applicato allo studio delle relazioni tra le dimensioni della soddisfazione lavorativa dei professori nelle scuole italiane superiori di secondo grado.*

Key words: Ultrametric matrices, hierarchical structures, teachers' job satisfaction, confirmatory analysis, dimensionality reduction

---

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

Maurizio Vichi
Department of Statistical Sciences, University of Rome La Sapienza, Rome, Italy
e-mail: `maurizio.vichi@uniroma1.it`

Giorgia Zaccaria
Department of Statistical Sciences, University of Rome La Sapienza, Rome, Italy
e-mail: `giorgia.zaccaria@uniroma1.it`

1 Introduction

Multidimensional phenomena are often characterised by a hierarchy of nested latent concepts (dimensions) with different levels of abstraction, from the most specific to the most general ones. The study of these phenomena needs specific models since the traditional ones, usually used to reconstruct the relationships among variables (e.g., Factor Analysis, FA, [1]), fail in the definition of a hierarchical structure over them. Cavicchia et al. [4] introduced a parsimonious simultaneous model, named Ultrametric Correlation Model (UCM), to reconstruct a nonnegative data correlation matrix of order p via an ultrametric correlation one. The ultrametric property allows both to identify a partition of variables in $Q \leq p$ groups and the relationships among them by defining two difference features: the within-concept consistency and the correlation between groups.

In this paper, we introduce a new model, called *Ultrametric Covariance Model* (UCovM), to reconstruct a nonnegative covariance matrix by extending the one proposed by Cavicchia et al. [4] for nonnegative correlation matrices. Similarly to UCM, UCovM defines a hierarchy of latent concepts by pinpointing a variable partition in Q groups characterised by three features: the variance of a group, the covariance within the group and the covariance between groups. Since a decreasing order is imposed on these features, two variables belonging to the same group are more concordant than two belonging to different groups. Although the nonnegativity assumption might seem restrictive, it turns out to be realistic in many real-data applications. We apply UCovM to Teachers Job Satisfaction data set [7] in order to investigate the hierarchical relationships between the six dimensions defining the job satisfaction for teachers.

2 Background

Let us recall the definition of an ultrametric matrix [6, pp. 58-59], which differs from an ultrametric distance matrix even if there exists a relationship between the two.

Definition 1. A nonnegative matrix **U** of order p is said to be ultrametric if

(i) $u_{jl} = u_{lj}$ for all $j,l = 1,\dots,p$ (symmetry);
(ii) $u_{jj} \geq \max\{u_{lj} : l = 1,\dots,p\}$ for $j = 1,\dots,p$ (column pointwise diagonal dominance);
(iii) $u_{jl} \geq \min\{u_{ji}, u_{il}\}$, for $i,j,l = 1,\dots,p$ (ultrametric inequality).

Every ultrametric matrix turns out to be positive semi-definite, as demonstrated by Dellacherie et al. [6, pp. 60-61]. Considering a nonnegative data covariance matrix **S** of order p, with elements $s_{jl} \in \mathbb{R}^+$ (the set of nonnegative real numbers), $j,l = 1,\dots,p$, it is (i) symmetric and (ii) positive semi-definite by definition. If conditions (ii) and (iii) hold, **S** is an ultrametric covariance matrix.

3 Methodology

Let $\mathbf{S}$ be a nonnegative data covariance matrix of order p. The problem we want to deal with can be formalised as

$$\mathbf{S} = \mathbf{S}_{\text{u}} + \mathbf{E}, \tag{1}$$

where $\mathbf{S}_{\text{u}}$ is an ultrametric covariance matrix of order p and $\mathbf{E}$ is an error matrix of the same order.

