A copula-based approach to joint modelling of multiple longitudinal responses with multimodal structures

Mahdiyeh, Zahra; Kazemi, Iraj & VERBEKE, Geert (2022) A copula-based approach to joint modelling of multiple longitudinal responses with multimodal structures. In: STATISTICAL MODELLING, 22(4), p. 327-348.

DOI: 10.1177/1471082X20967168
Handle: http://hdl.handle.net/1942/33502

A Copula-based Approach to Joint Modeling of Multiple Longitudinal Responses with Multimodal Structures

Zahra Mahdiyeh [1], Iraj Kazemi [1] and Geert Verbeke [2, 3]

[1] Department of Statistics, University of Isfahan, Iran

[2] Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven, B-3000 Leuven, Belgium

[3] Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Universiteit Hasselt, Diepenbeek, Belgium

---

Address for correspondence: Iraj Kazemi, Department of Statistics, University of Isfahan, Hezarjerib, Isfahan 81746, Iran .

E-mail: i.kazemi@stat.ui.ac.ir.

Phone: (+98) 313 793 4596.

Fax: (+98) 313 793 4600.

Abstract: This paper introduces a flexible modeling strategy to extend the familiar mixed-effects models for analyzing longitudinal responses in the multivariate setting. By initiating a flexible multivariate multimodal distribution, the strategy relaxes the imposed normality assumption of related random effects. We use copulas to construct a multimodal form of elliptical distributions. It can deal with the

multimodality of responses and the nonlinearity of dependence structure. Moreover, the proposed model can flexibly accommodate clustered subject-effects for multiple longitudinal measurements. It is useful when several sub-populations exists but cannot be directly identified. Since the implied marginal distribution is not in the closed form, to approximate the associated likelihood functions, we suggest a computational methodology based on the Gauss-Hermite quadrature that consequently enables us to implement standard optimization techniques. We conduct a simulation study to highlight the main properties of the theoretical part and make a comparison with regular mixture distributions. Results confirm that the new strategy deserves to receive attention in practice. We illustrate the usefulness of our model by the analysis of a real-life data set taken from a low-back pain study.

Key words: Clustered random effects; Copula function; Gaussian quadrature; Low-back pain; Multiple longitudinal responses; Multimodality; Non-linear dependence.

1 Introduction

Linear mixed-effects (LME) models have been progressively extended in recent studies to analyze some correlated data, including longitudinal or clustered, wherein a set of subjects are repeatedly measured on different conditions or periods (Laird and Ware, 1982). A routine assumption in fitting various mixed models is the normality of underlying subject effects though it may violate in practical applications. For example at the presence of outliers, the random effects may follow a distribution with heavier tails than normal. Another realistic situation involves the existence of latent

subpopulations in the data generating process especially when important categorical covariates are omitted in the fixed part of models.

Although the conventional likelihood-based estimates of fixed effects might be robust to non-normality of random effects (Butler and Louis, 1992), the same is not true for the prediction of random effects (Zhang and Davidian, 2001). Also, the ML estimate of model parameters including variance components suffers from the loss of efficiency and incorrect computation of standard errors (Pinheiro et al., 2001). In recent years, certain problems about choosing suitable distributions for the random effects and calibrating them efficiently to the observed data set have extensively discussed by several authors. In some applications, to avoid misleading inference, it is suggested that the collected measurements must be classified based on the adoption of a multimodal distribution for random effects. The choice of statistical methods in the literature to set up a multimodal structure has mostly concentrated on a mixture of multiple unimodal components. An example of mixture distributions with normal components in linear mixed-effects models is given by Verbeke and Lesaffre (1996) and described further by Verbeke and Molenberghs (2000). Another application of mixture distributions using the skew-t components is proposed by Lin (2010) to allow the accommodation of both skewness and thick tails for random effects.

In this paper, we extend the common mixed-effects models to the analysis of multiple longitudinal responses by utilizing an innovative modeling strategy to cover possible multimodality of data. We propose to apply the separate LME models to all responses and linking them by allowing a suitable multivariate multimodal distribution for the random effects of assorted responses.

In practical applications, there are several issues in using mixture distributions in

fitting LME models for multiple responses. One is the identifiability (Hennig, 2000) due to the unstructured forms of covariance matrices and a large number of unknown parameters which makes the use of basic estimation procedures complicated. Any mixture distribution to simplify the execution of computational procedures requires the convincing prior information on the choice of the true number of components. Fixing the number of components to avoid overfitting and the same distribution for all clusters are strong restrictions. Other constraints comprise the linearity of dependence structure between variables and the similarity of all marginal distributions.

The importance of dealing with these challenges motivated us to investigate alternative strategies that offer great flexibility in jointly modeling of multimodal data. We construct a new multivariate distribution by a combination of copula functions (Sklar, 1959) and a member of elliptical distributions (Fang et al., 1990) called the Double Gamma (DG) distribution. It is a suitable choice for analyzing longitudinal data that exhibit multimodality since it can cover most distributional peaks through a limited number of parameters without the need for strong prior information. Moreover, this option overtakes mixture distributions that tend to enforce additional components or parameters to capture more peaks. Furthermore, a copula is using to separate the dependence structure of a multivariate distribution from the individual marginal distributions by looking at its underlying copula form. Through the study of copula the analyst can be better aware of various associations between variables and recognize tails of the related distributions. For instance, a copula may be applied to observed data indicating correlation in the extreme tails but not elsewhere in the distribution.

The proposed strategy constructs straightforwardly a multimodal structure for the underlying random-effects based on the DG distribution and a suitable copula. It

involves several main advantages, such as i) facilitating the fitting of general mixed models with multiple responses, ii) presenting great flexibility to model the nonlinear correlations between responses, iii) allowing various marginals for each response, iv) covering several types of responses with multimodal behaviors, v) managing the impact of hidden subpopulations with different behaviors in terms of peaks that cannot be directly observed through the value of responses, vi) being useful for analyzing clustered data, vii) avoiding incorrect inference when the normality of random effects is violated or clustering of responses occurs and some important categorical covariates are omitted.

The maximum likelihood approach is used to fit the proposed multivariate mixed models. The corresponding likelihood functions appear in the non-closed form due to complicated integrals. Hence a numerical integration method using the Gauss-Hermite quadrature is employed to approximate integrals. In addition, familiar numerical optimization techniques, such as Newton Raphson, are utilized to maximize the underlying likelihoods in user-friendly software packages, such as SAS and R.

