Electronic Journal of Statistics
Vol. 14 (2020) 4457–4488
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1784

A Bayesian hierarchical copula model

Haoxin Zhuang and Liqun Diao*

Department of Statistics and Actuarial Science
University of Waterloo
200 University Avenue West, Waterloo, ON, Canada
e-mail: h9zhuang@uwaterloo.ca; l2diao@uwaterloo.ca

Grace Y. Yi

Department of Statistical and Actuarial Sciences
Department of Computer Science
University of Western Ontario
1151 Richmond Street, London, ON, Canada
e-mail: gyi5@uwo.ca

Abstract: Dependent data with hierarchical structures arises commonly from a variety of application, and analysis of such data is often challenging due to the complexity in modeling dependence structures and the computation intensity. In this paper, we propose a Bayesian hierarchical copula model (BHCM) to accommodate hierarchical structures of dependent data, where the subject-level dependence is modeled by the copula-based model and the hierarchical structure is described using random dependence parameters. We introduce a layer-by-layer sampling scheme for conducting Bayesian inferences. Our proposed BHCM enjoys the flexibility of modeling various complex association structures, while retaining manageable computation. Extensive simulation studies show that our proposed estimators outperform conventional likelihood-based estimators in a variety of finite sample settings. We apply the BHCM to analyze the Vertebral Column dataset arising from UCI Machine Learning Repository.

Keywords and phrases: Bayesian hierarchical model, copula, dependence modeling, MCMC.

Received July 2020.

Contents

*Corresponding Author.

1. Introduction

Data with complex structures arises commonly in modern scientific research. Examples include data with a hierarchical nesting structure, data collected from different research centers, and data configured at multiple locations or multiple time points, etc. With the development of Markov Chain Monte Carlo (MCMC) algorithms [13, 20] and the rapidly increasing computation power, Bayesian hierarchical models have become useful in multiple areas including environment [25], genetics [5], machine learning [11, 18], etc. A large body of the literature on Bayesian hierarchical models has been available; see [7], for example, for a comprehensive introduction and discussion.

Using the Bayesian hierarchical framework, conventional single-level models can be generalized to accommodate complex hierarchical structures of data by allowing randomness and a hierarchical structure of the parameters. For instance, the multilevel model [46], also known as the hierarchical linear model, is an extension of the general linear regression model to facilitate the hierarchical nesting structure of data. Another example is the time-varying autoregressive model [41, 32, 37], which is a hierarchical generalization of the autoregressive model and allows for random coefficients. Our interest in the paper is, while using the same framework, to study dependence modeling using copula models.

Copula [45] is a useful tool to model the dependence between multivariate random variables, allowing separate modeling of the marginal distributions and the dependence structure. A comprehensive introduction to copula can be found in [27, 29] and [36]. Several methods have been employed for estimation of the copula parameters. The maximum likelihood (ML) method [27, 9, 50] is the most commonly-used method, though it requires a large computational resource in the presence of a large number of parameters. A computationally friendly but less efficient alternative is the inference functions for margin (IFM)

method proposed by [30] with the asymptotic properties studied by [28]. Another estimation method, the ranked-based method [27], is used to estimate a copula parameter utilizing its relationship with Kendall's τ. Such a method is restrictive to the single-parameter copula function, which allows for an explicit expression of the relationship between the dependence parameter and Kendall's τ.

[48] provided a summary of applications of Bayesian methods in copula. Bayesian approaches are mainly used for three objectives. First, they offer an alternative to the likelihood-based and rank-based methods we mentioned previously. For instance, [35] used a Bayesian method to estimate copula forms for D-Vine. Secondly, Bayesian approaches can be used for model selection. For instance, [47] used a D-Vine copula to model longitudinal data and proposed a Bayesian approach for estimating parameters and identifying independent bivariate pairs in the vine structure. [21] and [22] discussed a sequential method and a simultaneous method, respectively, for selecting copula forms in a regular vine structure using the reversible jump MCMC. Thirdly, the Bayesian hierarchical model can be used to accommodate the covariate information. For example, Chapter 4 in [19] modeled spatial dependence by a multivariate copula function. A Bayesian hierarchical model was proposed to relate parameters in the marginal models to the covariates and the copula parameters were assumed to have noninformative priors.

While various methods are available in the literature, studies of copula models under the Bayesian framework are relatively limited. Moreover, the use of available methods is often hindered by its complexity in modeling and the intensity of implementation. To circumvent these issues, we propose a Bayesian hierarchical copula model (BHCM) for which we are particularly interested in the scenario with hierarchical structured data illustrated by Figure 1. The nodes at the data level represent subjects and those at the intermediate level represent clusters which form the top node at the population level. Data in possession of this hierarchical structure arises commonly in practice. Examples include multicenter medical studies conducted at m sites, meta-analyses of m studies, spatially configured data of m locations, longitudinal data from m subjects, time series with time varying dependence structures of m periods, etc., where m is a finite positive integer.

To account for a more complex hierarchical structure, the three-level structure can be easily extended by including more intermediate levels. Suppose that multivariate data are collected from each subject and the dependence modeling of the subject-level multivariate structure is of interest. To model the subject-level dependence, we propose a Bayesian hierarchical copula model (BHCM); such a model accounts for the hierarchical structure by allowing random dependence parameters and specifying multiple layers of prior and hyperprior distributions. This model unifies the ideas of the Bayesian hierarchical approach and the copula-based dependence modeling, and offers us the flexibility in facilitating various association structures, while allowing a straightforward implementation of inference procedures.

The rest of the paper is organized as follows. In Section 2, we describe the model formulation of the proposed BHCM. In Section 3, we examine issues concerning inferences and the sampling schemes. In Section 4, we discuss the selection of transformation functions and associated scaling parameters. In Section 5, we perform simulation studies to evaluate the finite sample performance of the proposed methods. In Section 6, we analyze the Vertebral Column Data [10] using the proposed BHCM. The paper ends with a discussion in Section 7.

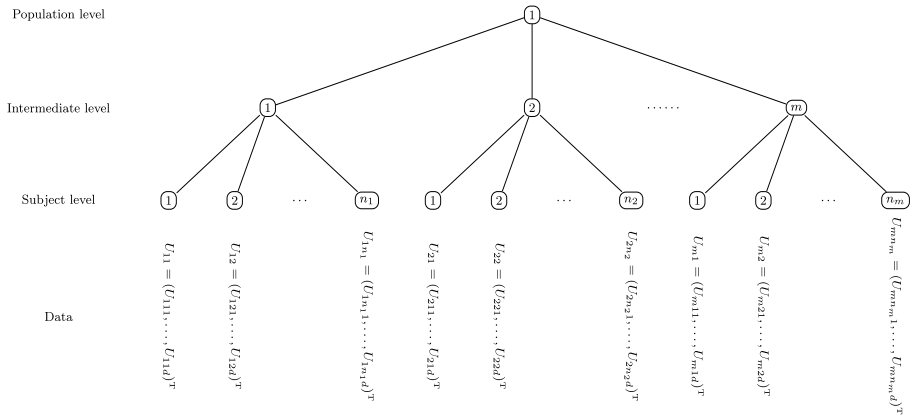


FIG 1. *A three-level hierarchical structure*

2. Model formulation

We consider a three-level hierarchical structure as illustrated in Figure 1. The single node at the top level represents the population level. The bottom level is the subject level in which each node corresponds to the data from a subject. The intermediate level contains m clusters to which the bottom-level subjects belong. Let $U_{ji} = (U_{ji1}, \ldots, U_{jid})^{\mathrm{T}}$ be the vector of d features for the ith subject in the jth cluster, where $i = 1, \ldots, n_j$, $j = 1, \ldots, m$, and n_j is a positive integer that may depend on j. Let $U_j = (U_{j1}^{\mathrm{T}}, \ldots, U_{jn_j}^{\mathrm{T}})^{\mathrm{T}}$ and $U = (U_1^{\mathrm{T}}, \ldots, U_m^{\mathrm{T}})^{\mathrm{T}}$. Let u_{jik}, u_{ji}, u_j and u represent the observed counterparts of U_{jik}, U_{ji}, U_j and U, respectively, for $i = 1, \ldots, n_j$, $j = 1, \ldots, m$, and $k = 1, \ldots, d$.

The copula formulation is advantageous in its separation of modeling marginal distributions and dependence structures, and much attention has been directed to modeling the dependence structures with a standard treatment of marginal distributions. Consistent with many authors [e.g., 1, 38, 39], we assume that U_{jik} follows a uniform distribution on $[0, 1]$ marginally and focus on dependence modeling of the subject-level data U_{ji} using copula-based models. In Section 2.1, we first use a copula-based approach to model the dependence structure among the d features of each subject and allow different structures for different clusters. In Section 2.2, we account for the hierarchical structure and continue our discussion in the framework of Bayesian hierarchical models.

2.1. Copula-based dependence models

According to [45], any joint cumulative distribution function (CDF) can be written as a copula function of its univariate marginal CDFs. A copula function on $[0,1]^d$, denoted by C, is defined as $C(u_1,\ldots,u_d) = P(U_1 \le u_1,\ldots,U_d \le u_d)$, for uniformly distributed random variables $U_1,\ldots,U_d$ on [0,1]. If the marginal distributions are all continuous, the copula C always exists and is unique. Here we assume that the joint distribution of d features in cluster j is governed by a multivariate copula function $C_j(u_{ji1},\ldots,u_{jid};\theta_j)$, for $i = 1,\ldots,n_j$, where $\theta_j = (\theta_{j1},\ldots,\theta_{jp_j})^{\mathrm{T}}$ is the vector of parameters indexing the copula function C_j, p_j is the number of parameters, and $j = 1,\ldots,m$. Let $\theta = (\theta_1^{\mathrm{T}},\ldots,\theta_m^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of all copula parameters. Common choices of multivariate copula C_j include multivariate Gaussian copula and multivariate t-copula from the elliptical copula family [12], and multivariate Clayton, Frank and Gumbel copulas from the Archimedean copula family [16, 17]. Copula functions in the Archimedean family contain only one parameter, while those in the elliptical family may contain multiple parameters. For $j = 1,\ldots,m$, let c_j denote the density function derived from C_j.

2.2. Bayesian hierarchical models

We construct a Bayesian hierarchical model to account for the 3-level hierarchical structure, illustrated in Figure 1, through the following 3-stage specifications of prior and hyperprior distributions [23, 33]. The first stage of the hierarchical model facilitates the vector $U_{ji} = (U_{ji1},\ldots,U_{jid})^{\mathrm{T}}$ by a copula-based dependence model as described in Section 2.1, where θ_j is of dimension p_j. As we allow the dependence structures to be distinct and governed by different copula functions across clusters, the association parameters θ_j may have different ranges for $j = 1,\ldots,m$. Before we specify a prior distribution for θ_j, we map each component θ_{jl} of θ_j to $\mathbb{R}$ or an interval on $\mathbb{R}$ through a proper transformation. When an explicit expression of Kendall's τ is available, a natural way of reparameterizing the parameters θ_{jl} is to invoke the Kendall's τ, together with the Fisher z-transformation [42]. However, transformations based on Kendall's τ are not always feasible, especially for copulas indexed with two or more parameters. For example, for a bivariate t-copula indexed by a correlation coefficient and a degree of freedom, the Kendall's τ relates to the correlation coefficient but has nothing to do with the degree of freedom. On the other hand, for bivariate copulas in the BB family [27, 29], which are indexed by two parameters, the Kendall's τ can be written as a function of the two parameters but it cannot be in one-to-one relation with the two copula parameters. Therefore, we cannot make inference about the copula parameters by modeling Kendall's τ and then transforming it back to the copula parameters. In the development here, we take an alternative by writing $\gamma_{jl} = \alpha_{jl} g_{jl}(\theta_{jl})$ for $l = 1,\ldots,p_j$ and $j = 1,\ldots,m$, where α_{jl} is a non-zero scaling parameter, whose inclusion allows us to quantify different variabilities across clusters, and $g(\cdot)$ is a monotonic function mapping

the parameter space, $\mathcal{A}$, of the dependence parameter θ, to $\mathbb{R}$ or an interval on $\mathbb{R}$.

The form of the transformation functions and the rationale behind rescaling are discussed in details in Section 4. Let $\gamma_j = (\gamma_{j1}, \ldots, \gamma_{jp_j})^{\mathrm{T}}$ denote the vector of transformed and scaled dependence parameters in cluster j, and let $\gamma = (\gamma_1^{\mathrm{T}}, \ldots, \gamma_m^{\mathrm{T}})^{\mathrm{T}}$.