The Ultrametric Covariance Model (UCovM) defines an ultrametric covariance matrix for modelling hierarchical latent concepts, which is formally specified as follows

$$\mathbf{S}_{\text{u}} = \mathbf{V}(\mathbf{S}_{\text{W}} + \mathbf{S}_{\text{B}})\mathbf{V}' - \mathbf{V}\mathbf{S}_{\text{W}}\mathbf{V}' \odot \mathbf{I}_p + \mathbf{V}\mathbf{S}_{\text{V}}\mathbf{V}' \odot \mathbf{I}_p, \tag{2}$$

subject to constraints

$$\mathbf{V} = [v_{jq} \in \{0,1\} : j = 1,\dots,p, q = 1,\dots,Q]; \tag{3}$$

$$\mathbf{V}\mathbf{1}_Q = \mathbf{1}_p \quad \text{i.e.} \quad \textstyle\sum_{q=1}^{Q} v_{jq} = 1 \quad j = 1,\dots,p; \tag{4}$$

$$\mathbf{S}_{\text{B}} = \mathbf{S}'_{\text{B}}, \text{diag}(\mathbf{S}_{\text{B}}) = \mathbf{0}, {}_{B}s_{qh} \geq \min\{{}_{B}s_{qt}, {}_{B}s_{ht}\}\ q,h,t = 1,\dots,Q, t \neq h \neq q; \tag{5}$$

$$\min\{{}_{W}s_{qq} : q = 1,\dots,Q\} \geq \max\{{}_{B}s_{qh} : q,h = 1,\dots,Q, h \neq q\}; \tag{6}$$

$${}_{V}s_{qq} \geq {}_{W}s_{qq}, q = 1,\dots,Q, \tag{7}$$

where $\mathbf{I}_p$ is an identity matrix of order p, $\odot$ is the Hadamard (element-wise) product and $\text{diag}(\mathbf{S}_{\text{B}})$ identifies the main diagonal of $\mathbf{S}_{\text{B}}$.

$\mathbf{S}_{\text{V}}$ and $\mathbf{S}_{\text{W}}$ are diagonal matrices, whose diagonal elements represent the variances of and the covariances within the Q variable groups, respectively, whereas the covariances between them are expressed by the off-diagonal elements of $\mathbf{S}_{\text{B}}$. Since constraint (5), (6) and (7) hold, an ordering between the elements of $\mathbf{S}_{\text{V}}$, $\mathbf{S}_{\text{W}}$ and $\mathbf{S}_{\text{B}}$ exists. This leads to a hierarchy of latent concepts, each one associated with a variable group, whose hierarchical levels are defined by the covariances within and between groups, i.e., the diagonal and off-diagonal elements of $\mathbf{S}_{\text{W}}$ and $\mathbf{S}_{\text{B}}$, respectively. Specifically, the higher the covariance among two variable groups (or variables themselves), the stronger the concordance among them and the earlier they are merged together.

UCovM allows pinpointing groups of variables, each one associated with a dimension, by reducing the dimensonality of the phenomenon under study, and identifying new latent concepts and the hierarchical relationships among them. Thus, UCovM is an exploratory, parsimonious and simultaneous model. If $\mathbf{V}$ is set a priori, i.e., the variable partition is fixed, then the model can be applied in a confirmatory approach.

The proposal is estimated in a least-squares framework and implemented with a coordinate descent algorithm.

Table 1: List of variables for each dimension of the Teachers' Job Satisfaction data set[a] and the corresponding Cronbach's α.

Dimension	Dimension ID	Variables	α
Communication	Comm	1, 2, 3, 4, 5	0.8136
External School Image	Imag	6, 7	0.8582
Involvement	Invo	8, 9, 10, 11, 12, 13	0.8999
Leadership	Lead	14, 15, 16, 17	0.9021
School Climate	Clim	18, 19, 20, 21, 22, 23	0.8817
Infrastructure	Infr	24, 25	0.7052