We examine the usefulness of our methodology in the analysis of low back pain (LBP) and its related disabilities, which have grown in most industrialized countries and are amongst the most frequent reasons for consulting a primary care physician. Prevention of LBP in primary stages is a major public health problem worldwide since it contributes to the prevention of disabilities because of back pain in progressive stages. To evaluate the contributions of some factors to the acute LBP, we apply our strategy to re-analyze a real-life data set taken from a prospective cohort study on low back pain (Park et al., 2010).

The remainder of this paper is organized as follows. Section 2 introduces the uni-

variate DG distribution and reports its main properties. Section 3 presents a short introduction of copula functions and how to construct a new multivariate multimodal distribution using copulas. Section 4 specifies the multivariate mixed-effects models and extends strategies to jointly analyzing multiple multimodal responses. Section 5 conducts a simulation study to evaluate the performance of our proposed model. Finally, in Section 6 we apply our methodologies to analyze a real-life data set taken from the low-back pain study and our proposed model will be compared to several competitors.

2 Double Gamma distribution

A special case of elliptical distributions (Fang et al., 1990) is the double Gamma (DG) distribution [??], defined as follows.

Definition 1 *The random variable X follows the DG distribution with parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\alpha > 0$, if its probability density function (PDF) is of the form*

$$f(x;\mu,\sigma,\alpha) = \frac{1}{2\Gamma(\alpha)\sigma^{\alpha}} |x-\mu|^{\alpha-1} \exp\left(-|x-\mu|/\sigma\right), x \in \mathbb{R}. \tag{2.1}$$

We denote $X \sim DG(\mu,\sigma,\alpha)$. By conducting basic statistical techniques the following properties hold: $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \alpha(\alpha+1)\sigma$, the kurtosis measure is $(\alpha+3)(\alpha+2)/\alpha(\alpha+1)$, and the cumulative distribution function (CDF) of X is

$$F(x) = \frac{1}{2}(1 + \mathrm{sign}(x-\mu) F_{\alpha}(|x-\mu|/\sigma)), x \in \mathbb{R},$$

where $F_{\alpha}(t)$ denotes the value of standard gamma density with parameter α integrated up to t and $\mathrm{sign}(t)$ equals 0 for $t = 0$, -1 for $t < 0$, and $+1$ for $t > 0$. The

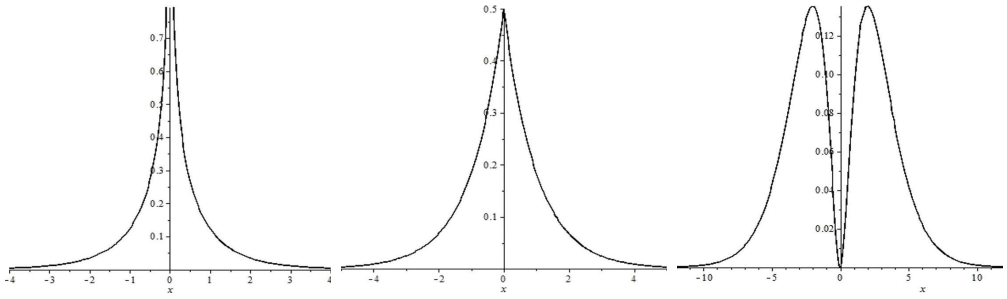

Figure 1: The density plot of $DG(0, 1, \alpha)$ for (left) $\alpha = 0.6$ (center) $\alpha = 1$ (right) $\alpha = 3$

univariate DG distribution is symmetric about μ and its shape depends on α. Figure 1 shows density plots of $DG(0, 1, \alpha)$ for some values of α. For $0 < \alpha < 1$ the density function (2.1) tends to infinity at $X = \mu$, whereas for $\alpha > 1$ it has a local minimum at μ with two modes. The special case $\alpha = 1$ refers to a generalization of the double-exponential distribution introduced by Gómez et al. (1998).

3 Multimodal Double-Gamma copula

Consider the random vector $(U_1, \ldots, U_p)^\top$ where each U_i, $i = 1, \ldots, p$, follows a uniform random variable over the unit interval $[0, 1]$. On the unit hyper-cube $[0, 1]^p$, the p-dimensional copula function C can be defined based on the joint CDF of $(U_1, \ldots, U_p)^\top$. Sklar (1959) shows that for any p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)^\top$ with joint CDF $F(x_1, \ldots, x_p)$ and continuous margins $F_1(x_1), \ldots, F_p(x_p)$, a unique copula function C exists on $\mathrm{RanF_1} \times \cdots \times \mathrm{RanF_p}$, where $\mathrm{RanF_k}$ denotes the range of F_k, such that $F(x_1, \ldots, x_p)$ can be represented through this copula and its margins as

$$F(x_1, \ldots, x_p) = C\left(F_1(x_1), \ldots, F_p(x_p)\right), \quad (x_1, \ldots, x_p)^\top \in \mathbb{R}^p. \tag{3.1}$$

Key idea is that any existing continuous multivariate distribution can be reformulated according to a copula function and conversely, using any kind of univariate margins, a multivariate distribution can be constructed by means of a copula. Using this fact, through the combination of copula functions with multimodal univariate distributions (e.g., DG), we can construct new flexible multivariate multimodal distributions that are useful for modeling multivariate multimodal measurements. To select underlying marginal distributions, we note that a unimodal distribution corresponds to an unclustered population while the existence of several distinct modes indicates a clustered population for measurements of each variable.

As already mentioned the PDF and CDF of the univariate DG distribution are in the closed-form. Thus, we introduce at least one DG distribution as the marginal of copula to consequently provide a collection of multivariate multimodal distributions. The dependence between related variables is then specified by making use of an assigned copula.

Copula functions can be determined completely by any scale-invariant dependence measure that remains unchanged under monotonically increasing transformations of marginal distributions. Therefore, we can express the dependence measures for proposed copulas in terms of the selected copula function C only. A well-known type of these measures is Kendall's τ given by

$$\tau_{_K} = 4 \int_0^1 \int_0^1 C\left(u_1, u_2\right) dC\left(u_1, u_2\right) - 1.$$

If the copula C and margins $F_1, \ldots, F_p$ are continuous and differentiable then the joint density function, corresponding to the joint distribution (3.1), is given by

$$f\left(x_1, \ldots, x_p\right) = c\left(F_1(x_1), \ldots, F_p(x_p)\right) \prod_{k=1}^{p} f_k(x_k), \quad \left(x_1, \ldots, x_p\right)^\top \in \mathbb{R}^p,$$

where $f_k(\cdot)$ is the density corresponding to the marginal CDF $F_k(\cdot)$ for $k = 1, \ldots, p$ and copula density c is the derivative of the copula C.