At the second stage of the hierarchical model, we specify the prior distribution for the parameters γ_{jl} as

$$\gamma_{jl}|(\mu_{jl}, \sigma_{jl}) \sim N(\mu_{jl}, \sigma_{jl}^2), \tag{2.1}$$

where μ_{jl} and σ_{jl} indicate the cluster location and variability of γ_{jl}, respectively, for $l = 1, \ldots, p_j$ and $j = 1, \ldots, m$. Let $\mu_j = (\mu_{j1}, \ldots, \mu_{jp_j})^{\mathrm{T}}$ be the vector of mean parameters, let $\sigma_j = (\sigma_{j1}, \ldots, \sigma_{jp_j})^{\mathrm{T}}$ be the vector of standard deviations (s.d.) of the jth cluster, and let $\mu = (\mu_1^{\mathrm{T}}, \ldots, \mu_m^{\mathrm{T}})^{\mathrm{T}}$ and $\sigma = (\sigma_1^{\mathrm{T}}, \ldots, \sigma_m^{\mathrm{T}})^{\mathrm{T}}$. We further specify the prior distributions for cluster-level location parameters μ_{jl} as

$$\mu_{jl}|(\varphi_l, \delta_l) \sim N(\varphi_l, \delta_l^2), \tag{2.2}$$

and the hyperprior distributions for the cluster-level variability parameters σ_{jl} as

$$\sigma_{jl} \sim \pi_\sigma,$$

for $l = 1, \ldots, p_j$ and $j = 1, \ldots, m$, where φ_l and δ_l indicate the population location and variability of μ_{jl}, respectively, and π_σ is the prior distribution of σ_{jl}.

Let $\varphi = (\varphi_1, \ldots, \varphi_{p^*})^{\mathrm{T}}$ and $\delta = (\delta_1, \ldots, \delta_{p^*})^{\mathrm{T}}$, where $p^* = \max(p_1, \ldots, p_m)$. This stage characterizes the cluster-level parameters, which corresponds to the intermediate level of the hierarchical structure in Figure 1.

At the third stage, we specify the hyperprior distribution for the population-level parameters φ and δ as

$$\varphi_l \sim \pi_\varphi \quad \text{and} \quad \delta_l \sim \pi_\delta, \tag{2.3}$$

for $l = 1, \ldots, p^*$, where π_φ and π_δ are prior distributions for φ_l and δ_l, respectively.

Combining (2.1) and (2.2) gives

$$\gamma_{jl}|(\varphi_l, \delta_l, \sigma_{jl}) \sim N(\varphi_l, \sigma_{jl}^2 + \delta_l^2), \tag{2.4}$$

for $l = 1, \ldots, p_j$ and $j = 1, \ldots, m$, where the variance of γ_{jl} includes the within-cluster variability σ_{jl}^2 and between-cluster variability δ_l^2.

For the parameters $(\varphi^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ at the population level and σ_j at the cluster level, we select a weak-informative prior, such as an inverse Gamma distribution, Inverse Gamma$(\varepsilon, \varepsilon)$, with a small ε, or a non-informative prior (e.g., an improper uniform prior [26]). For the construction of the Bayesian hierarchical model, we assume exchangeability for all levels of specification.

3. Bayesian inference

Here we aim to make Bayesian inferences about the vector of the dependence parameters $\theta = (\theta_1^{\mathrm{T}}, \ldots, \theta_m^{\mathrm{T}})^{\mathrm{T}}$. Since we have worked with the transformed and scaled dependence parameters γ in Section 2.2, we will continue our discussion in terms of γ and transform them back to their original scale θ. We first consider the posterior distribution of γ

$$f(\gamma|u) \propto f(u|\gamma)f(\gamma),$$

where $f(u|\gamma)$ stands for the copula density function with the data u and the transformed parameters γ specified as in Section 2.1, and $f(\gamma)$ is the prior distribution of γ, given in Section 2.2. The distribution of $f(\gamma)$ can be obtained by integrating the joint distribution of $f(\gamma, \sigma, \varphi, \delta)$ with respect to σ, φ and δ, where $f(\gamma, \sigma, \varphi, \delta)$ is determined by $f(\gamma|\sigma, \varphi, \delta)\pi(\sigma)\pi(\varphi)\pi(\delta)$. This calculation involves the integrals of dimension $\sum_{j=1}^{m} p_j + 2p^*$, which is generally difficult to implement. To overcome this difficulty, we employ an alternative strategy by sampling from the joint posterior distribution $f(\gamma, \sigma, \varphi, \delta|u)$. The posterior distributions that are used in the sampling algorithm is provided in Section 3.1 and the sampling algorithm is introduced in Section 3.2.

3.1. Posterior distributions

We start with the joint posterior distribution of $(\gamma^{\mathrm{T}}, \sigma^{\mathrm{T}}, \varphi^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$,

$$\begin{aligned} f(\gamma, \sigma, \varphi, \delta|u) &\propto f(u|\gamma)f(\gamma|\sigma, \varphi, \delta)\pi(\sigma)\pi(\varphi)\pi(\delta) \\ &= \prod_{j=1}^{m}\left[\prod_{i=1}^{n_j} c_j(u_{ji}; \gamma_j)\prod_{l=1}^{p_j}\phi(\gamma_{jl}|\varphi_l, \sigma_{jl}^2 + \delta_l^2)\right]\pi_\sigma\pi_\varphi\pi_\delta, \end{aligned} \tag{3.1}$$

where $\phi(\cdot|a, b^2)$ is the density function of the normal distribution with mean a and variance b^2.

The joint posterior distribution of $(\varphi^{\mathrm{T}}, \delta^{\mathrm{T}}, \sigma^{\mathrm{T}})^{\mathrm{T}}$ can be obtained by integrating (3.1) with respect to γ,

$$\begin{aligned} f(\sigma, \varphi, \delta|u) &= \int f(\gamma, \sigma, \varphi, \delta|u)d\gamma \\ &= \prod_{j=1}^{m}\int\prod_{i=1}^{n_j} c_j(u_{ji}; \gamma_j)\prod_{l=1}^{p_j}\phi(\gamma_{jl}|\varphi_l, \sigma_{jl}^2 + \delta_l^2)\pi_\sigma\pi_\varphi\pi_\delta d\gamma_j. \end{aligned} \tag{3.2}$$

As the integrand in (3.2) involves the product of various copula functions, the resulting integral generally has no analytical form, and thus a numerical algorithm, such as the Monte Carlo approach, is needed to numerically approximate the posterior density. Details on the implementation of Monte Carlo method is deferred to Section 5.1.

Finally, the conditional posterior distribution of parameters γ_j, given all the hyperprior parameters and $\gamma_{(-j)} = (\gamma_1^{\mathrm{T}}, \ldots, \gamma_{j-1}^{\mathrm{T}}, \gamma_{j+1}^{\mathrm{T}}, \ldots, \gamma_m^{\mathrm{T}})$, is of the form

$$\begin{aligned} f(\gamma_j|\sigma, \varphi, \delta, \gamma_{(-j)}, u) &= f(\gamma_j|\sigma, \varphi, \delta, u) \\ &\propto f(u_j|\gamma_j) f(\gamma_j|\varphi, \delta, \sigma) \\ &= \prod_{i=1}^{n_j} c_j(u_{ji}; \gamma_j) \prod_{l=1}^{p_j} \phi(\gamma_{jl}|\varphi_l, \sigma_{jl}^2 + \delta_l^2), \end{aligned} \tag{3.3}$$

where the first equality comes from that given $(\varphi^{\mathrm{T}}, \delta^{\mathrm{T}}, \sigma^{\mathrm{T}})^{\mathrm{T}}$, γ_j is independent of $\gamma_{(-j)}$.

3.2. Sampling scheme and Markov Chain Monte Carlo

To utilize the joint posterior distribution $f(\gamma, \sigma, \varphi, \delta|u)$ in (3.1), we let $\zeta = (\gamma^{\mathrm{T}}, \sigma^{\mathrm{T}}, \varphi^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ denote the vector of all the parameters. The Metropolis-Hasting (M-H) algorithm [34, 24] can be employed, in principle, to sample from $f(\zeta|u)$ directly. In the instance with a high dimensional ζ, directly applying the M-H algorithm to the joint posterior distribution (3.1) is challenging because it is not always straightforward to choose an appropriate proposal density function and to tune the parameters in the proposal density to get a good acceptance rate, and therefore the M-H can be inefficient or not even converge. It is not feasible to directly invoke the Gibbs sampler [15, 14] to sample from the joint posterior distribution $f(\gamma, \sigma, \varphi, \delta|u)$ in (3.1) by iteratively sampling from the conditional distributions $f(\sigma, \varphi, \delta|\gamma^{(t-1)}, u)$ and $f(\gamma|\sigma^{(t)}, \varphi^{(t)}, \delta^{(t)}, u)$, where $\gamma^{(t)}$, $\sigma^{(t)}$, $\varphi^{(t)}$ and $\delta^{(t)}$ represent the parameter samples in the tth iteration, since the conditional distribution of the hyper-parameters does not depend on the data,

$$f(\sigma, \varphi, \delta|\gamma^{(t-1)}, u) \propto f(\gamma^{(t-1)}|\sigma, \varphi, \delta)\pi_\sigma \pi_\varphi \pi_\delta.$$

To cope with the issue, we consider the following "layer by layer" sampling procedure.

1. Sample hyperprior parameters $(\sigma^{\mathrm{T}}, \varphi^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ from the posterior distribution $f(\sigma, \varphi, \delta|u)$ in (3.2) using the M-H algorithm.
2. Calculate the sample means of the sampled vectors in Step 1 and let them be the Bayesian estimates for σ, φ, and δ, denoted by $\hat{\sigma}$, $\hat{\varphi}$, and $\hat{\delta}$, respectively.
3. Sample parameters γ_j from the conditional posterior distribution $f(\gamma_j|\hat{\sigma}, \hat{\varphi}, \hat{\delta}, u)$ in (3.3) with the Bayesian estimates of the hyperprior parameters obtained from Step 2 plugged in. Apply the M-H algorithm to $f(\gamma_j|\hat{\sigma}, \hat{\varphi}, \hat{\delta}, u)$ and repeat this step for $j = 1, \ldots, m$.
4. Transform $\gamma_{jl}^{(t)}$ back to obtain $\theta_{jl}^{(t)}$ through a division by α_{jl} and the inverse transformation function $g_{jl}^{-1}(\cdot)$, for $l = 1, \ldots, p_j$, $j = 1, \ldots, m$, and $t = 1, \ldots, N$.
5. Compute the quantities of interest that are related to the parameters θ_{jl}, such as the posterior mean.

In Steps 1 and 3, we apply the random walk Metropolis algorithm, where the proposal distribution is a multivariate normal distribution with the mean vector determined by the sampled values from the previous iteration of the M-H algorithm and the covariance matrix specified as a diagonal matrix with diagonal elements tuned to achieve an acceptance rate for convergence of the sampling algorithm. Besides normal distributions, other distributions can also be considered as proposal distributions. For variance parameters, σ and δ, a truncated normal or a Gamma distributions can be good options as well [14]. If a range $[a, b]$ of each parameter can be determined beforehand, a truncated normal proposal can stabilize performance of the sampling procedure when the dependence is extremely strong. [42] and [43] contain some guidelines on determining the ranges for copula parameters.

In situations where the dimension of the parameters $(\sigma^{\mathrm{T}}, \varphi^{\mathrm{T}}, \delta^{\mathrm{T}})^{\mathrm{T}}$ is high and/or the convergence of the sampling algorithm is a concern, one may adopt a Gibbs Sampler [15, 14] in Step 1 and further decompose the joint posterior distribution (3.2) at the tth iteration as

$$\begin{aligned} &f(\sigma_j|\sigma_{(-j)}^{(t-1)}, \varphi^{(t-1)}, \delta^{(t-1)}, u) = f(\sigma_j|\varphi^{(t-1)}, \delta^{(t-1)}, u) \\ &\propto \int \prod_{i=1}^{n_j} c_j(u_{ji}; \gamma_j) \prod_{l=1}^{p_j} \phi(\gamma_{jl}|\varphi_l^{(t-1)}, \sigma_{jl}^2 + (\delta_l^{(t-1)})^2) \pi_{\sigma_j} \pi_\varphi \pi_\delta d\gamma_j \\ &f(\varphi|\sigma^{(t)}, \delta^{(t-1)}, u) \propto f(\sigma^{(t)}, \varphi, \delta^{(t-1)}|u) \\ &f(\delta|\sigma^{(t)}, \varphi^{(t-1)}, u) \propto f(\sigma^{(t)}, \varphi^{(t-1)}, \delta|u), \end{aligned} \tag{3.4}$$

where $\sigma_{(-j)} = (\sigma_1^{\mathrm{T}}, \ldots, \sigma_{j-1}^{\mathrm{T}}, \sigma_{j+1}^{\mathrm{T}}, \ldots, \sigma_m^{\mathrm{T}})^{\mathrm{T}}$, for $j = 1, \ldots, m$. Instead of sampling from the joint posterior (3.2), sampling from each of the conditional distributions in (3.4) improves the sampling efficiency in the sense that it facilitates a lower rejection rate yet a larger effective sample size. This gain is at the price of increasing the computation time which is basically caused by the calculation of the integration over γ.