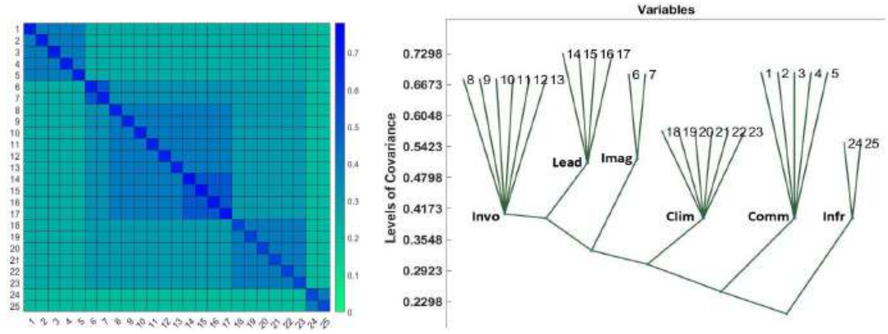
[a] See [7, Table 1] for a complete description of the variables.

4 Teachers' Job Satisfaction: differences between the overall ultrametric covariance structure and those by gender

Job Satisfaction is a multidimensional phenomenon characterised by different dimensions affecting feelings and emotions of employees towards their job. We apply the UCovM to study Teachers' Job Satisfaction (TJS) and investigate the hierarchical relationships among the factors that contribute to define TJS. The analysis is based upon the survey conducted by Sarnacchiaro et al. [7] in four Italian state secondary schools. Table 1 shows the dimensions of the TJS and the partition of variables in six groups, each one associated with the corresponding dimension. Cronbach's α [5] for each group is also computed and all dimensions result reliable. Moreover, Cronbach's α for the whole data set turns out to be 0.9528, revealing the strong reliability of the general latent concept, i.e. the TJS. Additional variables pertaining socio-demographic features are also measured; among them, we consider the variable *Gender* in order to compare the hierarchical structure defining TJS on the aforementioned data set with those estimated differently for female and male.

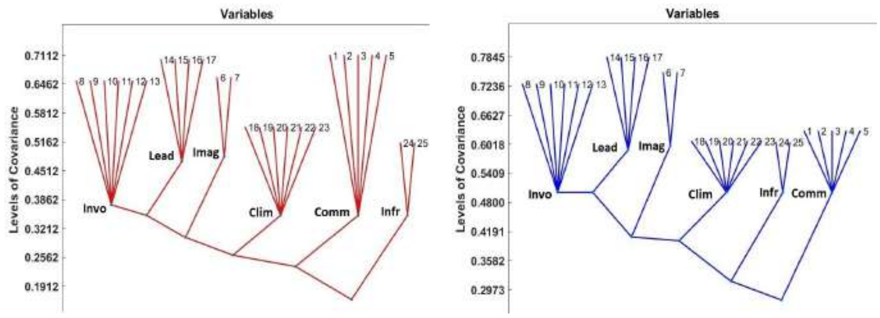
Firstly, we performed UCovM in a confirmatory approach on the covariance matrix - containing nonnegative values - of the whole data set. The partition in six groups of variables corresponding to the dimensions of the TJS is clearly visible in the covariance matrix (Figure 1a). The groups which are mostly concordant within them are those associated with *Leadership* and *External School Image*. As shown in Figure 1b, the first aggregation lumps together *Involvement* and *Leadership*, which, indeed, have a high impact on TJS [7]. The following aggregations show a constant trend by adding one at a time the remaining dimensions - connected with the school-based factors - to the first group, up to the *Infrastructure*, which is the less concordant dimension with the others (the covariance between the broader group with five dimensions and *Infrastructure* is equal to 0.2045).

Comparing these results with those obtained by implementing UCovM by gender - both the covariance matrices are nonnegative - we can notice some differences between TJS for female and male (Figure 2). The hierarchy over the six dimensions of TJS for female (Figure 2a) is similar to that obtained on the whole data set. Indeed, even if the covariances within and between groups are less strong than



(a) Heatmap of the covariance matrix (b) Path diagram representation resulting from UCovM

Fig. 1: Graphical representations of relationships among the dimensions of TJS for the whole data set.