For illustration, in the bivariate case, to show the multimodal feature of the offered copula, as an example Figure 2 demonstrate the contour of some well-known Archimedean copulas (Nelsen, 2006) by imposing the DG and normal margins. Note that the value of Kendall's τ for all given copulas equals 0.5. These figures evidently indicate that the number of peaks is a function of the assigned margins. Clearly, the copula function reflects only a particular dependence structure and the choice of copula directly controls what parts of the implied distribution are more associated. As mentioned by Frees and Valdez (1998), the Frank copula imposes a very specific radially symmetric dependence structure. For the Clayton copula the dependence is stronger in the lower-left region than in the upper-right region while for the Gumbel copula it is stronger in the upper-right region than in the lower-left region.

4 Specification of multimodal multivariate LME models

The multivariate linear mixed-effects (MLME) modeling is an appropriate technique to describe the variation in multiple responses that are measured repeatedly over time periods for each subject in terms of a set of fixed covariates. Let p responses be measured for N subjects. For each subject i, $i = 1, 2, \ldots, N$, denote the response vector $\mathbf{y}_i^k = \left(y_{i1}^k, \ldots, y_{in_i}^k\right)^\top$ corresponds to the k-th ($k = 1, \ldots, p$) response's measurements at n_i different time periods. The traditional linear mixed-effects model assumes that

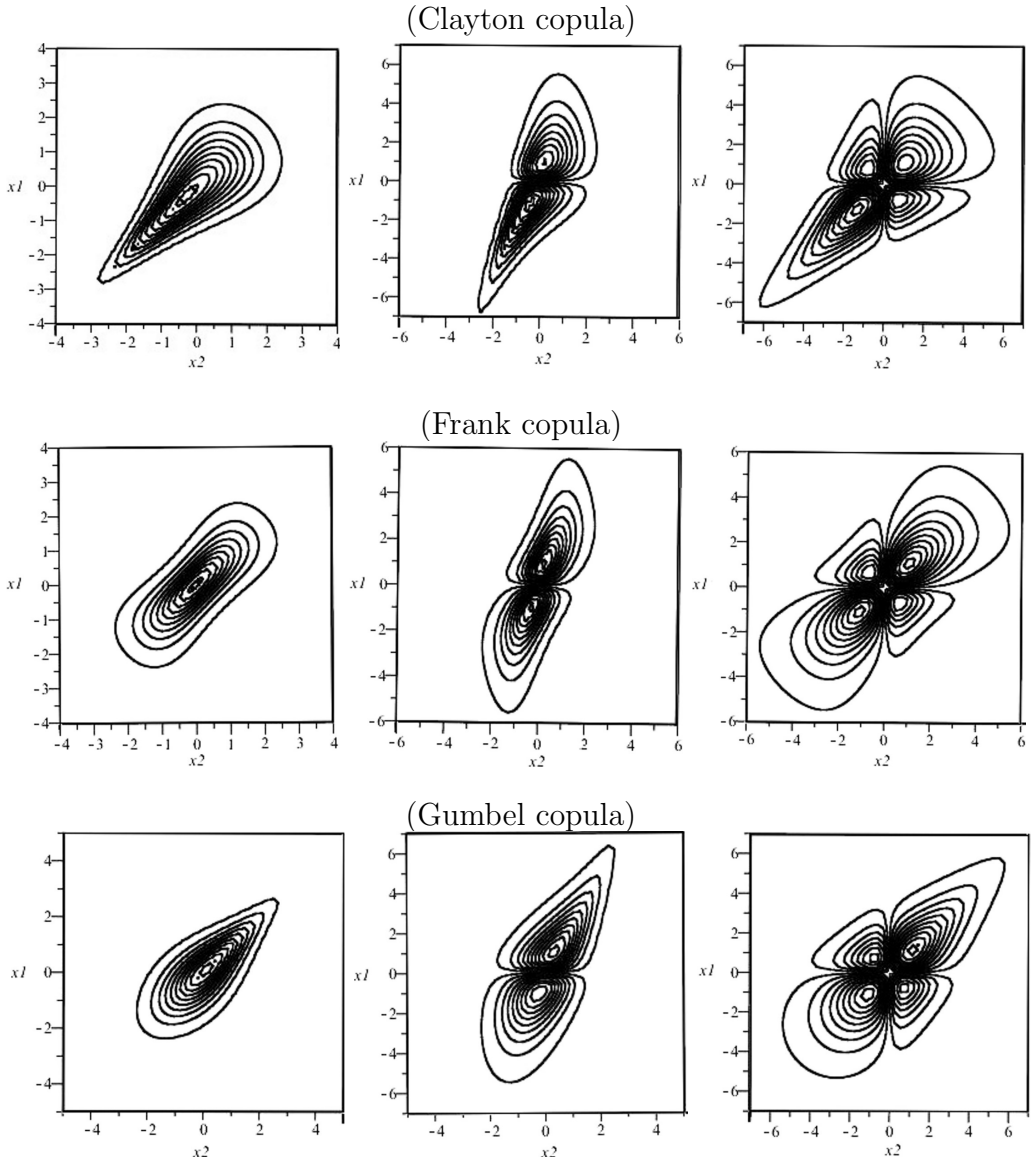

Figure 2: The density plot of the copula with the standard normal as both margins (left), the DG as one of two margins (center), and the DG as both margins (right).

the response vector $\mathbf{y}_i^k$ for each k satisfies

$$\mathbf{y}_i^k|\mathbf{b}_i^k \sim N_{n_i}\left(\mathbf{X}_i^k\boldsymbol{\beta}^k + \mathbf{Z}_i^k\mathbf{b}_i^k, \mathbf{D}_i^k\right),$$

where $\mathbf{X}_i^k$ and $\mathbf{Z}_i^k$ are $n_i \times r$ and $n_i \times q$ known covariates matrices related to the r-dimensional vector of unknown fixed regression coefficients $\boldsymbol{\beta}^k$ and the q-dimensional vector of random effects $\mathbf{b}_i^k = \left(b_{i1}^k, \ldots, b_{iq}^k\right)$, respectively, and $\mathbf{D}_i^k$ denotes an $n_i \times n_i$ covariance matrix (Fieuws and Verbeke, 2006). A usual assumption in fitting MLME models is that the measures $y_{i1}^k, \ldots, y_{in_i}^k$ for each k are conditionally independent given $\mathbf{b}_i^k$. It simply results in $\mathbf{D}_i^k = \sigma_k^2\mathbf{I}_{n_i}$, where $\mathbf{I}_{n_i}$ is a n_i-dimensional identity matrix . Furthermore, all random effects are assumed to be normally distributed. In practical applications, this naïve assumption can likely be violated if special classification exists for some responses.