While a large dimension of γ can considerably increase the computation time of the sampling procedure, Step 3 of the sampling procedure does not require an appreciable computation time, as the sampling from (3.3) is conducted within each cluster j which does not involve any integration. Although most of the computation time is required by Step 1 for the case with a large number of parameters, applications of our sampling algorithm are still feasible, because the most frequently-used copulas from the Archimedean and extreme-value families contain one or two parameters; even for copulas from the elliptical family, such as the Gaussian copula, which contain a high dimension of parameters, it is often common to impose certain correlation structures to the copula to facilitate a parsimonious model.

The evaluation of the posterior density distribution in (3.2) involves the integrals which generally do not have an analytically close form. To handle this issue, we suggest to use the random walk Metropolis algorithm [20] instead of the MCMC algorithms which require the gradient of the posterior distribution, such as Langevin MCMC or Hamiltonian MC [40].

Finally, we comment that standard Bayesian theory (e.g., [31], [8], [44] or [14]) can be adapted to establish the asymptotic results, such as the consistency and asymptotic normality, for the posterior distribution, provided regularity conditions.

4. Transformation of the dependence parameters

4.1. Transformation function

In this subsection, we discuss the selection of the transformation function $g(\cdot)$, which is a monotonic function mapping $\mathcal{A}$ to $\mathbb{R}$ or an interval on $\mathbb{R}$, where $\mathcal{A}$ is the parameter space for the dependence parameter θ. In Table 1, we give examples of transformation functions for some commonly-used copula functions, where L and U are the lower and upper bounds of $\mathcal{A}$, respectively.

TABLE 1
Transformation Functions for Copula Parameters

$\mathcal{A}$	Example of Copula Function	Tranformation Function
$[L, U]$	Gaussian Copula	$g(x) = \log\left(\frac{x-L}{U-x}\right)$
$[L, \infty)$	Clayton Copula	$g(x) = \log(x - L)$
$(-\infty, U]$	Rotated Clayton Copula	$g(x) = \log(U - x)$
$(-\infty, \infty) \setminus \{0\}$	Frank Copula	$g(x) = x$

For copula functions with an infinite range of parameters, we can impose a certain finite range $[L^*, U^*]$ and use the transformation function $g(x) = \log\left(\frac{x-L^*}{U^*-x}\right)$. For example, for the Frank copula, we may impose the range $[-100, 100]$ to cover the Kendall's τ from -0.96 to 0.96. In the simulation section, we compare the identity transformation function and the logit transformation function with end points as $[-100, 100]$ for the Frank copula.

4.2. Choice of scaling parameter

In this subsection, we discuss the choice of scaling parameter α_{jl}. First, we define $\gamma^*_{jl} = g_{jl}(\theta_{jl})$ as the dependence parameter mapped into $\mathbb{R}$ or an interval on $\mathbb{R}$ without scaling and write $\gamma^* = (\gamma^*_{j1}, \ldots, \gamma^*_{jp_j})^{\mathrm{T}}$. Then the scaled and unscaled parameters have the relationship $\gamma_{jl} = \alpha_{jl}\gamma^*_{jl}$, for $l = 1, \ldots, p_j$ and $j = 1, \ldots, m$.

We impose a normal prior on γ_{jl} in Section 2.2 in the form of

$$\gamma_{jl} \sim N(\mu_{jl}, \sigma^2_{jl}),$$

and further impose a normal prior on the cluster mean μ_{jl} as

$$\mu_{jl} \sim N(\varphi_l, \delta^2_l),$$

which is equivalent to imposing a normal prior on γ^*_{jl} of the form

$$\gamma^*_{jl} \sim N\left(\frac{\mu_{jl}}{\alpha_{jl}}, \frac{\sigma^2_{jl}}{\alpha^2_{jl}}\right),$$

together with the prior distribution for the cluster mean

$$\frac{\mu_{jl}}{\alpha_{jl}} \sim N\left(\frac{\varphi_l}{\alpha_{jl}}, \frac{\delta_l^2}{\alpha_{jl}^2}\right).$$

As $|\alpha_{jl}|$ gets larger, both the within-cluster and between-cluster variances assumed in the prior distributions become smaller. In other words, as $|\alpha_{jl}|$ increases, we impose a stronger prior on γ_{jl}^*.

Next, we describe a method of choosing suitable values of the α_{jl}. Suppose that we obtain the maximum likelihood estimate (MLE) of γ_j^*, denoted $\tilde{\gamma}_j^*$, by maximizing the likelihood function

$$L(\gamma_j^*|u_j) = \prod_{i=1}^{n_j} c_j(u_{ji}|\gamma_j^*).$$

The asymptotic covariance matrix of $\tilde{\gamma}_j^*$ can be estimated by $I^{-1}(\tilde{\gamma}_j^*)$, where $I(\tilde{\gamma}_j^*)$ is the observed information matrix

$$I(\tilde{\gamma}_j^*) = -\frac{\partial^2 \log L(\gamma_j^*|u_j)}{\partial \gamma_j^* \partial \gamma_j^{*\mathrm{T}}}\bigg|_{\gamma_j^*=\tilde{\gamma}_j^*}.$$

Let $\widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)$ denote the estimated asymptotic standard deviation of $\tilde{\gamma}_{jl}^*$, which is calculated as the square root of the lth diagonal element of $I^{-1}(\tilde{\gamma}_j^*)$. By the invariance property of MLE, the MLE of $\gamma_{jl} = \alpha_{jl}\gamma_{jl}^*$, denoted $\tilde{\gamma}_{jl}$, is $\alpha_{jl}\tilde{\gamma}_{jl}^*$, and its estimated asymptotic s.d. is $\widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}) = |\alpha_{jl}|\,\widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)$. We aim to choose the α_{jl} such that the resultant 95% confidence intervals of the $\tilde{\gamma}_{jl}$ are of the same length, say, L, for all $l = 1, \ldots, p_j$ and $j = 1, \ldots, m$, where $L = 2 \times 1.96 \times \widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}) = 2 \times 1.96 \times |\alpha_{jl}| \times \widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)$. Therefore, we set

$$\alpha_{jl} = \frac{L}{3.92 \times \widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)} \times \mathrm{sign}(\tilde{\gamma}_{jl}^*),$$

which has the same sign as $\tilde{\gamma}_{jl}^*$; α_{jl} is the ratio of the target width of a 95% confidence interval of $\tilde{\gamma}_{jl}$ to the width of the 95% confidence interval of $\tilde{\gamma}_{jl}^*$. Consequently, the within-cluster mean can be approximated by

$$\tilde{\gamma}_{jl} = \alpha_{jl}\tilde{\gamma}_{jl}^* = \mathrm{sign}(\tilde{\gamma}_{jl}^*) \times \frac{L}{3.92 \times \widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)} \times \tilde{\gamma}_{jl}^*,$$

and the within-cluster s.d. can be approximated by

$$\widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}) = |\alpha_{jl}|\,\widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*) = \frac{L}{3.92}, \tag{4.1}$$

a constant value shared by all clusters. The population mean can be approximated by

$$\bar{\gamma}_l := \frac{1}{m}\sum_{j=1}^{m} \tilde{\gamma}_{jl} = \sum_{j=1}^{m} \mathrm{sign}(\tilde{\gamma}_{jl}^*) \times \frac{L}{m \times 3.92 \times \widehat{\mathrm{sd}}(\tilde{\gamma}_{jl}^*)} \times \tilde{\gamma}_{jl}^*, \tag{4.2}$$

and the between-cluster s.d. can be approximated by

$$\frac{1}{m-1}\sum_{j=1}^{m}(\tilde{\gamma}_{jl}-\bar{\gamma}_l)^2=\frac{1}{m-1}\left[\sum_{j=1}^{m}\alpha_{jl}^2(\tilde{\gamma}_{jl}^*)^2-m\bar{\gamma}_{jl}^2\right]. \tag{4.3}$$

Scaling the transformed dependence parameters has the following effects. First, it standardizes how much the observations of the subjects within the same cluster vary from the cluster mean. As we derive in (4.1), all clusters share the same within-cluster s.d.. Secondly, the population mean in (4.2) can be viewed as a weighted average of the unscaled γ_{jl}^*'s. If a cluster has a larger within-cluster variability in terms of γ_{jl}^*, a smaller weight is often assigned to this cluster (see Appendix A for a detailed discussion). Therefore, the population mean will be less affected by the clusters with large variabilities and then becomes more stable. The same argument applies to the calculation of between-cluster variance in (4.3). Thirdly, the term $\text{sign}(\tilde{\gamma}_{jl}^*)$ in α_{jl} ensures the positivity of the estimates of scaled parameters, which reduces the between-cluster variability. Based on the simulation results in Section 5, we suggest to use $L=4$ as "a rule of thumb" to avoid an overwhelmingly strong or weak prior distribution.

5. Simulation studies

In this section, we conduct simulation studies to examine the finite sample performance of the Bayesian estimators of the dependence parameter θ under the proposed BHCM; the examination is taken in contrast to the performance of the likelihood-based estimators, which are used conventionally for estimating the parameters of copula models. Though the interpretation for the Bayesian and the likelihood estimators is not the same, such comparisons can shed lights on the performance of our proposed BHCM, because with non-informative priors for the parameters θ, the Bayesian estimators would be numerically close to the likelihood estimators.

5.1. Simulation settings

We consider a three-level hierarchical structure with $m=4$ clusters at the intermediate level, and the sample size is taken as $n=200$ or 400. A vine copula structure [3, 1] is utilized to simulate dependent hierarchical data. While various dependence structures can be obtained by choosing different types of vines, changing the order of the nodes in the vine structure, and adopting different bivariate copulas on different levels of the vine structure, here we generate data from a D-Vine copula structure as illustrated in Figure 2, where the bivariate copulas in the vine structure higher than level 1 are all assumed to be independent. In Figure 2, the dependence strength between U_{ji1} and U_{ji2} is of interest. The bivariate copula between U_{1i2} and U_{2i1} is the connecting structure between clusters 1 and 2. Similarly, $C(u_{2i2},u_{3i1})$ connects clusters 2 and 3, and $C(u_{3i2},u_{4i1})$ connects clusters 3 and 4.

We consider five simulation settings. The copula forms and the parameter values are summarized in Table 2. Settings 1 and 2 have the same copula forms for different clusters, and Settings 3, 4 and 5 allow different dependence structures. In Settings 1 and 3, the difference between the strength of dependence is moderate across clusters, while the difference is more obvious in Settings 2, 4 and 5. To demonstrate the capability of our proposed BHCM in handling the setting with multiple copula parameters, in Setting 5, we further consider copulas with a single parameter in clusters 1 and 2, and copulas with two parameters in clusters 3 and 4. A moderate dependence between clusters is introduced in all settings and the linking copulas are set to be Gaussian(0.71).

