An ultrametric model for building a composite indicator system to study climate change in European countries

Un modello ultrametrico per la costruzione di un sistema di indicatori compositi per lo studio del cambiamento climatico nei paesi europei

Giorgia Zaccaria and Pasquale Sarnacchiaro

Abstract Nowadays climate change is one of the most urgent topics in public debate, also included into the 2030 Agenda for Sustainable Development adopted by the United Nations, because of its risks for human life. It represents a multidimensional phenomenon defined by different dimensions that pertain to greenhouse gas emissions, human causes of climate change, impacts on humans and natural systems, and efforts of human to avoid and adapt to the consequences, each of which is described by a set of variables directly observed. In this paper, we introduce a new model-based approach to study this phenomenon and its different characterization in European countries. The proposal aims at building a system of composite indicators in order to understand the main determinants of this phenomenon, and to provide guidelines for policy decisions to combat climate change in the European Union framework.

Abstract *Il cambiamento climatico è uno dei temi più urgenti nel dibattito pubblico, incluso nell'Agenda 2030 per lo sviluppo sostenibile adottata dalle Nazioni Unite, a causa dei rischi che può comportare per la vita umana. Esso rappresenta un fenomeno multidimensionale definito da diverse dimensioni, ognuna delle quali misurata da un insieme di variabili direttamente osservabili, che riguardano le emissioni di gas serra, le cause del cambiamento climatico, le sue conseguenze per gli umani e la natura e gli sforzi messi in campo dal genere umano per evitare e adattarsi a tali conseguenze. In questo lavoro, introduciamo un nuovo approccio basato su modello per studiare questo fenomeno e le sue diverse caratterizzazioni nei paesi europei. La proposta metodologica ha l'obiettivo di costruire un sistema di indicatori compositi per identificare i principali fattori che determinano il cam-*

Giorgia Zaccaria
Department of Law and Economics, University of Rome Unitelma Sapienza, Rome, Italy
e-mail: giorgia.zaccaria@unitelmasapienza.it

Pasquale Sarnacchiaro
University of Naples Federico II, Naples, Italy
e-mail: sarnacch@unina.it

biamento climatico e fornire di conseguenza indicazioni per attuare politiche nel contesto europeo volte a combattere il cambiamento climatico.

Key words: Ultrametric models, hierarchical structures, model-based clustering, composite indicators, climate change, sustainable development goals

1 Introduction

Climate change is a compelling topic in public debate because of its risks for human life. One of the main causes of this phenomenon pertain human activities, e.g., industrial processes, energy consumption by fossil fuels, agricultural exploitation of lands. An urgent action to combat these changes is needed. For these reasons, the United Nations included "climate action" within the goals of the 2030 Agenda for Sustainable Development, adopted in 2015 [6], that are monitored at European level by Eurostat throughout indicators and statistics.

In order to study and evaluate multidimensional phenomena as climate change, composite indicators are often used since they are able to summarize a big amount of information via a synthetic measure. Their construction is characterized by different choices that are usually normative, i.e., based upon experts' opinions or a theory known a priori [8]. For these reasons, methods for composite indicator construction are often criticized and considered to be not statistically rigorous [7]. In this paper, we discuss a new model-based approach to build a system of composite indicators for evaluating climate change at European level. In order to study this phenomenon and its different characterization in Europe, we consider the model introduced by Cavicchia, Vichi and Zaccaria [3]. The latter is an ultrametric Gaussian Mixture Model (GMM), where each component of the mixture has an ultrametric structure which is one-to-one associated with a hierarchy of latent dimensions. The idea herein is to extend the ultrametric GMM proposed by [3] to build a system of composite indicators, i.e., including the quantification of latent concepts for each level of the hierarchy, in order to study on which dimensions and in which countries policies to address climate change should be focused.

The paper is organized as follows. In Section 2 the essential background to introduce the methodology used in the paper is provided. Section 3 illustrates the ultrametric Gaussian Mixture Model, which is then applied to study climate change in European countries (Section 4). A final discussion completes the paper in Section 5.

2 Background

In order to define the parsimonious parameterization of the covariance matrix in Section 3, let us introduce the definition of a strict extended ultrametric covariance matrix [3].