In this paper, we propose an extension of MLME models which is promoted to analyze multiple clustered responses. We construct a new model by allowing the response vector $\mathbf{y}_i^k$, conditioned on the random effects $\mathbf{b}_i^k$, follows a known distribution with PDF $g^k(\mathbf{y}_i^k|\mathbf{b}_i^k;\boldsymbol{\theta}^k)$, where the vector of unknown parameters $\boldsymbol{\theta}^k$ possibly depends on some covariates. By assuming that all elements of $\mathbf{y}_i^k$ are independent, given $\mathbf{b}_i^k$, we introduce a multivariate distribution based on utilizing a proposed multimodal copula for the random effects to take into account correlated responses. It suggests that fitting a separate linear mixed-effects model for each response can appropriately specify the joint model by successively combining the multimodal copula distribution for all random effects. This strategy also allows choosing any marginal density with the bimodal/multimodal property, such as the univariate Double Gamma, for each random effect, to construct a multivariate multimodal density.

In the regression modeling methodology, the marginal expectation of responses is

commonly assumed to depend only on the covariates, i.e., $\mathrm{E}(\mathbf{Y}_i^k) = \mathbf{X}_i^k\boldsymbol{\beta}^k$. It is quite desirable to keep this property even for our proposed model by assuming that the marginal mean of each random effect is zero. Also, under offered assumptions, k-th LME model clearly dictates the marginal cross correlation structures between k-th response's measurements at two time points $j \neq s$, within subject i, as $\mathrm{Corr}(Y_{ij}^k, Y_{is}^k) = \sigma_{i(j,s)}^k / \sqrt{\sigma_{i(j,j)}^k \sigma_{i(s,s)}^k}$, where $\sigma_{i(j,s)}^k = \mathbf{Z}_{ij}^{k\top}\mathrm{Cov}(\mathbf{b}_i^k)\mathbf{Z}_{is}^k + \sigma_k^2 I(j=s)$, with $\mathbf{Z}_{ij}^{k\top}$ being the j-th row of the matrix $\mathbf{Z}_i^k$ and $I(\cdot)$ denotes the indicator variable. Moreover, the role of the dependence between the responses-specific random effects generates the correlation structure between the measurements of different responses k and l, to be measured as $\mathrm{Corr}(Y_{ij}^k, Y_{is}^l) = \sigma_{i(j,s)}^{(k,l)} / \sqrt{\sigma_{i(j,j)}^k \sigma_{i(s,s)}^l}$, where $\sigma_{i(j,s)}^{(k,l)} = \mathbf{Z}_{ij}^{k\top}\mathrm{Cov}(\mathbf{b}_i^k, \mathbf{b}_i^l)\mathbf{Z}_{is}^l$ with $\mathrm{Cov}(\mathbf{b}_i^k, \mathbf{b}_i^l)$ is computed by the defined copula and Hoeffding's Lemma (Pumi and Lopes, 2012). For illustration, consider the following simple model

$$\mathbf{y}_i^k | b_i^k \sim N_{n_i}\left(\mathbf{X}_i^k\boldsymbol{\beta}^k + b_i^k\mathbf{J}_{n_i}, \sigma_k^2\mathbf{I}_{n_i}\right),$$

where $\mathbf{J}_{n_i}$ denotes an n_i-dimensional vector of ones. For all l and $k = 1, \ldots, p$ we have

$$\mathrm{Corr}(Y_{ij}^k, Y_{is}^l) = \mathrm{Cov}(b_i^k, b_i^l) / \sqrt{\mathrm{Var}(Y_{ij}^k)\mathrm{Var}(Y_{is}^l)},$$

with $\mathrm{Var}(Y_{ij}^k) = \mathrm{Var}(b_i^k) + \sigma_k^2$, where $\mathrm{Var}(b_i^k)$ can be obtained from the marginal distribution of the random intercept b_i^k and

$$\mathrm{Cov}(b_i^k, b_i^l) = \int_0^1 \int_0^1 \frac{C(u_k, u_l) - u_k u_l}{f_k\left(F_k^{-1}(u_k)\right) f_l\left(F_l^{-1}(u_l)\right)} du_k du_l,$$

where F_k and F_l denote the marginal CDFs and f_k and f_l denote the marginal PDFs of b_i^k and b_i^l, respectively.

The above expressions reveal that the correlation between measurements of responses is directly related to the correlation between responses-specific random effects. It

also shows that measures within of each response may be correlated even if measures between two mixed responses are uncorrelated.

The inference for the vector of unknown model parameters $\boldsymbol{\Theta}$ (includes the vector parameters $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^p$, the parameters of the marginal distributions of random effects and the parameter of selected copula function C) in fitting the proposed model is based on the Log-likelihood function $\ell(\boldsymbol{\Theta}|\mathbf{y}) = \sum_i \ln[f(\mathbf{y}_i; \boldsymbol{\Theta})]$, where $f(\mathbf{y}_i; \boldsymbol{\Theta})$ is the marginal density function of the response vector $\mathbf{y}_i = (\mathbf{y}_i^1, \ldots, \mathbf{y}_i^p)^\top$ which can be obtained by integrating out the random-effects vector $\mathbf{b}_i = (\mathbf{b}_i^1, \ldots, \mathbf{b}_i^p)^\top$ as

$$\begin{aligned} f(\mathbf{y}_i; \boldsymbol{\Theta}) \quad = \quad & \int \prod_{k=1}^{p} \left\{ g^k(\mathbf{y}_i^k | b_{i1}^k, \ldots, b_{iq}^k; \boldsymbol{\theta}^k) \prod_{h=1}^{q} f_{kh}(b_{ih}^k) \right\} \\ & \times c(F_{11}(b_{i1}^1), \ldots, F_{pq}(b_{iq}^p)) db_{i1}^1 \ldots db_{iq}^1 \ldots db_{i1}^p \ldots db_{iq}^p, \end{aligned} \tag{4.1}$$

where $F_{kh}(\cdot)$ and $f_{kh}(\cdot)$ for $k = 1, \ldots, p$ and $h = 1, \ldots, q$, are PDF and CDF of the presumed marginal distribution for random effect b_{ih}^k respectively, and c is the density function of the selected copula C.