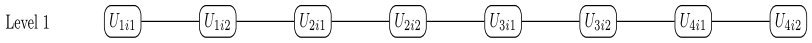


FIG 2. *The top level of a D-Vine structure*

TABLE 2
Simulation Settings: Copula Forms and Parameters

	Setting 1	τ[1]	Setting 2	τ	Setting 3	τ	Setting4	τ	Setting5	τ
Cluster 1	Clayton(1.33)	0.40	Clayton(1.33)	0.40	Clayton(3.00)	0.60	Clayton(3.00)	0.60	Gumbel(2.50)	0.60
Cluster 2	Clayton(1.64)	0.45	Clayton(2.00)	0.50	Gumbel(2.50)	0.60	Gumbel(4.00)	0.75	Joe(2.50)	0.45
Cluster 3	Clayton(2.00)	0.50	Clayton(3.00)	0.60	Gaussian(0.81)	0.60	Gaussian(0.60)	0.41	BB1(5.00,3.00)[2]	0.90
Cluster 4	Clayton(2.44)	0.55	Clayton(4.67)	0.70	Frank(7.93)	0.60	Frank(13.00)	0.73	BB7(3.00,5.00)[3]	0.73
Between-cluster	Gaussian(0.71)	0.50	Gaussian(0.71)	0.50	Gaussian(0.71)	0.50	Gaussian(0.71)	0.50	Gaussian(0.71)	0.50

[1] Kendall's τ
[2] Clayton-Gumbel Copula
[3] Joe-Clayton Copula

We construct the following BHCM. For $i = 1, \ldots, n$, $j = 1, 2, 3, 4$ and $l = 1, 2$ (for setting 5), assume that

$$\begin{aligned} U_{ji} = (U_{ji1}, U_{ji2}) &\sim C_j(u_{ji1}, u_{ji2}; \theta_j) \\ \gamma_{jl} &= \alpha_{jl} g_{jl}(\theta_{jl}), \\ \gamma_{jl} | \mu_{jl}, \sigma_{jl} &\sim N(\mu_{jl}, \sigma_{jl}^2), \\ \mu_{jl} | \varphi_l, \delta_l &\sim N(\varphi_l, \delta_l^2), \end{aligned}$$

and all the hyperprior parameters have non-informative uniform priors. Sampling $N = 6000$ from the posterior distribution and setting the Normal density with mean $\zeta^{(t-1)}$ and variance as the stepsize as the proposal density $q(\zeta' | \zeta^{(t-1)})$, we use the M-H algorithm and the layer-by-layer sampling strategy described in Section 3.2 to sample θ. The posterior sample mean is used as the point Bayesian estimators for the parameters. In comparison, we also obtain the MLE of θ by maximizing the likelihood function

$$L(\theta) = \prod_{i=1}^{n} \left[\prod_{j=1}^{4} c_j(u_{ji1}, u_{ji2}; \theta_j) \times \prod_{k=1}^{3} c_{k,k+1}(u_{ki2}, u_{k+1,i1}) \right],$$

where c_j is the copula density governing the subject-dependence within cluster j for $j = 1, \ldots, 4$, and $c_{k,k+1}$ denotes the copula densities that connect between clusters for $k = 1, 2, 3$.

While the sampling algorithm is implemented on the R platform, we handle the integrals in the posterior distribution (3.2) by employing C++ through Monte Carlo approximations of size 15000, which is computationally fast yet the resulting approximation is fairly accurate. Simulations are repeated 200 times for each setting.

5.2. Evaluation metrics

We use the following metrics to evaluate the Bayesian estimators and MLEs.

1. *Empirical Bias (EBias)*: The EBias is calculated as the average of the point estimates obtained from 200 simulations subtracting the true parameter values.
2. *Empirical Standard Error (ESE)*: The sample standard deviation of the 200 estimates.
3. *Asymptotic Standard Error (ASE)*: The average of the estimated asymptotic standard deviations obtained from the 200 simulations. The estimated asymptotic s.d. for a Bayesian estimator is calculated as the sample s.d. of the sampled sequence, and that of a maximum likelihood estimator is calculated from the inversion of the observed information matrix.
4. *95% Interval*: Left and right endpoints of an equal-tailed 95% Bayesian credible interval are computed as the 2.5th percentile and the 97.5th percentile of a sampled sequence, respectively. A 95% confidence interval for the MLE is computed by MLE$\pm$1.96$\times$the estimated asymptotic s.d.. A 95% interval is computed by averaging the left and right endpoints of 200 simulations [6].
5. *Empirical Coverage Probability (ECP)*: ECP is the percentage of the 95% credible intervals or 95% confidence intervals that contain the true value of the parameter out of 200 simulations.

5.3. Simulation results

We summarize the simulation results for Setting 5 in Table 3, and those for Settings 1-4 in Tables 7-10 in the Appendix.

The findings for all the settings reveal consistent patterns, as commented below. We tune L, the target length of a 95% confidence interval of $\tilde{\gamma}_{jl}$, to be 1, 4, 10 and 20 for comparison (results for $L = 1$ and 10 not shown). For the point estimates of the copula parameters under all simulation settings, the EBias of estimates obtained from the proposed BHCM are compatible with or smaller than those from the likelihood-based estimates. The Bayesian estimators with $L = 1$ have similar ESE's and ASE's to those of the likelihood-based estimates; the standard error of the Bayesian estimators gets smaller, as L gets larger. For interval estimates of the copula parameters, 95% Bayesian credible intervals of the proposed BHCM are shorter than the likelihood-based 95% intervals when L is set to be 4, 10 or 20. When L is set to be a large number, there are

TABLE 3
Simulation Results for Setting 5

				n=200					n=400				
Cluster	Copula	Parameter	L	Ebias	ESE	ASE	95% interval	ECP	Ebias	ESE	ASE	95% interval	ECP
Bayesian Estimation													
1	Gumbel	2.5	4	0.020	0.150	0.133	(2.264,2.789)	0.940	0.003	0.093	0.095	(2.321,2.694)	0.950
2	Joe	2.5	4	-0.029	0.166	0.158	(2.172,2.795)	0.950	-0.024	0.118	0.111	(2.263,2.701)	0.940
3	BB1	5.0	4	0.111	0.617	0.478	(4.116,5.998)	0.920	-0.141	0.383	0.316	(4.227,5.474)	0.925
		3.0	4	0.083	0.247	0.270	(2.552,3.617)	0.970	0.021	0.172	0.171	(2.689,3.364)	0.960
4	BB7	3.0	4	0.005	0.279	0.253	(2.551,3.550)	0.940	0.043	0.232	0.185	(2.717,3.403)	0.905
		5.0	4	0.038	0.486	0.497	4.079,6.036)	0.970	0.021	0.364	0.347	(4.359,5.727)	0.940
1	Gumbel	2.5	20	-0.037	0.176	0.113	(2.249,2.694)	0.820	-0.015	0.087	0.072	(2.346,2.631)	0.870
2	Joe	2.5	20	-0.075	0.191	0.133	(2.175,2.699)	0.845	-0.013	0.111	0.090	(2.315,2.669)	0.910
3	BB1	5.0	20	0.091	0.835	0.407	(4.207,5.811)	0.765	-0.152	0.321	0.218	(4.401,5.262)	0.820
		3.0	20	0.048	0.233	0.208	(2.647,3.469)	0.890	0.021	0.109	0.083	(2.860,3.188)	0.900
4	BB7	3.0	20	0.042	0.442	0.228	(2.632,3.529)	0.780	0.113	0.198	0.126	(2.872,3.369)	0.810
		5.0	20	0.051	0.470	0.451	(4.178,5.956)	0.950	0.053	0.322	0.250	(4.565,5.552)	0.920
Maximum Likelihood Estimation													
1	Gumbel	2.5	-	0.024	0.141	0.147	(2.236,2.812)	0.960	0.020	0.106	0.104	(2.317,2.723)	0.950
2	Joe	2.5	-	0.036	0.169	0.178	(2.187,2.885)	0.960	0.030	0.126	0.131	(2.284,2.776)	0.940
3	BB1	5.0	-	-0.373	0.541	0.863	(2.936,6.318)	0.940	-0.289	0.448	0.627	(3.483,5.940)	0.930
		3.0	-	0.234	0.365	0.472	(2.309,4.159)	0.960	0.178	0.300	0.332	2.527,3.830)	0.930
4	BB7	3.0	-	0.060	0.285	0.296	(2.479,3.640)	0.930	0.062	0.252	0.209	(2.654,3.471)	0.910
		5.0	-	0.072	0.520	0.547	(4.005,6.149)	0.980	0.060	0.389	0.384	(4.308,5.812)	0.920

unignorable gaps between the ESE's and ASE's, and ECP deviates from the 95% nominal level. This is attributed to the strong prior imposed on γ_{jl}^* as we discussed in Section 4.2, so that the posterior distribution is highly peaked and deviated from the normal distribution. We recommend against choosing L to be too small (close to 1) or too large (greater than 10). The former imposes a weak prior and leads to results similar to maximum likelihood estimates, and the latter imposes a too strong prior and leads to an underestimated standard deviation and a possibly inflated bias.

As the sample size increases from 200 to 400, both the proposed BHCM and MLE provide estimates with smaller biases, a better agreement between ESE's and ASE's, and the coverage rates closer to the 95% nominal level. The improvement in the standard error of the BHCM estimates, compared to the likelihood-based estimates, is reduced, since Bayesian estimation tends to perform better with a smaller sample size and the two estimation methods have the same limiting distribution, which, therefore, has similar performance as the sample size gets larger. The gaps between ESE's and ASE's of Bayesian estimates with a large L are getting closer as the sample size increases, showing that the posterior distributions get closer to normality with a larger sample.

For the Frank copula with the range $(-\infty, \infty)\backslash\{0\}$ in Settings 3 and 4, we report the results of two different choices of transformation functions in Tables 9 and 10 in Appendix, respectively. The identity transformation function $g(\theta) = \theta$ performs poorly with a small sample size, compared to the transformation function $g(\theta) = \log(\frac{100+\theta}{100-\theta})$. As the sample size increases from 200 to 400, the two transformation functions seem to work equally well.

Above all, with $L = 4$ across all settings, the BHCM provides reasonable point estimates and interval estimates of copula parameters, and smaller EBias

and shorter 95% intervals than those from the maximum likelihood method. The benefit of using the BHCM is more obvious if the clusters share more similarity in the subject-level dependence structures (e.g., Setting 1). The BHCM exhibits the capability of handling settings with copula structures containing both one- and two-parameter copulas and large differences in dependence strength.

For the BHCM with $L = 4$ in Setting 5, we also report the sample trace plots and sample density plots for the results of the mean parameters φ_l and μ_{jl} and those for the copula parameters θ_{jl} for $j = 1, 2, 3, 4$ and $l = 1, 2$, respectively, in Figures 3 and 4. In all the sample trace plots, the samples of the mean parameters and the copula parameters vary closely around the posterior mean, and the sample densities are all close to a bell shape, indicating the convergence of the M-H algorithm.

6. Data analysis

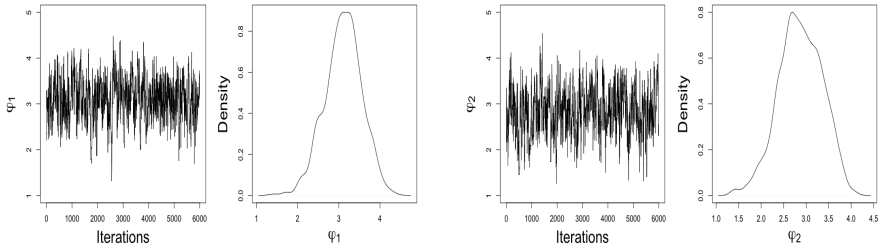
We now apply the proposed BHCM to analyze the Vertebral Column dataset from the UCI Machine Learning Repository (`http://archive.ics.uci.edu/ml/datasets/vertebral+column`). This is a biomedical dataset collected by Dr. Henrique da Mota during a medical residence at Lyon, France. The dataset contains biomedical features of 60 patients with disk hernia, 150 patients with spondylolisthesis and 100 healthy volunteers. The three groups of people are labeled as $j = 1, 2, 3$, respectively. The biomechanical features include angle of pelvic incidence `(PI)`, angle of pelvic tilt `(PT)`, lumbar lordosis angle `(LL)`, sacral slope `(SS)`, pelvic radius `(PR)`, and degree of spondylolisthesis `(DS)`, which are labeled as $k = 1, 2, 3, 4, 5$ and 6, respectively. For $j = 1, 2, 3$, $i = 1, \ldots, n_j$, and $k = 1, 2, 3, 4, 5$, let Y_{ijk} denote the kth biomedical features of the ith subject from the jth group of people, where $n_1 = 60$, $n_2 = 150$, and $n_3 = 100$.

In medical research, `PR` describes the pelvic lordosis angle, and `PI`, `PT` and `SS` describe the shape and orientation of the pelvis. They represent two different approaches to characterize the pelvis. For the latter one, `PI` is defined as "the angle between a line perpendicular to the sacral plate and a line joining the sacral plate to the axis of the femoral heads" and is the arithmetic summation of `PT` and `SS` [4]. We are interested in examining the dependence of `PI` versus `PT` and of `PI` versus `SS`. `DS` is the degree of slipping and can take negative values. We are interested in understanding its association with the characteristics of pelvis including `PI`, `PT` and `PR`, and that of lumbar `LL`.

6.1. Marginal model

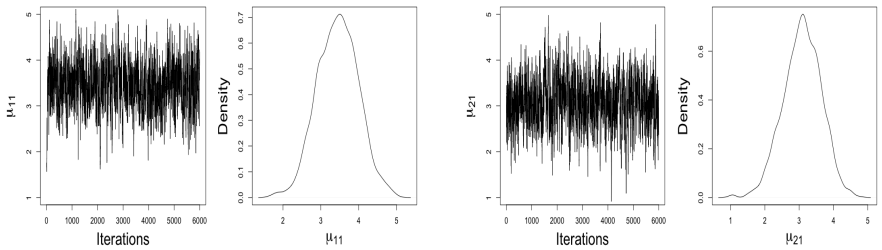
The histograms of the six biomedical features in three groups are displayed in Figure 5 in Appendix C.1, all showing uni-modal but possibly skewed distributions. As a result, we use a generalized skewed-t distribution to model the marginal distributions of the features to account for the possible skewness.