Definition 1 (Strict Extended Ultrametric Covariance Matrix). A matrix $\boldsymbol{\Sigma}$ is said to be a *strict extended ultrametric covariance matrix* if all its elements $\sigma_{jl} \in \mathbb{R}$, for $j,l = 1,\dots,p$, and the following conditions hold:

(i) *symmetry*: $\sigma_{jl} = \sigma_{lj}$ for $j,l = 1,\dots,p$;
(ii) *positivity of the diagonal*: $\sigma_{jj} > 0$ for all $j = 1,\dots,p$;
(iii) *ultrametric inequality*: $\sigma_{jl} \geq \min\{\sigma_{jh}, \sigma_{lh}\}$, for $j,l,h = 1,\dots,p$;
(iv) *strictly diagonal dominance*: $\sigma_{jj} > \sum_{\substack{l=1 \\ l\neq j}}^{p} |\sigma_{jl}|$ for $j = 1,\dots,p$.

Condition (iv), together with conditions (i) and (ii), is sufficient for the positive definiteness of a matrix as shown by Brouwer and Haemers [2]. Definition 1 extends the definition of an ultrametric matrix introduced by Dellacherie et al. [4, pp. 60-61], which is limited to nonnegative matrices.

In the next section, we model the component covariance structure of a Gaussian Mixture Model via a strict extended ultrametric covariance matrix.

3 Methodology

Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n)$ be a random sample - where $\boldsymbol{x}_i$ is a p-dimensional random vector - which is drawn from a population composed of G subpopulations. Suppose that, conditional on the membership to the subpopulation, the density of $\boldsymbol{x}_i$ is the following

$$f(\boldsymbol{x}_i|\boldsymbol{\Psi}) = \sum_{g=1}^{G} \pi_g\, \phi\big(\boldsymbol{x}|\boldsymbol{\mu}_g, \mathbf{V}_g(\boldsymbol{\Sigma}_{\mathrm{W}g} + \boldsymbol{\Sigma}_{\mathrm{B}g})\mathbf{V}'_g - \mathrm{diag}(\mathbf{V}_g\boldsymbol{\Sigma}_{\mathrm{W}g}\mathbf{V}'_g) + \mathrm{diag}(\mathbf{V}_g\boldsymbol{\Sigma}_{\mathrm{V}g}\mathbf{V}'_g)\big), \tag{1}$$

where $\pi_g, g = 1,\dots,G$, are the mixing proportions (prior probabilities), and the component covariance structure is parameterized according to a strict extended ultrametric covariance matrix subject to the following constraints

$$\mathbf{V} = [v_{jq} \in \{0,1\} : j = 1,\dots,p, q = 1,\dots,Q]; \tag{2}$$

$$\mathbf{V}\mathbf{1}_Q = \mathbf{1}_p \quad \text{i.e.} \quad \sum_{q=1}^{Q} v_{jq} = 1 \quad j = 1,\dots,p; \tag{3}$$

$$\boldsymbol{\Sigma}_{\mathrm{B}} = \boldsymbol{\Sigma}'_{\mathrm{B}}, \mathrm{diag}(\boldsymbol{\Sigma}_{\mathrm{B}}) = \mathbf{0}, {}_B\sigma_{qh} \geq \min\{{}_B\sigma_{qs}, {}_B\sigma_{hs}\}\ q,h,s = 1,\dots,Q,\ s \neq h \neq q; \tag{4}$$

$$\min\{{}_W\sigma_{qq} : q = 1,\dots,Q\} \geq \max\{{}_B\sigma_{qh} : q,h = 1,\dots,Q, h \neq q\}; \tag{5}$$

$${}_V\sigma_{qq} > |{}_W\sigma_{qq}|\Big(\sum_{l=1}^{p} v_{lq} - 1\Big) + \sum_{\substack{h=1 \\ h\neq q}}^{Q} |{}_B\sigma_{qh}| \sum_{l=1}^{p} v_{lh} \quad q = 1,\dots,Q. \tag{6}$$