5 Maximum likelihood estimation

To carry out the inference of $\boldsymbol{\Theta}$, the direct maximization of the Log-likelihood function may involve solving complex integrals using advanced numerical techniques. In this paper, the application of a numerical technique using the Gauss-Hermite quadrature is proposed to approximate the likelihood function and to make inference on parameters in user-friendly software packages, such as SAS or R. Gaussian quadrature can be used to approximate integrals with respect to a given kernel by a weighted average of the integrand evaluated at predetermined points, called nodes. The known weights and

nodes for most kernels can be obtained by tables provided by Abramowitz and Stegun (1964) or by using an algorithm proposed by Golub (1973). Gaussian quadrature for multiple integrals are numerically complicated (Davis and Rabinowitz, 2007).

In fitting LME models for non-normal responses, the Gaussian quadrature technique can approximate the marginal density function by a weighted average of the integrand directly when the distribution of random effects is normal and the dimension of the random-effects vector is not large (Lesaffre and Spiessens, 2001; McCulloch and Searle, 2001; Gueorguieva, 2001). Thus, the estimation process is not straightforward when we use the Gaussian quadrature for fitting our proposed model. Nevertheless, using a statistical trick followed by Liu and Yu (2007) we can multiply and divide the integrand in (4.1) by a standardized multivariate normal density and reformulate the resulting function over the normal random effects $\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_i^1, \ldots, \boldsymbol{\alpha}_i^p)^\top$, where $\boldsymbol{\alpha}_i^k = \left(\alpha_{i1}^k, \ldots, \alpha_{iq}^k\right)$ for $k = 1, \ldots, p$, as

$$
\begin{aligned}
f(\mathbf{y}_i; \boldsymbol{\Theta}) \;=\; & \int \prod_{k=1}^{p} \left\{ g^k(\mathbf{y}_i^k | \alpha_{i1}^k, \ldots, \alpha_{iq}^k; \boldsymbol{\theta}^k) \prod_{h=1}^{q} f_{kh}(\alpha_{ih}^k) \right\} \\
& \times c(F_{11}(\alpha_{i1}^1), \ldots, F_{pq}(\alpha_{iq}^p)) \frac{\phi_{pq}(\boldsymbol{\alpha}_i; \mathbf{0}, \mathbf{I}_{pq})}{\phi_{pq}(\boldsymbol{\alpha}_i; \mathbf{0}, \mathbf{I}_{pq})} d\alpha_{i1}^1 \ldots d\alpha_{iq}^1 \ldots d\alpha_{i1}^p \ldots d\alpha_{iq}^p,
\end{aligned}
$$

where $\phi_{pq}\left(\boldsymbol{\alpha}_i; \mathbf{0}, \mathbf{I}_{pq}\right)$ denotes the normal density function of the pq-dimensional vector $\boldsymbol{\alpha}_i$ with the zero mean vector and the identity covariance matrix. Thus, the Gaussian quadrature technique can easily be applied to approximate the integrand

$$
\prod_{k=1}^{p} \left\{ g^k(\mathbf{y}_i^k | \alpha_{i1}^k, \ldots, \alpha_{iq}^k; \boldsymbol{\theta}^k) \prod_{h=1}^{q} f_{kh}(\alpha_{ih}^k) \right\} c(F_{11}(\alpha_{i1}^1), \ldots, F_{pq}(\alpha_{iq}^p)) / \phi_{pq}(\boldsymbol{\alpha}_i; \mathbf{0}, \mathbf{I}_{pq}).
$$

This technique only requires that the PDF of random effects can be appeared in the closed form as is available for our proposed multimodal copulas.

6 Simulation studies

We conduct two simulation studies to highlight the performance of our modeling methodology in comparison with normal and mixture models. To obtain the maximum likelihood estimation of model parameters, the Gauss–Hermite quadrature in the NLMIXED procedure of SAS is used.

To design the first simulation study a specific mixed effects model is considered for illustrative purposes. In particular, we generate 100 data sets from the bivariate LME model

$$y_{ij}^k = \beta_0^k + \beta_1^k x_j + \beta_2^k x_i^k + b_i^k + e_{ij}^k, \tag{6.1}$$

for $k = 1, 2$, $i = 1, \ldots, 100$, $j = 1, \ldots, 5$, where e_{ij}^1 and e_{ij}^2 are independent and identically distributed (iid) that follow $N(0, 1)$ and $N(0, 4)$ respectively, the covariate $x_j = j - 3$ contains values changing within subjects and the same for all subjects, and x_i^1 and x_i^2 are assumed to be the subject level covariates and are drawn uniformly in the range $(10, 20)$. True values of fixed parameters are set to $\beta_0^1 = 20$, $\beta_1^1 = 4$, $\beta_2^1 = 3$, $\beta_0^2 = 10$, $\beta_1^2 = 6$, and $\beta_2^2 = 7$. To show usefulness of the proposed multimodal copulas for accommodating the multimodality, as an illustrative case, we select the Clayton copula and generate the random intercepts b_i^1 and b_i^2 from a bivariate distribution according to the Clayton copula with margins $DG(0, 1, 3)$ and $DG(0, 2, 5)$. We set $\theta = 2$ to correspond to Kendal's $\tau = 0.5$.

To generate b_i^1 and b_i^2, we first draw variants (u_1, u_2) from the following process (Nelsen, 2006):

1. Generate two independent uniform random variables u_1 and v.

2. Set $u_2 = \left(\left(v^{-\theta/(1+\theta)} - 1\right) u_1^{-\theta} + 1\right)^{-1/\theta}$.

Then, we generate b_i^k for $k = 1, 2$, by computing the quantile function of $DG(\mu_k, \sigma_k, \alpha_k)$ given by $b_k = F_k^{-1}(u_k) = \mu_k + \sigma_k \text{sign}(u_k - 0.5) F_{\alpha_k}^{-1}(|u_k - 0.5|)$, where F_α^{-1} denotes the quantile of $Gamma(\alpha, 1)$ distribution.