The estimates of the marginal parameters are obtained by maximizing the marginal likelihood function, and the results are summarized in Table 11 in Appendix C. The six biomedical features are transformed to copula data $u_{jik} \in$



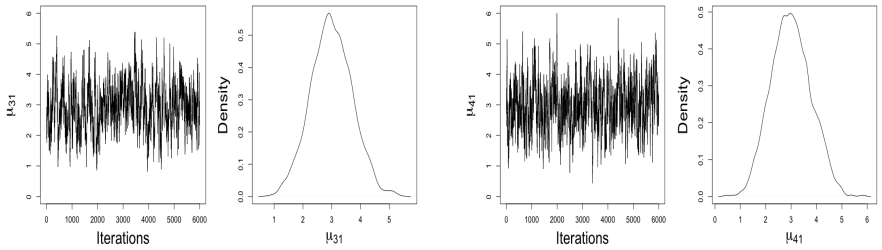
(1) φ_1

(2) φ_2



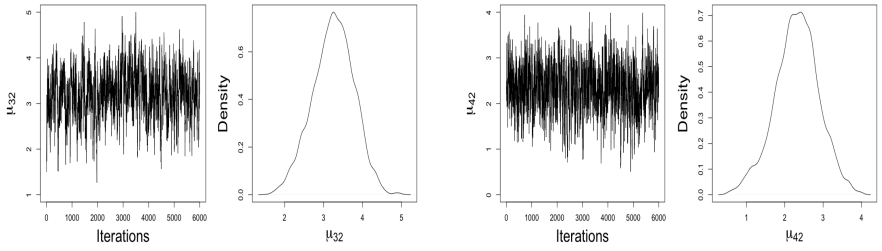
(3) μ_{11}

(4) μ_{21}



(5) μ_{31}

(6) μ_{41}



(7) μ_{32}

(8) μ_{42}

FIG 3. *Sample trace plots and sample density plots of mean parameters* φ_l *and* μ_{jl} *for* $j = 1, 2, 3, 4$ *and* $l = 1, 2$ *of the BHCM with* $L = 4$ *in Setting 5*

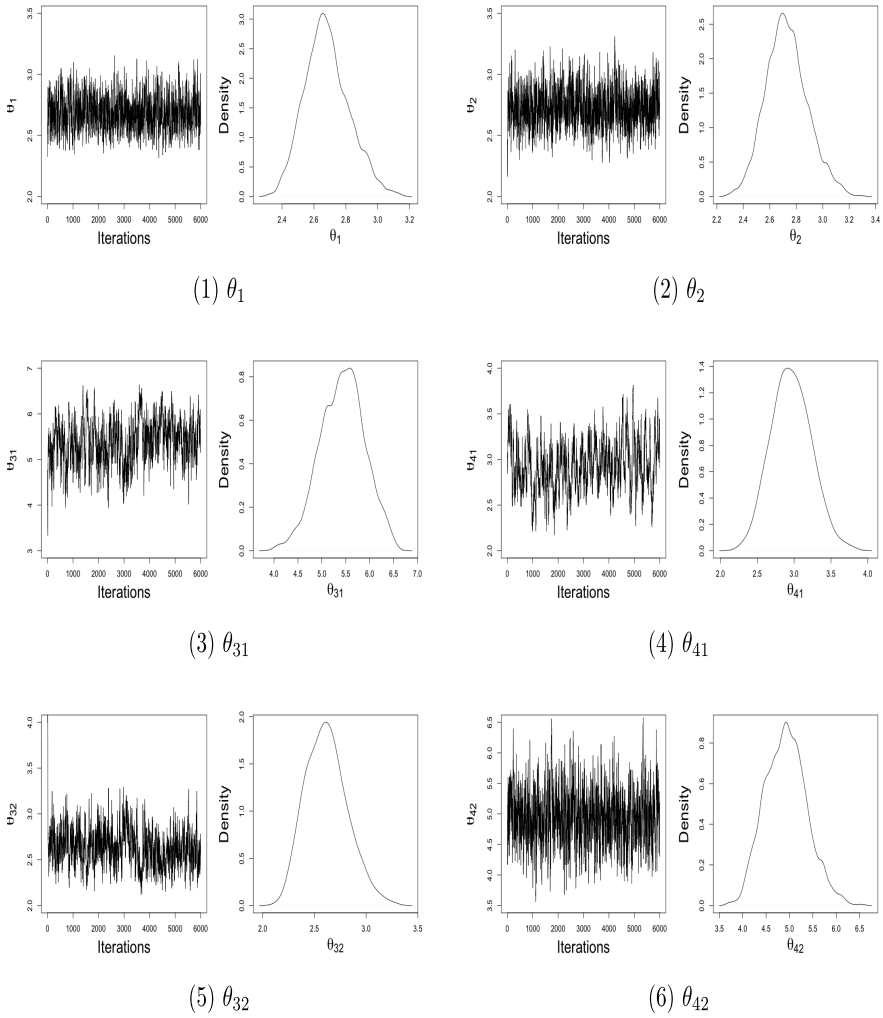


FIG 4. *Sample trace plots and sample density plots of copula parameters* θ_{jl} *for* $j = 1, 2, 3, 4$ *and* $l = 1, 2$ *of the BHCM with* $L = 4$ *in Setting 5*

$[0, 1]$ by applying the fitted marginal CDF to the observed values of the corresponding feature. Let U_{jik} denote the transformed uniformed random variable for the kth feature of the ith subject in group j for $j = 1, 2, 3$, $i = 1, \ldots, n_j$ and $k = 1, \ldots, 6$.

6.2. Dependence model

We are interested in studying the dependence between the following 6 pairs of variables: `PI` versus `PT`, `PI` versus `SS`, `DS` versus `PI`, `DS` versus `PT`, `DS` versus `PR`,

and DS versus LL. The scatter plots for those pairs are displayed in Figure 6 in Appendix C.2.

We construct a set of parametric copula functions, including the commonly-used copulas in the Archimedean family (Clayton, Gumbel, Frank and Joe copulas), Gaussian copula and their rotated versions. The specific copula function forms are selected based on the AIC criterion [2], which is conducted using the BiCopSelect function in the R package VineCopula [43]. For each bivariate feature, we construct a BHCM for the three groups of individuals.

For comparison, we consider two benchmark models. The first one is a multivariate copula model (MCopula), which takes the same marginal and dependence models as the BHCM, i.e., the marginals are generalized skewed-t distributions and copula models are selected using AIC as reported in Table 5. The second one is a multivariate Gaussian model (MVN), in which the marginal distributions are all specified as a Gaussian distribution and the copulas of the interested six pairs are also specified as the Gaussian copula. The parameters in both benchmark models are estimated using the maximum likelihood method.

6.3. Results

We compare the performance of the three models, BHCM, MCopula and MVN, in terms of log-likelihood values and the Deviance Information Criterion (DIC) [49], and summarize the results in Table 4. The BHCM has the smallest overall DIC, thus being the best to fit the data. For the clusters of patients with Spondilolisthesis and being healthy, the marginal distributions of some features, for instance, DS, are highly skewed as shown in Figure 5. MVN provides a poor fit of the data, yielding the smallest log-likelihood and the largest DIC. For the cluster of patients with Disk Hernia, the skewness in the marginal distributions is mild and most of the bivariate copulas selected are the Gaussian copula as shown in Table 5. The BHCM and MCopula produce log-likelihood values similar to that of MVN but smaller DIC than MVN does, which is partially attributed to the fact that the BHCM and the MCopula are penalized by extra parameters in their marginal generalized skewed-t distributions.

TABLE 4
Log-likelihood and DIC of three models for each cluster

	Disk Hernia		Spondilolisthesis		Healthy		Total	
	log-likelihood	DIC	log-likelihood	DIC	log-likelihood	DIC	log-likelihood	DIC
BHCM	-1209.79	2464.05	-3639.85	7322.6	-2060.90	4166.29	-6910.54	13952.96
MCopula	-1209.90	2467.78	-3637.70	7323.46	-2062.70	4173.45	-6910.34	13964.68
MVN	-1212.80	2461.66	-3686.70	7409.31	-2079.30	4194.61	-6978.79	14065.58

Tables 5 shows the point estimates and interval estimates under the proposed BHCM with $L = 4$ together with the results obtained from the likelihood-based method. Once the Frank copula selected, we use the logit transformation function, which leads to more stable results than the identity transformation function

when the sample size is small. It is seen that PI has a positive dependence on PT and SS, which aligns with the reports in the medical literature [4]. Across different groups, the dependence strengths of PI versus PT and PI versus SS show similar Kendall's τ ranging from 0.4 to 0.6. The dependence between DS and other pelvic and lumbar characteristics show an obvious distinction across groups. For patients with disease disk hernia and healthy people, DS has a weak dependence on other four features. However, for patients with Spondylolisthesis, DS has a much stronger positive dependence on the four features.

The BHCM with $L = 4$ produces similar point estimates to those obtained from the likelihood-based method, but smaller standard errors. 95% credible interval of the BHCM with $L = 4$ are narrower than 95% confidence intervals obtained from the likelihood-based method. For the cluster of patients with Spondilolisthesis, the DS feature is highly right-skewed as shown in Figure 5, thus the MVN model fails to fit the data well.

7. Discussion

In this paper, we propose a Bayesian hierarchical copula model to characterize the subject-level dependence for data with a hierarchical association structure. The model is flexible to account for data coming from multiple sources with different sample sizes. We use a "layer-by-layer" sampling scheme, combined with the Metropolis-Hasting algorithm, to sample from the posterior distribution. Simulation studies and data analysis are conducted to compare the estimators obtained from our proposed BHCM to the likelihood-based estimators. The results show that the BHCM outperforms the maximum likelihood methods when the sample size is small. The proposed model captures the between-cluster variability and facilitate the information shared across clusters through delineating the hierarchical structures.

In forming the copula models here, the marginal distributions are assumed to be uniform over the unit interval $[0, 1]$. However, this assumption is not essential. Other parametric models, such as the normal distribution and generalized skewed-t distribution, can be considered to reflect various data features. Furthermore, nonparametric models can also be considered as robust alternatives. It is interesting to extend the proposed method to accommodate these settings.

We comment that our BHCM differs from the Hierarchical Archimedean Copula (HAC) proposed by [38]. Since an Archimedean copula function can be defined through the generator function of the copula [e.g., 36], an HAC is built by applying the generator function to a lower level HAC in a recursive manner. An HAC overcomes some disadvantages of a regular Archimedean copula. However, it is not designed to handle a hierarchical structure as the one in Figure 1. Though our proposed BHCM does not necessarily feature an HAC as the fundamental building block, our proposed framework is general enough to cover the structures that the HAC can handle.

