Journal of Multivariate Analysis 155 (2017) 317–333

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Multivariate dependence modeling based on comonotonic factors

Lei Hua [a,*], Harry Joe [b]

[a] Division of Statistics, Northern Illinois University, DeKalb, IL 60115, USA
[b] Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2

ARTICLE INFO

Article history:
Received 14 July 2016
Available online 17 January 2017

AMS subject classifications:
62H25
62H99

Keywords:
Bi-factor
Copula
Dependence clusters/groups
Laplace Transform
Parsimonious dependence models

ABSTRACT

Comonotonic latent variables are introduced into general factor models, in order to allow non-linear transformations of latent factors, so that various multivariate dependence structures can be captured. Through decomposing each univariate marginal into several components, and letting some components belong to different sets of comonotonic latent variables, a great variety of multivariate models can be constructed, and their induced copulas can be used to model various multivariate dependence structures. The paper focuses on an extension of Archimedean copulas constructed by Laplace Transforms of positive random variables. The corresponding comonotonic factor models with one set of comonotonic latent variables and multiple sets of comonotonic latent variables are studied. In particular, we propose several parametric comonotonic factor models that are useful in accommodating both within-group and between-group dependence with possible asymmetric tail dependence. Numerical methods for estimation with the resulting copula models are discussed. There is an application using a dataset of body composition measurements to demonstrate the usefulness of the proposed parsimonious dependence models.

1. Introduction

In classical statistics based on the multivariate normal distribution, factor models are popular choices for parsimonious dependence modeling when the dimension of data is large. The simplest factor model is a conditional independence model when the observed variables are conditionally independent given the latent variables.

In multivariate non-Gaussian modeling based on copulas, the modeling of univariate marginal distributions is separated from that of the dependence structure. The vine copula or pair-copula construction [1,2,19,28] has become a flexible modeling approach. This construction extends multivariate normal dependence models where the correlation matrix has been reparametrized to partial correlation vines. In order to achieve parsimonious dependence for a large number of variables, truncated vine [4] models have been proposed.

Recently, there has been much interest in developing copula models with factor structures, in order to allow for tail dependence, and upper and lower tail asymmetry. Examples of copula models with factor structures include: (a) extensions of Archimedean copulas in [23,30,32]; (b) linear forms based on Student-t random variables by [13,27]; (c) general factor copulas based on vines rooted at latent variables by [17,18]; and (d) grouped normal mixture models in [5,20,24].

* Corresponding author.
E-mail addresses: Lhua@niu.edu (L. Hua), Harry.Joe@ubc.ca (H. Joe).

http://dx.doi.org/10.1016/j.jmva.2017.01.008
0047-259X/

In this paper, we employ sets of comonotonic latent variables in constructing multivariate models. In particular, we focus on an extension of exchangeable Archimedean copulas. Despite the extensive studies from a theoretical point of view, Archimedean copulas are generally not flexible enough for dependence modeling when the dimension exceeds 5, due to its exchangeable feature. The extension of exchangeable Archimedean copulas under the proposed framework has some better properties than the LT-Archimedean factor copula of [30,23]. For modeling dependence among many variables, our 1-factor parsimonious copula based on comonotonic latent variables is a reasonable first order model that provides more flexibility. Our construction can also extend to multiple factors, including structured factor models, as well as a version of the Gaussian oblique factor model when variables are divided into non-overlapping groups.

The essential idea of the proposed comonotonic latent variable model is the following. For a d-dimensional random vector $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$, we introduce $p \in \{1, \ldots, d\}$ sets of comonotonic latent variables. The latent variables are comonotonic within each set, while the dependence structure between the sets of comonotonic latent variables can be either independent or dependent. First, decompose each component Y_i of $\boldsymbol{Y}$ into several components, and then let some components belong to different sets of the p sets of latent variables. For such a model, roughly speaking, the dimension of the complexity of the question can be effectively reduced from d to p, the number of sets of comonotonic latent variables. In what follows, we refer to models that have such a structure as *comonotonic factor models*.

The idea was motivated by the fact that many dependence structures in statistical modeling are based on a scale mixture representation $\boldsymbol{Y} = V\boldsymbol{H}$, where the dependence is introduced by the random vector $\boldsymbol{Y}$, V is the scale random variable, $\boldsymbol{H}$ is a vector of random elements that can either be independent or dependent, and V and $\boldsymbol{H}$ are independent. For example, elliptical distributions are based on scale mixtures with a linear transformation of a random vector that is uniformly distributed on the surface of a hypersphere; Archimedean copulas constructed by Laplace Transforms (LTs) of positive random variables can be derived from a stochastic representation of scale mixtures of independent unit Fréchet distributions. For the scale mixture models, $(V, \ldots, V)^\top$ can be simply treated as a trivial comonotonic random vector, and it motivates us to employ a more general comonotonic random vector to replace the scale random variable V.

In this paper, we focus on a subclass of the comonotonic factor models, and assume that for each $i \in \{1, \ldots, d\}$, the univariate marginal Y_i can be written as

$$Y_i = V_i H_i, \tag{1}$$

where $V_i \perp H_i$, V_i is a positive random variable, and H_i follows the standard Fréchet distribution with cumulative distribution function (cdf) $F(x) = e^{-1/x}$, $x > 0$. Such a representation leads to a tractable univariate marginal distribution for the Y_i's. We are interested in copula structures that can be retrieved from $\boldsymbol{Y}$, and the dependence between the Y_i's is introduced through the dependence between the V_i's and the dependence between the H_i's. We will begin with the easy case of which the V_i's are comonotonic, and the H_i's are mutually independent. Then, multiple sets of comonotonic latent variables are introduced to generate more structured dependence patterns among the Y_i's, where decompositions of either V_i or H_i is required. For example, one can write $Y_i = (V_{i1}+V_{i2})H_i$ for each $i \in \{1, \ldots, d\}$, and let the V_{i1}'s be comonotonic and the V_{i2}'s be comonotonic, respectively, while the random vectors $(V_{11}, \ldots, V_{1d})^\top$ and $(V_{21}, \ldots, V_{2d})^\top$ are independent. In Section 3, we will discuss several different types of decompositions, and derive the corresponding dependence patterns for the resulting $\boldsymbol{Y}$.

The remainder of the paper is organized as follows. In Section 2, we first introduce the notation system used in the paper, and then the notion of comonotonicity. In Section 3, comonotonic factor models are studied, along with discussions on a variety of approaches of constructing parametric comonotonic factor models; numerical methods for maximum likelihood estimation are discussed accordingly. In Section 4, we focus on comonotonic bi-factor models that account for both within-group and between-group dependence. Applications of the copulas of comonotonic factor models for flexible dependence modeling of a real dataset are reported in Section 5, where response variables are naturally divided into several dependence clusters, of which dependence can be intuitively explained by latent variables. Section 6 concludes the paper with further discussions.

2. Preliminaries

In this section, we will first introduce the notation system, and then the basic concept of comonotonicity and its relevant properties for the proposed model.

2.1. Notation

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_d)^\top$ be a random vector with a joint cdf F and continuous univariate marginal distribution functions $F_1, \ldots, F_d$. Then by the Sklar's theorem [31], F can be written as $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$, where the copula function $C : [0, 1]^d \to [0, 1]$ is uniquely determined by

$$C(u_1, \ldots, u_d) = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \tag{2}$$

with F_i^{-1} standing for the inverse function of F_i for each $i \in \{1, \ldots, d\}$.

A bold symbol represents a matrix or a vector, and a vector always denotes a column vector unless otherwise specified. For a multivariate distribution, we often transform each univariate marginal to be standard Normal distributed to compare