Histograms of the generated random intercepts in Figure 3, clearly demonstrate the existence of two modes for each one. Also, the related scatter plot and the distribution surface of random intercepts in Figure 4 (left) and (center) show the existence of two partitions with nonlinearity of the dependence structure. For each of 100 generated

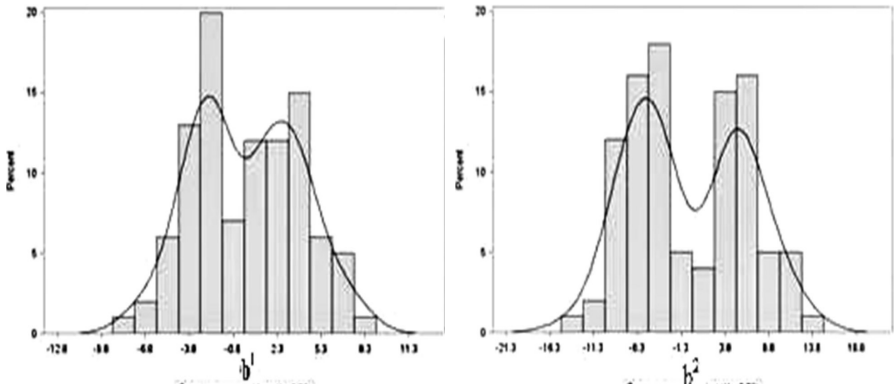

Figure 3: Histograms of the generated random effects from the Clayton copula with DG margins.

data sets, Model (6.1) was fitted by assuming that $e_{ij}^k \overset{iid}{\sim} N\left(0, \sigma_{e^k}^2\right)$ for $k = 1, 2$, and the random intercepts b_i^1 and b_i^2 distributed by

M1: The bivariate normal $N_2\left(\mathbf{0}, \mathbf{\Sigma}_b\right)$.

M2: The mixture distribution $\sum_{j=1}^2 \pi_j \phi\left(\boldsymbol{\mu}_j, \mathbf{\Sigma}_b\right)$ with $\sum_{j=1}^2 \pi_j = 1$. Here, the condition $\sum_{j=1}^2 \pi_j \boldsymbol{\mu}_j = 0$ is required to let the mean value of random effects being zero. Also, it is necessary to assume a common covariance matrix for all com-

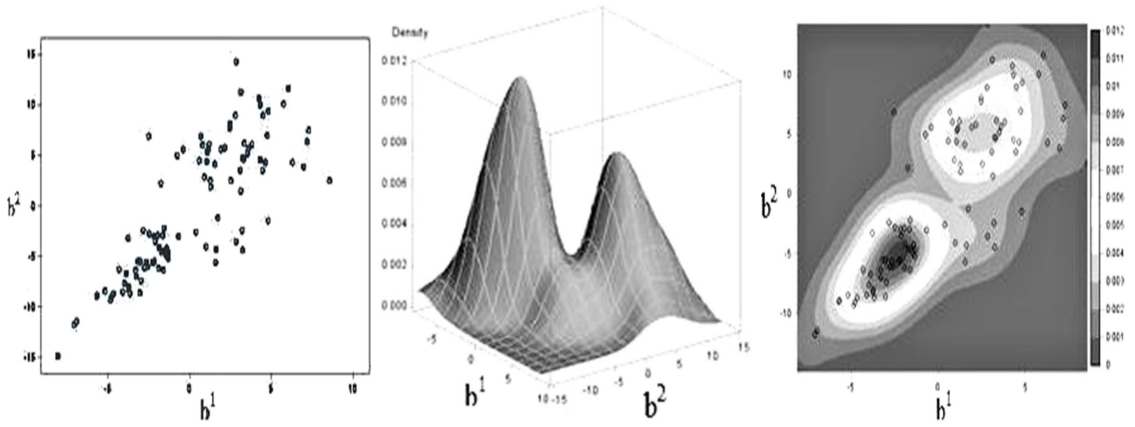

Figure 4: The scatter plot (left), the surface plot (center) and the contour plot (right) of the generated random effects from the Clayton copula with DG margins.

ponents to avoid unbounded likelihood (Böhning, 1999; Verbeke and Lesaffre, 1996).

M3: The Clayton copula with margins $DG(0, \sigma_1, \alpha_1)$ and $DG(0, \sigma_2, \alpha_2)$.

M4: The Gaussian copula with margins $DG(0, \sigma_1, \alpha_1)$ and $DG(0, \sigma_2, \alpha_2)$.

To make a comparative study, we report the parameter estimates and their standard errors of each model in Table 1(a). We also compute the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) to select the best-fitted model. These values show that the Clayton copula with the DG margins better fits for all generated data sets. It assures the correctness of our simulation process. The estimate of shape parameters for the DG margins are also significant.

A comparison of various models shows that the most parameter estimates are nearly unbiased and the same for all fitted models. In model M3, the biases and standard errors are very small and the efficiency of β_2^1 and β_2^2 estimates that are associated with the subject-level covariates x_i^1 and x_i^2 are improved in comparison with the case

Table 1: Simulation results based on 100 generated data sets of model (6.1) when the random effects have been generated from (a) Clayton copula and (b) bivariate normal. Parameter estimates (Est) and their standard errors (SE) are reported.

	M1		**M2**		**M3**		**M4**	
Parameters	Est	SE	Est	SE	Est	SE	Est	SE
(a) Multimodal random effects								
$\beta_0^1 = 20$	19.821	0.798	19.873	0.792	20.081	0.764	20.153	0.799
$\beta_0^2 = 10$	9.867	0.689	10.113	0.668	9.922	0.524	9.899	0.759
$\beta_1^1 = 4$	4.014	0.271	4.013	0.264	4.013	0.231	4.019	0.285
$\beta_1^2 = 6$	5.986	0.235	6.014	0.219	5.988	0.187	5.898	0.311
$\beta_2^1 = 3$	3.067	0.172	2.944	0.167	2.957	0.155	3.61	0.188
$\beta_2^2 = 7$	6.895	0.379	7.098	0.369	7.056	0.348	7.212	0.423
$\sigma^2_{e^1} = 1$	1.008	0.153	1.006	0.149	1.004	0.144	1.009	0.229
$\sigma^2_{e^2} = 4$	4.011	0.149	4.009	0.146	4.008	0.145	4.23	0.161
(b) Normal random effects								
$\beta_0^1 = 20$	19.935	0.349	20.069	0.613	20.132	0.745	19.878	0.512
$\beta_0^2 = 10$	9.951	0.371	10.166	0.581	10.187	0.654	10.067	0.423
$\beta_1^1 = 4$	3.996	0.093	4.005	0.105	4.012	0.116	3.994	0.101
$\beta_1^2 = 6$	5.997	0.106	6.009	0.109	5.991	0.114	6.007	0.107
$\beta_2^1 = 3$	3.019	0.098	2.948	0.161	3.065	0.192	2.963	0.123
$\beta_2^2 = 7$	7.008	0.027	6.949	0.192	7.073	0.193	7.045	0.179
$\sigma^2_{e^1} = 1$	0.999	0.127	1.008	0.229	0.988	1.241	1.009	0.221
$\sigma^2_{e^2} = 4$	4.004	0.214	4.006	0.248	4.011	0.254	4.006	0.232

of normally distributed random-effects. This evidence shows that the use of the proposed multimodal multivariate distribution based on copulas is more plausible and the adoption of incorrect assumptions (e.g., normality) for random-effects distributions may reduce the efficiency of the regression parameter estimates. Similar findings have been addressed in McCulloch and Neuhaus (2011) for the estimate of intercepts. Our results show that the efficiency of β_0^1 and β_0^2 estimates may be degraded when the random effects distribution is far from normal while the normality is assumed. The efficiency of β_1^1 and β_1^2 estimates are nearly equal in all models which shows that the distribution of random effects does not considerably influence the estimate of

longitudinal effects. This fact was already addressed by Verbeke and Lesaffre (1996) in a specific linear mixed-effects model. The estimate of scale parameters shows a discrepancy for all fitted models but are not comparable because of owning different scales.