The proposed method invokes different regularization on the estimation of the copula parameters by the tuning parameter L and the estimates of the

TABLE 5. *Copula Functions and Estimates for Six Interested Dependence of Three Health Groups*

			BHCM with $L = 4$			MCopula			MVN			
Group	Dependence Relations	Copula	Estimates	s.d.	95% Interval	Estimates	s.d.	95% Interval	Copula	Estimates	s.d.	95% Interval
Disk Hernia	PI v.s. PT	Gaussian	0.696	0.046	(0.599,0.775)	0.694	0.055	(0.586,0.801)	Gaussian	0.710	0.052	(0.608,0.812)
	PI v.s. SS	Gaussian	0.726	0.040	(0.633,0.793)	0.766	0.042	(0.683,0.849)	Gaussian	0.756	0.044	(0.670,0.842)
	DS v.s. PI	Gaussian	0.161	0.098	(-0.031,0.339)	0.150	0.125	(-0.095,0.395)	Gaussian	0.144	0.125	(-0.101,0.389)
	DS v.s. PT	Frank	-0.511	0.577	(-1.489,0.522)	-0.226	0.753	(-1.702,1.250)	Gaussian	0.044	0.129	(-0.209,0.297)
	DS v.s. LL	Gaussian	0.244	0.103	(0.031,0.435)	0.246	0.118	(0.015,0.477)	Gaussian	0.231	0.119	(-0.002,0.464)
	DS v.s. PR	Gaussian	-0.055	0.113	(-0.263,0.175)	-0.060	0.128	(-0.312,0.191)	Gaussian	-0.051	0.129	(-0.304,0.202)
Spondilolisthesis	PI v.s. PT	Frank	5.718	0.505	(0.599,0.775)	5.594	0.622	(4.375,6.814)	Gaussian	0.601	-	-
	PI v.s. SS	Gumbel	1.729	0.099	(1.554,1.943)	1.736	0.113	(1.515,1.958)	Gaussian	0.665	-	-
	DS v.s. PI	Frank	3.427	0.431	(2.552,4.245)	3.453	0.535	(2.404,4.502)	Gaussian	0.533	-	-
	DS v.s. PT	S Clayton[1]	0.887	0.143	(0.608,1.174)	0.905	0.153	(0.605,1.206)	Gaussian	0.439	-	-
	DS v.s. LL	Frank	3.230	0.426	(2.437,4.104)	3.155	0.527	(2.121,4.189)	Gaussian	0.324	-	-
	DS v.s. PR	Joe	1.466	0.115	(1.265,1.698)	1.481	0.123	(1.239,1.723)	Gaussian	0.329	-	-
Healthy	PI v.s. PT	Gaussian	0.633	0.038	(0.555,0.699)	0.636	0.051	(0.537,0.735)	Gaussian	0.634	0.051	(0.534,0.734)
	PI v.s. SS	Gumbel	2.574	0.178	(2.239,2.910)	2.599	0.214	(2.179,3.018)	Gaussian	0.839	0.023	(0.794,0.884)
	DS v.s. PI	Frank	1.822	0.430	(0.936,2.632)	1.714	0.628	(0.483,2.945)	Gaussian	0.200	0.094	(0.016,0.384)
	DS v.s. PT	Gaussian	0.242	0.080	(0.085,0.401)	0.244	0.091	(0.065,0.423)	Gaussian	0.182	0.095	(-0.004,0.368)
	DS v.s. LL	Frank	1.409	0.570	(0.335,2.538)	1.511	0.600	(0.334,2.687)	Gaussian	0.261	0.090	(0.085,0.437)
	DS v.s. PR	Gaussian	-0.111	0.093	(-0.289,0.065)	-0.107	0.098	(-0.299,0.086)	Gaussian	-0.058	0.099	(-0.252,0.136)

[1] Survival Clayton Copula

hyperprior parameters $\hat{\sigma}$, $\hat{\varphi}$, and $\hat{\delta}$. While the hyperprior parameters bring in information "borrowed" from other clusters, the tuning parameter L controls the strength that the hyperprior parameters can influence the copula parameters, as discussed in Section 4 and shown in Section 5.

Finally, we comment that the BHCM model does not have to be restricted to existing forms of multivariate copulas but takes the advantage of flexible vine copulas to formulate the subject-level multivariate dependence structures. Considering different vine structures and copula selection strategies for each bivariate component, one may develop the vine-copula-based BHCM, which can be an interesting topic to be explored in depth. One challenge that may hinder the inference about the vine-copula-based BHCM concerns the high dimension of the posterior density, and developing more efficient sampling algorithms is generally required.

Acknowledgement

The authors thank the Associate Editor and the review team for their helpful comments on the initial submission. This research is supported by the Natural Science and Engineering Research Council of Canada (NSERC; Grant ID: 299493 (YY) and 1280961 (LD)). Yi is a Canada Research Chair in Data Science (Tier 1). Her research was undertaken, in part, thanks to funding from the Canada Research Chairs Program.

References

[1] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198. MR2517884

[2] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer. MR1486823

[3] Bedford, T. and Cooke, R. M. (2002). Vines: a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068. MR1926167

[4] Berthonnaud, E., Dimnet, J., Roussouly, P., and Labelle, H. (2005). Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters. *Clinical Spine Surgery*, 18(1):40–47.

[5] Broët, P., Richardson, S., and Radvanyi, F. (2002). Bayesian hierarchical model for identifying changes in gene expression from microarray experiments. *Journal of Computational Biology*, 9(4):671–683.

[6] Chen, M.-H. and Shao, Q.-M. (1999). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 8(1):69–92. MR1705909

[7] Congdon, P. D. (2010). *Applied Bayesian Hierarchical Methods*. Chapman and Hall/CRC. MR2723193
[8] DeGroot, M. H. (2005). *Optimal Statistical Decisions*. John Wiley & Sons. MR2288194
[9] Dissmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59(1):52–69. MR3000041
[10] Dua, D. and Graff, C. (2017). UCI machine learning repository.
[11] Fei-Fei, L. and Perona, P. (2005). A Bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 524–531.
[12] Frahm, G., Junker, M., and Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statistics & Probability Letters*, 63(3):275–286. MR1986327
[13] Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409. MR1141740
[14] Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC. MR3235677
[15] Geman, S. and Geman, D. (1987). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. In *Readings in Computer Vision*. Elsevier.
[16] Genest, C. and MacKay, J. (1986a). The joy of copulas: bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283. MR0866908
[17] Genest, C. and MacKay, R. J. (1986b). Copules Archimédiennes et families de lois bidimensionnelles dont les marges sont données. *Canadian Journal of Statistics*, 14(2):145–159. MR0849869
[18] George, D. and Hawkins, J. (2005). A hierarchical Bayesian model of invariant pattern recognition in the visual cortex. In *2005 IEEE International Joint Conference on Neural Networks, Proceedings.*, volume 3.
[19] Ghosh, S. (2010). *Copula Based Hierarchical Bayesian Models*. PhD thesis, University of Texas A&M. MR2714038
[20] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). Introducing Markov Chain Monte Carlo. *Markov Chain Monte Carlo in Practice*, 1:19. MR1397966
[21] Gruber, L., Czado, C., et al. (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis*, 10(4):937–963. MR3432245
[22] Gruber, L. F., Czado, C., et al. (2018). Bayesian model selection of regular vine copulas. *Bayesian Analysis*, 13(4):1107–1131. MR3855365
[23] Gustafson, P., Hossain, S., and Macnab, Y. C. (2006). Conservative prior distributions for variance parameters in hierarchical models. *Canadian Journal of Statistics*, 34(3):377–390. MR2328550
[24] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov

chains and their applications. *Biometrika*, 57(1):97–109. MR3363437
[25] Hughes, J. P. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, 30(5):1535–1546. MR2689842
[26] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461. MR0017504
[27] Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. Chapman and Hall/CRC. MR1462613
[28] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419. MR2167922
[29] Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman and Hall/CRC. MR3328438
[30] Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.
[31] LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related bayes estimates. *University of California Publications in Statistics*, 1:277–330. MR0054913
[32] Lee, S. (1998). Coefficient constancy test in a random coefficient autoregressive model. *Journal of Statistical Planning and Inference*, 74(1):93–101. MR1665122
[33] Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B*, 34(1):1–18. MR0415861
[34] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
[35] Min, A. and Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546. MR3051264
[36] Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media. MR2197664
[37] Nicholls, D. F. and Quinn, B. G. (1980). The estimation of random coefficient autoregressive models. *Journal of Time Series Analysis*, 1(1):37–46. MR0605573
[38] Okhrin, O., Okhrin, Y., and Schmid, W. (2013a). On the structure and estimation of hierarchical Archimedean copulas. *Journal of Econometrics*, 173(2):189–204. MR3019689
[39] Okhrin, O., Okhrin, Y., and Schmid, W. (2013b). Properties of hierarchical Archimedean copulas. *Statistics & Risk Modeling with Applications in Finance and Insurance*, 30(1):21–54. MR3041504
[40] Radford, N., Brooks, S., Gelman, A., Jones, G., and Meng, X. (2010). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 31:32. MR2742422
[41] Sáfadi, T. and Morettin, P. (2003). A Bayesian analysis of autoregressive

models with random normal coefficients. *Journal of Statistical Computation and Simulation*, 73(8):563–573. MR1998669

[42] Schamberger, B., Gruber, L. F., and Czado, C. (2017). Bayesian inference for latent factor copulas and application to financial risk forecasting. *Econometrics*, 5(2):21.

[43] Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T. (2018). *VineCopula: Statistical Inference of Vine Copulas*. R package version 2.1.4.

[44] Shen, X. and Wasserman, L. (2001). Rates of convergence of posterior distributions. *The Annals of Statistics*, 29(3):687–714. MR1865337

[45] Sklar, A. (1959). Functions de repartition an dimension set leursmarges. *Publications de L'In-stitut de Statistique de L'Universite de Paris*, 8:229–231. MR2724323

[46] Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman and Hall/CRC. MR2059021

[47] Smith, M., Min, A., Almeida, C., and Czado, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association*, 105(492):1467–1479. MR2796564

[48] Smith, M. S. (2011). Bayesian approaches to copula modelling. *arXiv preprint arXiv:1112.4204*. MR3221171

[49] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):485–493. MR3210727

[50] Stober, J. and Schepsemeier, U. (2013). Estimating standard errors in regular vine copula models. *Computational Statistics*, 28:2679–2707. MR3141358

Appendix A: Variability of the transformed dependence parameters

We explore the within-cluster variability of $\tilde{\gamma}_{jl}^*$, which relates to the choice of rescaling parameter α_{jl}. We use simulations to show how the standard error of the transformed dependence parameter, $\widehat{\text{sd}}(\tilde{\gamma}_{jl})$, varies with respect to the copula form, the sample size, and the strength of dependence. Five commonly-used copula forms, Clayton copula, Gumbel copula, Joe copula, Frank copula, and Gaussian copula, are considered with Kendall's τ varing from 0.1 to 0.9 and the sample size $n = 200$ or 400. In each scenario, simulation is repeated 500 times, and the transformed dependence parameters are estimated using maximum likelihood estimation with standard errors calculated from the inverse of the observed information. We report the results in Table 6.

The results show that the copula form, the true parameter values, the sample size, and the transformation function all affect the standard error of the transformed parameter γ_{jl}^*.

TABLE 6
Empirical standard error of the MLE of transformed dependence parameter under various copula functions

Kendall's	Clayton		Gumbel		Joe		Gaussian		Frank		Frank†	
τ	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$	$n=200$	$n=400$
0.1	0.590	0.348	0.844	0.413	0.632	0.347	0.142	0.103	0.431	0.307	0.009	0.006
0.2	0.229	0.160	0.291	0.197	0.262	0.175	0.137	0.098	0.436	0.311	0.009	0.006
0.3	0.152	0.108	0.192	0.134	0.177	0.121	0.130	0.093	0.453	0.324	0.009	0.007
0.4	0.118	0.083	0.146	0.103	0.136	0.093	0.123	0.087	0.487	0.350	0.010	0.007
0.5	0.098	0.069	0.119	0.084	0.111	0.076	0.116	0.083	0.548	0.395	0.011	0.008
0.6	0.085	0.061	0.101	0.071	0.096	0.066	0.110	0.079	0.651	0.471	0.013	0.009
0.7	0.076	0.054	0.087	0.061	0.085	0.059	0.106	0.076	0.837	0.604	0.017	0.012
0.8	0.069	0.050	0.077	0.054	0.076	0.053	0.104	0.074	1.225	0.883	0.025	0.018
0.9	0.064	0.046	0.069	0.048	0.068	0.048	0.101	0.073	2.429	1.730	0.057	0.041

† Using transformation function $g(x) = \alpha \log(\frac{x+100}{100-x})$

Appendix B: Additional simulation results

TABLE 7
Simulation Results for Setting 1

			$n=200$					$n=400$				
Cluster	Copula	L	EBias	ESE	ASE	95% Interval	ECP	EBias	ESE	ASE	95% Interval	ECP
Bayesian Estimation												
1	Clayton(1.33)	4	-0.001	0.125	0.127	(1.091,1.589)	0.950	-0.009	0.073	0.089	(1.154,1.503)	0.960
2	Clayton(1.64)	4	-0.002	0.140	0.157	(1.335,1.949)	0.970	-0.015	0.103	0.109	(1.413,1.839)	0.950
3	Clayton(2.00)	4	0.009	0.170	0.181	(1.665,2.373)	0.970	0.001	0.121	0.127	(1.756,2.253)	0.955
4	Clayton(2.44)	4	0.023	0.198	0.203	(2.081,2.876)	0.930	-0.010	0.155	0.145	(2.155,2.724)	0.940
1	Clayton(1.33)	20	0.039	0.070	0.058	(1.245,1.474)	0.850	0.013	0.044	0.037	(1.275,1.420)	0.905
2	Clayton(1.64)	20	-0.005	0.108	0.106	(1.432,1.848)	0.910	-0.014	0.085	0.070	(1.489,1.762)	0.890
3	Clayton(2.00)	20	0.003	0.150	0.147	(1.729,2.303)	0.890	< 0.001	0.122	0.093	(1.820,2.185)	0.900
4	Clayton(2.44)	20	-0.024	0.172	0.181	(2.076,2.786)	0.865	-0.014	0.152	0.123	(2.197,2.680)	0.910
Maximum Likelihood Estimation												
1	Clayton(1.33)	-	0.021	0.147	0.158	(1.043,1.664)	0.970	-0.004	0.098	0.111	(1.112,1.547)	0.960
2	Clayton(1.64)	-	0.023	0.165	0.176	(1.315,2.003)	0.955	-0.013	0.113	0.123	(1.382,1.864)	0.965
3	Clayton(2.00)	-	0.012	0.191	0.196	(1.628,2.396)	0.965	0.006	0.127	0.139	(1.734,2.277)	0.945
4	Clayton(2.44)	-	0.028	0.216	0.223	(2.035,2.908)	0.955	-0.007	0.155	0.156	(2.131,2.743)	0.950