$\boldsymbol{\Psi} = \{\pi_g, \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_{\mathrm{V}g}, \boldsymbol{\Sigma}_{\mathrm{W}g}, \boldsymbol{\Sigma}_{\mathrm{B}g}, \mathbf{V}_g : g = 1, \ldots, G\}$ is the overall parameter vector, where $\mathbf{V}_g$ is a membership matrix which defines a partition of the variable space into a reduced number of groups Q, each one associated with a latent dimension, $\boldsymbol{\Sigma}_{\mathrm{V}g}$ is a diagonal matrix of order Q whose diagonal entries represent the variance of the variable groups, $\boldsymbol{\Sigma}_{\mathrm{W}g}$ is a diagonal matrix of order Q whose diagonal values identify the covariance within the variable groups, and $\boldsymbol{\Sigma}_{\mathrm{B}g}$ is the matrix of order Q whose off-diagonal elements represent the covariance between the groups. The ultrametric paramaterization of the component covariance structure leads to a hierarchy of latent concepts which can assume different characterizations in the components (subpopulations) of the mixture.

The proposal is estimated via a grouped coordinate ascent algorithm [9, 1] by maximizing the objective function defined by Hathaway [5], which is equivalent to maximize the log-likelihood function of (1).

4 A composite indicator system for evaluating climate change in European countries

Every year, Eurostat collects data for monitoring progress towards the Sustainable Development Goals (SDGs) at European level. In order to study climate change, impacts on humans on it and actions to combat it, indicators related to different goals and measured for the 27 European countries and the United Kingdom can be considered. Indeed, other than the indicators defined to monitor Goal 13 - *Climate Action*, e.g., greenhouse gas emissions, we consider variables for measuring the level of pollution in groundwater and rivers (Goal 6 - *Clean Water and Sanitation*), energy consumption and renewable energy production (Goal 7 - *Affordable and Clean Energy*), exposure to air pollution (Goal 11 - *Sustainable Cities and Communities*) and generation of waste (Goal 12 - *Responsible Consumption and Production*). In detail, we include in our analysis indicators from different goals of the SDGs measured on the 27 European countries and the United Kingdom, and we use their population size (i.e., number of inhabitants) to normalize the variables in order to perform comparison among countries.

The motivation of this study lies on the assumption that it is crucial to combine information from different goals in order to define a general composite indicator for measuring climate change, and a system of specific composite indicators that define it. Indeed, the researcher could be interested in identifying dimensions of the phenomenon under study to which policies should be addressed in order to progress towards climatic goal.

5 Conclusions

In this paper, we provide a new hierarchical model-based approach to build a system of composite indicators, and inspect the nested relationships among specific indicators that define the general latent concept under study. The proposal is based upon the ultrametricity notion, which is related to hierarchical (tree) structures, and its implementation in Gaussian mixture models [3]. In particular, we focus on the analysis of climate change by identifying its dimensions and its different characterizations in European countries. The proposal aims at providing guidelines for policy decisions to combat climate change in the European Union framework.

References

1. Bezdek, J., Hathaway, R., Howard, R., Wilson, C., Windham, M.: Local convergence analysis of a grouped variable version of coordinate descent. J. Optim. Theory. Appl. **54**(3), 471–477 (1987)
2. Brouwer, A., Haemers, W.: Spectra of graphs. Springer, New York (2012)
3. Cavicchia, C., Vichi, M., Zaccaria, G.: Gaussian mixture model with an extended ultrametric covariance structure. Adv Data Anal Classif, doi: 10.1007/s11634-021-00488-x (2022)
4. Dellacherie, C., Martinez, S., Martin, J.S.: Inverse M-matrices and ultrametric matrices. Lecture Notes in Mathematics. Springer International Publishing (2014)
5. Hathaway, R.: Another interpretation of the EM algorithm for mixture distributions. Stat. Probab. Lett. **4**(2), 53–56 (1986)
6. Heads of State and Government and High Representatives: Transforming our world: the 2030 agenda for sustainable development, a/res/70/1. Tech. rep., United Nations (2015). URL `https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf`
7. Mazziotta, M., Pareto, A.: Methods for constructing composite indices: One for all or all for one? Rivista Italiana di Economia Demografia e Statistica **67**(2), 67–80 (2013)
8. OECD-JRC: Handbook on Constructing Composite Indicators. Methodology and User Guide. OECD Publisher, Paris (2008)
9. Zangwill, W.: Nonlinear programming: a unified approach. Prentice-Hall, Englewood Cliffs (1969)