Figure 4(right) displays the scatter plot of estimated random intercepts from model M3, with the super-imposed contour plots of the fitted Clayton copula with DG margins. It demonstrates that the additional flexibility afforded by the proposed distribution is sufficient to capture quite accurately the true multimodal underlying feature of the random intercepts.

Afterward, we design the second simulation by assuming that the random intercepts follow a bivariate normal and illustrate that the proposed model still provides reasonable estimation results. This reveals that the proposed model deserves to be used in practical applications as a reliable alternative even if the classical model is correct. Specifically, we let b_i^1 and b_i^2 being generated by a bivariate normal distribution with mean $\mathbf{0}$ and the covariance matrix $\begin{pmatrix} 1 & 1 \\ 1 & 4 \end{pmatrix}$.

For each of 100 generated data sets, we again, fit Model (6.1) by assuming that b_i^1 and b_i^2 follow M1-M4. As expected, the AIC and BIC values choose the normal as the best fitted model. Results are given in Table 1(b) show that the parameter estimates are, for the most parameters, relatively unbiased in all models. The estimate of fixed-effects parameters in model M4 is extremely close to model M1. The Gaussian copula evidently can cover the linear dependence structure between the generated normal random intercepts. In this way, there is no efficiency loss associated when using the Gaussian copula. The comparison of findings for M3 and M4 shows that

changing of the copula can increase the bias and standard error of some parameter estimates. As a result, proper specification of the random-effects distribution and the copula function are important. The choice of better fitness depends necessarily on using various copulas with specific margins to construct suitably a joint random effects-distribution.

7 Data analysis strategies for the low back pain study

We reanalyze a real-life data set, which is taken from a prospective cohort study on low back pain (Park et al., 2010), to illustrate the usefulness of our proposed MLME model. The main aim of the study was to explore the effect of a treatment package composed of herbal medicine, acupuncture, bee venom acupuncture, and a Korean version of spinal manipulation (Chuna) on low back pain. We show that our methodology is useful when a complex structure involving the multimodality of bivariate responses is to be analyzed.

7.1 Data description

The institutional review boards (IRBs) of both the University of North Carolina and Jaseng hospital in Korea has organized the low back pain study. The collection of measurements was from November 2006 to October 2007. In total 127 patients were selected. They had not previously treated for low back pain at the Jaseng hospital. Some specific cases were deleted from the sample due to some exclusion criteria, such as back pain caused by non-spinal or soft tissue issues, pregnancy, spinal tumor, rheumatoid arthritis, the history of back surgery, vertebral fracture,

dislocation, suspected concurrent severe neurological symptoms, and major organ. The control of treatment was at baseline and followed-up measurements were at weeks 4, 8, 12, 16, 20 and 24. Patients were 34.7±8.4 years old (mean ± standard deviation) with 41.6 percent female.

In our modeling process, we will jointly analyze the visual analog scale (VAS) (0-10) of back pain (Jensen et al., 1986) and the Oswestry Disability Index (ODI) (Beurskens et al., 1996). The model includes several medical and demographic factors such as the patients' age, sex, body mass index, surgery recommendation (0=recommended and 1=not recommended), baseline measures of two responses and the quality of life variables according to different subcategories mental health and physical health. These two main summary measures are aggregated from 8 sub-scale items (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional and mental health) of the SF-36 Health-Related Quality of Life Questionnaire (Ware et al., 1995) and are defined as scores ranging from 0 to 100 wherein the higher score indicates an improved level of health.

7.2 Data analysis

The individual profiles plot, not shown here, shows that both ODI and VAS levels increase over time for most patients and substantial inter-patient variation exists. Thus, we first fit two separate univariate normal random-intercepts for responses $(y_{ij}^1, y_{ij}^2) \equiv (VAS_{ij}, ODI_{ij})$ as

$$y_{ij}^k = \mathbf{X}_{ij}\boldsymbol{\beta}^k + b_i^k + e_{ij}^k, \tag{7.1}$$

for $i = 1, \ldots, 127$ and $j = 1, \ldots, 6$, where $e_{ij}^k \overset{iid}{\sim} N\left(0, \sigma_{e^k}^2\right)$ and $b_i^k \overset{iid}{\sim} N\left(0, \sigma_k^2\right)$ for $k = 1, 2$. We observed that any evidence of interaction of covariates by time was not

significant and hence only main effects are included for both responses. Also, to select a suitable covariance structure for within subject residual terms in the vector $\mathbf{e}_i^k = (e_{i1}^k, \ldots, e_{i6}^k)^\top$, we fitted four LME models based on four different structures, including the unstructured (UN), the first-order autoregressive (AR(1)), the first-order antedependence (ANTE(1)) and the standard variance component (VC) (Thiébaut et al., 2002). According to the AIC values, the VC with structure $\sigma^2_{e^k}\mathbf{I}_6$, which assumes independence of the residual measurements, seems to be the best one between others.

The empirical correlation between the measures of two responses ODI and VAS was 0.62, suggesting that a bivariate model may significantly be fitted better than two separate univariate LME models. Results of the fitted bivariate model show that the correlation between the prediction of random intercepts of two separated models for the ODI and the VAS is close to one (0.83), which may suggest that a model with one shared random intercept should also fit well. Comparison of two fitted models with shared and separated random-intercepts show that the sharing strategy makes no better fit based on the smallest AIC and BIC values.

A preliminary descriptive analysis, based on some categories factors, shows that a hidden classification may exist in the structure of collected data. Furthermore, based on the histograms of the predicted random intercepts, shown in Figure 5, we observed that the random intercept associated with VAS deviates from the normality and multimodal shape, whereas the random intercept associated with ODI may be normally distributed. Also, the related density surface and scatter plot of the predicted intercepts, shown in Figure 6 (left) and (center), obviously reveal that the joint distribution of intercepts (b_i^1, b_i^2) may be bimodal.