TABLE 8
Simulation Results for Setting 2

			$n=200$					$n=400$				
Cluster	Copula	L	EBias	ESE	ASE	95% Interval	ECP	EBias	ESE	ASE	95% Interval	ECP
Bayesian Estimation												
1	Clayton(1.33)	4	0.020	0.147	0.150	(1.087,1.677)	0.940	-0.004	0.071	0.081	(1.174,1.490)	0.945
2	Clayton(2.00)	4	0.002	0.169	0.177	(1.665,2.358)	0.965	-0.012	0.115	0.127	(1.744,2.243)	0.960
3	Clayton(3.00)	4	-0.013	0.220	0.233	(2.543,3.456)	0.945	0.003	0.161	0.167	(2.682,3.337)	0.935
4	Clayton(4.67)	4	-0.041	0.346	0.319	(4.018,5.269)	0.945	-0.021	0.249	0.230	(4.203,5.104)	0.940
1	Clayton(1.33)	20	0.054	0.139	0.102	(1.195,1.595)	0.815	0.021	0.052	0.043	(1.265,1.403)	0.895
2	Clayton(2.00)	20	0.038	0.162	0.124	(1.806,2.260)	0.830	0.004	0.094	0.089	(1.834,2.183)	0.910
3	Clayton(3.00)	20	-0.047	0.204	0.178	(2.615,3.313)	0.845	-0.014	0.159	0.142	(2.721,3.279)	0.880
4	Clayton(4.67)	20	-0.071	0.338	0.283	(4.054,5.164)	0.810	-0.033	0.242	0.218	(4.216,5.069)	0.910
Maximum Likelihood Estimation												
1	Clayton(1.33)	-	0.017	0.148	0.158	(1.040,1.660)	0.965	-0.002	0.099	0.111	(1.113,1.548)	0.965
2	Clayton(2.00)	-	0.009	0.191	0.196	(1.626,2.393)	0.950	-0.014	0.126	0.138	(1.716,2.256)	0.960
3	Clayton(3.00)	-	-0.010	0.232	0.253	(2.494,3.486)	0.950	0.004	0.168	0.179	(2.652,3.355)	0.955
4	Clayton(4.67)	-	-0.036	0.351	0.349	(3.947,5.316)	0.940	-0.019	0.253	0.248	(4.162,5.133)	0.940

TABLE 9
Simulation Results for Setting 3

			$n = 200$					$n = 400$				
Cluster	Copula	L	EBias	ESE	ASE	95% Interval	ECP	EBias	ESE	ASE	95% Interval	ECP
Bayesian Estimation												
1	Clayton(3)	4	0.004	0.215	0.218	(2.587,3.444)	0.955	0.011	0.141	0.154	(2.705,3.307)	0.945
2	Gumbel(2.5)	4	-0.002	0.140	0.128	(2.253,2.755)	0.920	-0.004	0.075	0.093	(2.321,2.686)	0.915
3	Gaussian(0.81)	4	-0.001	0.018	0.017	(0.772,0.840)	0.955	< 0.001	0.009	0.012	(0.784,0.833)	0.940
4	Frank(7.93)	4	0.011	0.571	0.542	(6.886,9.010)	0.925	0.020	0.406	0.395	(7.148,8.696)	0.930
1	Clayton(3)	4	0.012	0.182	0.219	(2.591,3.449)	0.965	0.001	0.149	0.157	(2.701,3.315)	0.970
2	Gumbel(2.5)	4	-0.010	0.135	0.129	(2.247,2.752)	0.935	-0.002	0.101	0.093	(2.319,2.684)	0.905
3	Gaussian(0.81)	4	-0.001	0.017	0.017	(0.772,0.840)	0.940	< 0.001	0.012	0.012	(0.784,0.833)	0.950
4	Frank(7.93)†	4	0.035	0.570	0.550	(6.931,9.088)	0.955	-0.017	0.440	0.394	(7.143,8.689)	0.925
1	Clayton(3)	20	-0.011	0.155	0.120	(2.761,3.231)	0.850	0.003	0.105	0.079	(2.853,3.161)	0.875
2	Gumbel(2.5)	20	-0.018	0.140	0.088	(2.312,2.657)	0.835	-0.011	0.098	0.064	(2.368,2.618)	0.860
3	Gaussian(0.81)	20	< 0.001	0.016	0.012	(0.788,0.832)	0.845	0.001	0.010	0.008	(0.795,0.826)	0.855
4	Frank(7.93)	20	0.040	0.558	0.341	(7.296,8.633)	0.850	0.015	0.419	0.315	(7.317,8.543)	0.855
1	Clayton(3)	20	-0.010	0.158	0.130	(2.738,3.250)	0.875	-0.008	0.102	0.081	(2.831,3.149)	0.905
2	Gumbel(2.5)	20	-0.014	0.136	0.100	(2.292,2.686)	0.850	-0.010	0.094	0.064	(2.360,2.610)	0.870
3	Gaussian(0.81)	20	0.001	0.015	0.013	(0.784,0.835)	0.845	0.001	0.011	0.008	(0.796,0.828)	0.855
4	Frank(7.93)†	20	0.024	0.549	0.321	(7.325,8.583)	0.820	-0.009	0.404	0.312	(7.318,8.542)	0.860
Maximum Likelihood Estimation												
1	Clayton(3)	-	0.013	0.253	0.254	(2.515,3.511)	0.955	< 0.001	0.164	0.179	(2.649,3.351)	0.965
2	Gumbel(2.5)	-	-0.004	0.147	0.145	(2.211,2.781)	0.940	-0.003	0.105	0.103	(2.296,2.699)	0.920
3	Gaussian(0.81)	-	-0.001	0.019	0.019	(0.772,0.846)	0.960	< 0.001	0.013	0.013	(0.784,0.836)	0.950
4	Frank(7.93)	-	-0.016	0.643	0.643	(6.653,9.715)	0.960	-0.046	0.456	0.454	(6.995,8.774)	0.930

(†) Using transformation function $g(\theta) = \log\left(\frac{\theta+100}{100-\theta}\right)$

TABLE 10
Simulation Results for Setting 4

			$n = 200$					$n = 400$				
Cluster	Copula	L	EBias	ESE	ASE	95% Interval	ECP	EBias	ESE	ASE	95% Interval	ECP
Bayesian Estimation												
1	Clayton(3)	4	0.015	0.224	0.229	(2.577,3.477)	0.955	0.012	0.160	0.164	(2.697,3.341)	0.950
2	Gumbel(4)	4	0.018	0.220	0.215	(3.609,4.452)	0.940	-0.021	0.151	0.153	(3.684,4.285)	0.940
3	Gaussian(0.6)	4	-0.006	0.034	0.035	(0.520,0.658)	0.950	-0.004	0.026	0.026	(0.543,0.643)	0.945
4	Frank(13)	4	0.016	0.865	0.804	(11.472,14.623)	0.940	0.087	0.651	0.613	(11.899,14.300)	0.930
1	Clayton(3)	4	0.021	0.225	0.227	(2.589,3.480)	0.955	0.012	0.162	0.163	(2.700,3.339)	0.965
2	Gumbel(4)	4	0.023	0.219	0.212	(3.620,4.450)	0.940	-0.022	0.151	0.152	(3.686,4.280)	0.940
3	Gaussian(0.6)	4	-0.006	0.035	0.035	(0.519,0.658)	0.945	-0.003	0.026	0.025	(0.545,0.643)	0.960
4	Frank(13)†	4	-0.010	0.856	0.786	(11.469,14.551)	0.930	0.090	0.651	0.577	(11.973,14.236)	0.920
1	Clayton(3)	20	0.008	0.189	0.165	(2.697,3.344)	0.905	0.003	0.135	0.116	(2.780,3.234)	0.925
2	Gumbel(4)	20	0.009	0.190	0.157	(3.710,4.327)	0.895	-0.020	0.131	0.117	(3.749,4.208)	0.900
3	Gaussian(0.6)	20	< 0.001	0.028	0.023	(0.552,0.643)	0.835	-0.003	0.021	0.017	(0.562,0.629)	0.895
4	Frank(13)	20	0.009	0.876	0.600	(11.844,14.197)	0.815	0.084	0.639	0.443	(12.220,13.958)	0.850
1	Clayton(3)	20	0.012	0.193	0.168	(2.693,3.350)	0.905	0.002	0.126	0.114	(2.780,3.228)	0.915
2	Gumbel(4)	20	0.005	0.186	0.159	(3.700,4.323)	0.890	-0.015	0.130	0.112	(3.770,4.209)	0.890
3	Gaussian(0.6)	20	-0.002	0.026	0.023	(0.551,0.641)	0.865	-0.002	0.020	0.017	(0.563,0.630)	0.905
4	Frank(13)†	20	-0.021	0.837	0.594	(11.828,14.155)	0.830	0.067	0.645	0.445	(12.204,13.955)	0.845
Maximum Likelihood Estimation												
1	Clayton(3)	-	0.026	0.242	0.254	(2.524,3.521)	0.960	0.013	0.171	0.180	(2.661,3.365)	0.960
2	Gumbel(4)	-	0.022	0.234	0.236	(3.559,4.486)	0.960	-0.022	0.158	0.165	(3.654,4.302)	0.945
3	Gaussian(0.6)	-	-0.004	0.038	0.039	(0.519,0.673)	0.940	-0.003	0.028	0.028	(0.543,0.651)	0.960
4	Frank(13)	-	-0.076	0.869	0.908	(11.144,14.704)	0.960	0.054	0.658	0.647	(11.785,14.322)	0.935

(†) Using transformation function $g(\theta) = \log\left(\frac{\theta+100}{100-\theta}\right)$

Appendix C: Additional results for data analysis

C.1. Marginal distribution of six features in three health groups

The marginal density of the k-th biomedical feature in the j-th group of people, $f_{jk}(y_{jik})$, is given by

$$\frac{p_{jk}}{2k_{jk}\sigma_{jk}q_{jk}^{1/p_{jk}}B\left(\frac{1}{p_{jk}},q_{jk}\right)\left(\frac{|y_{jk}-\mu_{jk}+r_{jk}|^{p_{jk}}}{q_{jk}(s_{jk}\sigma_{jk})^{p_{jk}}(\lambda_{jk}\mathrm{sign}(y_{jik}-\mu_{jk}+r_{jk})+1)^{p_{jk}}}+1\right)^{\frac{1}{p_{jk}}+q_{jk}}},$$