(a) Female (b) Male

Fig. 2: Path diagram representation of the TJS resulting from UCovM by gender.

those on the whole data set, the aggregations are the same. This happens also because the percentage of women in the data set is greater than that of men. On the other hand, the six dimensions of TJS for male show a slightly different hierarchical structure (Figure 2b). The first aggregation lumps together *Involvement* and *Leadership* as well; therefore, the variables pertaining *Involvement* are merged with those associated with *Leadership* such that the covariance within the former group is equal to that between the two variable groups. Looking at Figure 1a, the difference between the covariance magnitude of these two variable groups and that of the variable group associated with *Involvement* seems to be slight. The other aggregations show a constant trend, with covariances between *Involvement, Leadership, External School Image* and *School Climate* greater than 0.38. The last two aggregations are reversed with respect to those for female.

5 Conclusions

The model proposed herein, called Ultrametric Covariance Model, is an extension of the Ultrametric Correlation Model, introduced by Cavicchia et al. [4], to covariance matrices. It aims at reconstructing the hierarchical relationships existing among variables by modelling a nonnegative covariance matrix via an ultrametric covariance one.

UCovM was applied on a real data set in order to study the hierarchical relationships among the six dimensions of the Teachers' Job Satisfaction. The analysis is conducted on the overall data set [7] and differently by gender. The hierarchy of the TJS dimensions is slightly different between male and female. Comparing the results obtained by UCovM with those attained by UCM, we can highlight that for the whole data and the males' ones the second and the third aggregations are swapped, whereas the hierarchy remains the same for the females' data. Conversely to UCM, UCovM allows to inspect the variability of each group of the variable partition. Some comparisons with other methodologies, as Higher-Order models [2] and hierarchical clustering methods, were carried out: in both cases the models' fit pointed out that a simultaneous methodology was needed. Cavicchia et al. [3] in turn demonstrated that hierarchical clustering techniques had some limitations in detecting hierarchical relationships among variables if compared to simultaneous methodologies as UCM.

Our goals for future studies are to implement a bootstrap test to assess if the difference between the parameters of the UCovM estimated by gender is statistically significant; to study the TJS according to other socio-demographic features and to build an R and/or Matlab package to implement the proposal.

Acknowledgements The authors would like to thank Prof. Pasquale Sarnacchiaro, and the authors of [7], to have shared with us their data.

References

1. Anderson, T., Rubin, H.: Statistical inferences in factor analysis. Proceedings of the Third Symposium on Mathematical Statistics and Probability **5**, 111–150 (1956)
2. Cattell, R.: The scientific use of factor analysis in behavioral and life sciences. Plenum (1978)
3. Cavicchia, C., Vichi, M., Zaccaria, G.: Exploring hierarchical concepts: theoretical and application comparison. In: T. Imaizumi, A. Nakayama, S. Yokoyama (eds.) Advanced Studies in Behaviormetrics and Data Science. Behaviormetrics: Quantitative Approaches to Human Behavior, vol. 5, pp. 315–328. Springer, Singapore (2020)
4. Cavicchia, C., Vichi, M., Zaccaria, G.: The ultrametric correlation matrix for modelling hierarchical latent concepts. Adv Data Anal Classif **14**(4), 837–853 (2020)
5. Cronbach, L.: Coefficient alpha and the internal structure of tests. Psychometrika **16**(3), 297–334 (1951)
6. Dellacherie, C., Martinez, S., Martin, J.S.: Inverse M-matrices and ultrametric matrices. Lecture Notes in Mathematics. Springer International Publishing (2014)
7. Sarnacchiaro, P., Scippacercola, S., Malafronte, P.: A statistical model for self-evaluation of teacher's satisfaction: a study in an Italian secondary school. Electron. J. Appl. Stat. Anal. **13**(3), 637–656 (2019)