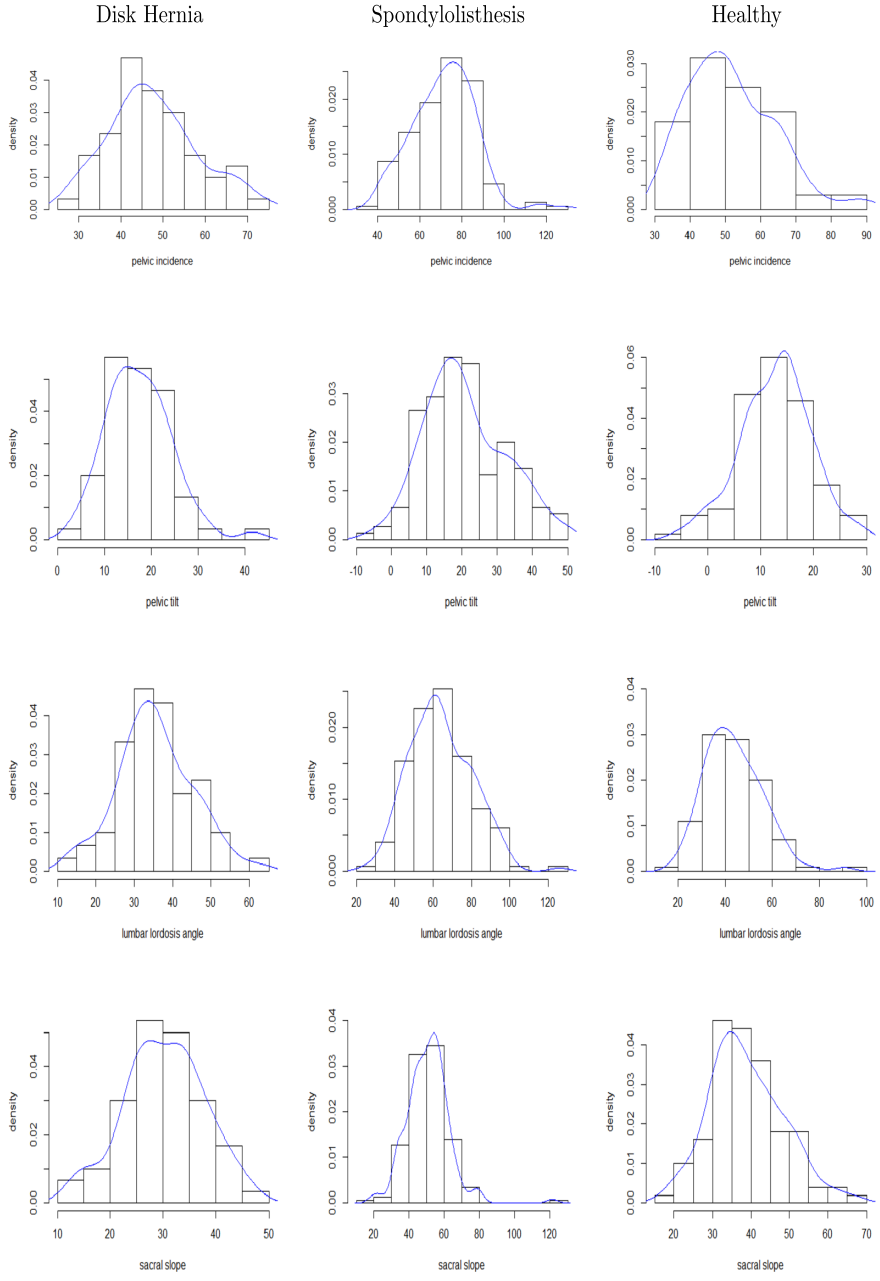
where $B(\cdot)$ is the Beta function, μ is the location parameter, σ is the scale parameter, $\lambda \in (-1,1)$ is the skewness parameter, p and q are kurtosis parameters, and r_{jk} and s_{jk} are given by

$$r_{jk} = \frac{2v_{jk}\sigma_{jk}\lambda_{jk}q_{jk}^{1/p_{jk}}B(\frac{2}{p_{jk}},q_{jk}-\frac{1}{p_{jk}})}{B(\frac{1}{p_{jk}},q_{jk})}$$

$$s_{jk} = \frac{q_{jk}^{1/p_{jk}}}{\sqrt{(3\lambda_{jk}^2+1)\frac{B(\frac{3}{p_{jk}},q_{jk}-\frac{2}{p_{jk}})}{B\left(\frac{1}{p_{jk}},q_{jk}\right)}-4\lambda_{jk}^2\frac{B(\frac{2}{p_{jk}},q_{jk}-\frac{1}{p_{jk}})^2}{B\left(\frac{1}{p_{jk}},q_{jk}\right)^2}}}.$$

TABLE 11
MLE of marginal parameters in the generalized skewed-t distributions

Groups	Features	Skewed t distribution			Normal distribution	
		μ	σ	λ	μ	σ
Disk Hernia	PI	47.711	10.581	0.238	47.638	10.608
	PT	17.431	6.942	0.314	17.398	6.958
	LL	35.522	9.677	0.101	35.464	9.686
	SS	30.261	7.495	-0.095	30.239	7.492
	PR	116.337	9.237	-0.190	116.475	9.277
	DS	2.470	5.483	-0.141	2.480	5.485
Spondylolisthesis	PI	71.538	15.056	0.065	71.514	15.059
	PT	20.821	11.436	0.279	20.748	11.468
	LL	64.100	16.346	0.256	64.110	16.342
	SS	50.993	12.207	0.204	50.766	12.278
	PR	114.599	15.517	0.087	114.519	15.528
	DS	51.897	35.119	0.629	51.897	39.974
Healthy	PI	51.401	12.577	0.635	51.685	12.306
	PT	12.789	6.739	-0.108	12.821	6.745
	LL	43.643	12.239	0.392	43.543	12.299
	SS	38.921	9.551	0.276	38.863	9.576
	PR	123.893	8.969	0.015	123.891	8.969
	DS	2.583	6.043	0.410	2.187	6.276


Disk Hernia
Spondylolisthesis
Healthy
density
pelvic incidence
pelvic tilt
lumbar lordosis angle
sacral slope

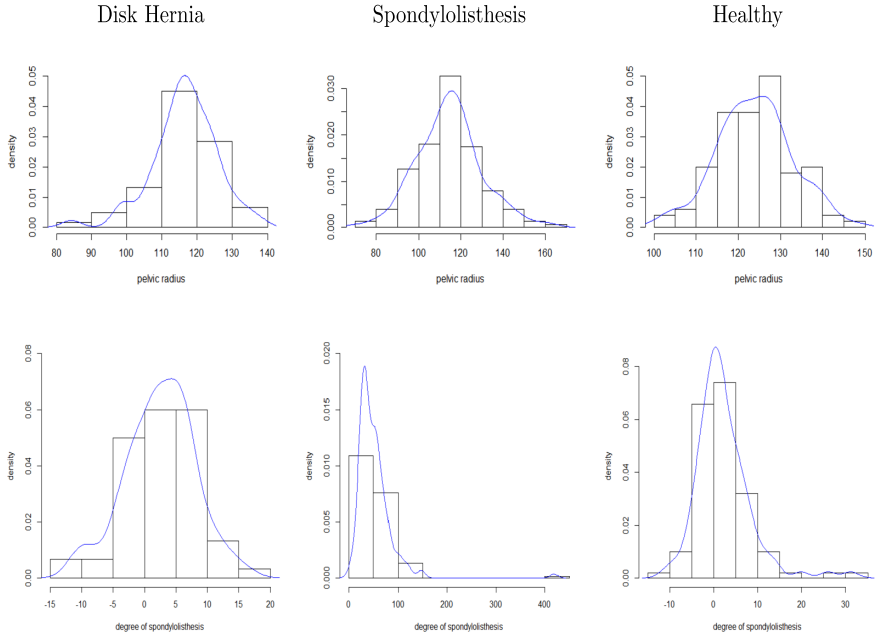
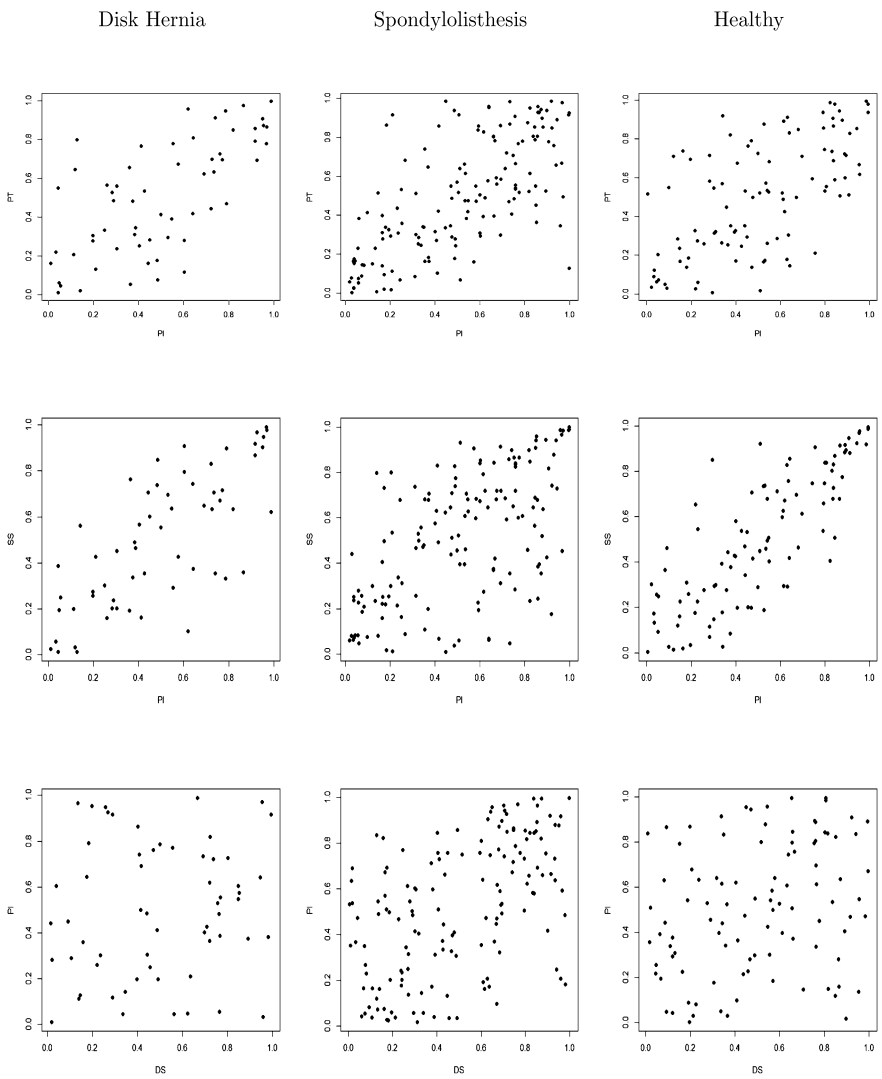


FIG 5. *Histogram of six biomedical features on three groups*

C.2. Dependence model



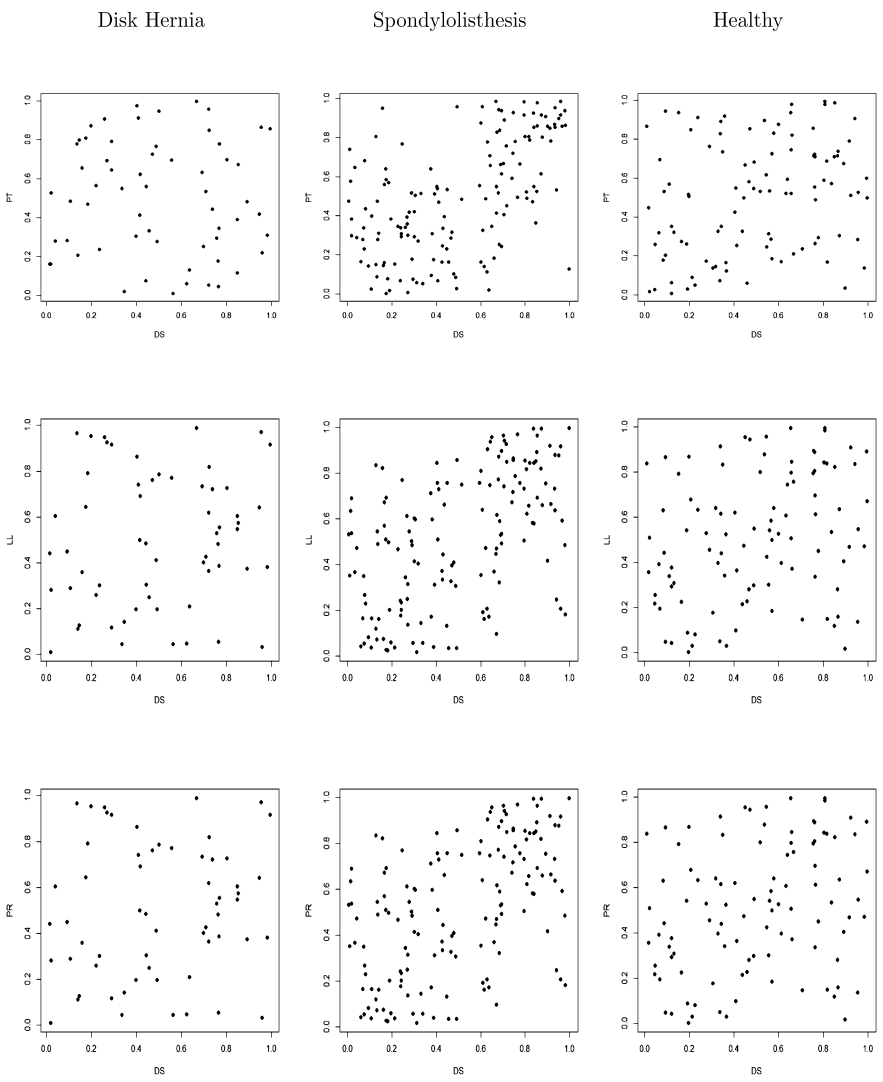


FIG 6. *Scatter plots of six pairs of bivariate dependence in 3 health groups*