The above evidence motivates us to examine the ability of our proposed strategy

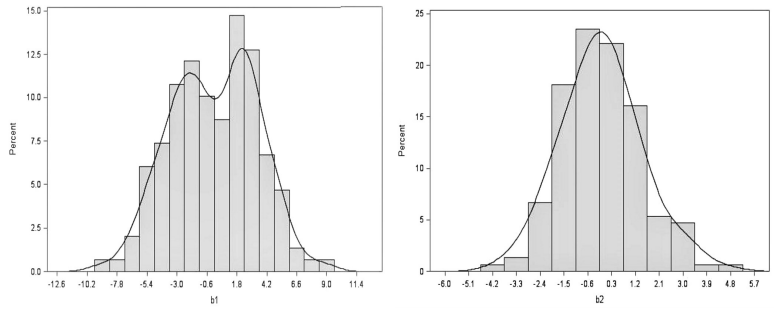

Figure 5: Histograms of the estimated random intercepts from model M5 in low back pain study.

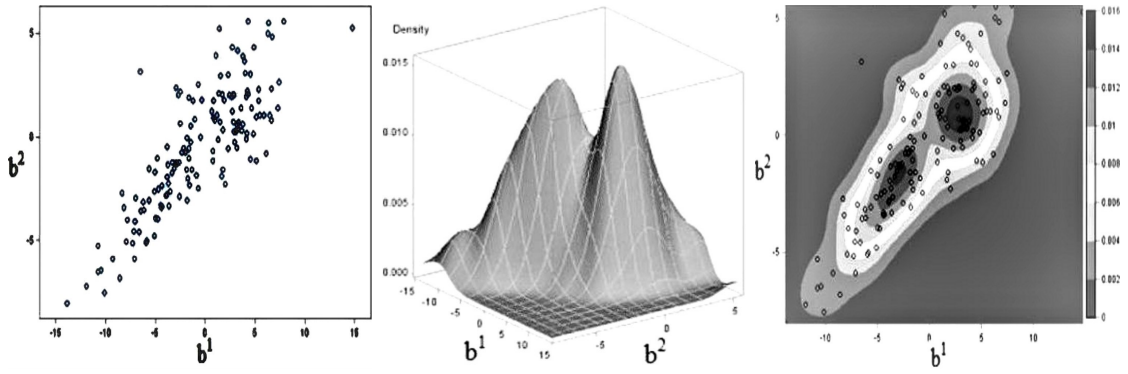

Figure 6: (left)The scatter plot (center) The surface plot of the estimated random intercepts from model M5. (right) The contour plots of the Clayton copula with DG margins in the low back pain study.

to classifying patients. A strong dependence observed in the lower-left region of the predicted intercepts. It may be cover by the Clayton copula. Thus, we specify a multimodal bivariate distribution for the random intercepts by utilizing the univariate DG distribution for each random intercept and the Clayton copula to join them. Because the joint distribution of random intercepts is multimodal, we fit a bivariate LME model specified by the finite mixture distribution with normal components and the bivariate DG distribution for random intercepts.

For comparison, we fit the mixed-effects model (7.1) by assuming that the random

intercepts be distributed as the already introduced specification for Models M1–M3 and two following models.

M4: The Clayton copula with margins $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$.

M5: The Clayton copula with margins $DG(0, \sigma_1, \alpha_1)$ and $N(0, \sigma_2^2)$.

Table 2 shows the estimation results. We report the SAS code for M5, as an example, in Appendix. Values of model selection criteria show that M5 is the best-fitting model while M3 is the second-best one. The dependence parameter estimate of Clayton copula and the shape parameter estimate of the DG margin in M5 are significant. We observe that the standard errors of fixed effects associated with most covariates in the normal model M1 are larger than those models assuming multimodality and are smaller in the selected model M5. The same is true for the estimate of random-effects variances.

Figure 6 (right) displays the scatter plot of the predicted intercepts with the superimposed contour plots of the fitted model M5. This figure indicates that the more flexibility, offered by our proposed model, is sufficient to capture the multimodality and the nonlinear dependence between the random intercepts.

7.3 Medical results based on the best fitted model

Results of the best-fitted model show that both responses significantly unrelated to the gender and surgery recommendation. A large amount of pain or disability at

Table 2: Estimate (standard error) of model parameters under the fitted models M1–M5 for the low back pain study. ODI= Oswestry Disability Index; VAS= Visual Analogue Scale.

	Estimate (standard error)				
	M1	**M2**	**M3**	**M4**	**M5**
Fixed effects parameters					
Baseline ^{ODI}	0.59 (0.67)	0.45 (0.54)	0.43 (0.52)	0.48 (0.59)	0.43 (0.33)
Female^{ODI}	0.38 (0.91)	0.53 (0.53)	0.46 (0.46)	0.36 (0.62)	0.54 (0.42)
Age^{ODI}	0.92 (0.74)	0.85 (0.52)	0.75 (0.42)	0.78 (0.57)	0.72 (0.36)
Body mass index^{ODI}	1.99 (1.39)	2.09 (1.26)	1.84 (1.11)	2.41 (1.29)	2.89 (0.91)
Surgery recommendation^{ODI}	0.13 (1.82)	0.33 (1.19)	0.52 (1.13)	0.46 (1.25)	0.42 (1.07)
Physical health^{ODI}	−0.64 (0.84)	−1.44 (0.76)	−1.09 (0.53)	−1.14 (0.69)	−1.05 (0.31)
Mental health^{ODI}	−0.78 (0.64)	−1.34 (0.46)	−1.11 (0.23)	−1.17 (0.47)	−1.98 (0.14)
Baseline^{VAS}	0.09 (0.09)	0.09 (0.07)	0.03 (0.06)	0.07 (0.08)	0.09 (0.05)
Female^{VAS}	0.78 (1.35)	0.45 (1.27)	0.76 (1.11)	0.38 (1.31)	0.58 (1.04)
Age^{VAS}	0.27 (0.37)	0.34 (0.29)	0.59 (0.14)	0.51 (0.32)	0.64 (0.11)
Body mass index^{VAS}	0.03 (0.89)	0.19 (0.64)	0.09 (.34)	0.07 (0.68)	0.13 (0.29)
Surgery recommendation^{VAS}	1.46 (1.91)	1.92 (1.36)	1.55 (1.28)	1.25 (1.56)	1.56 (1.13)
Physical health^{VAS}	−0.07 (0.09)	−0.02 (0.06)	−0.08 (0.05)	−0.01 (0.06)	−0.09 (0.04)
Mental health^{VAS}	−0.55 (0.36)	−0.58 (0.19)	−0.86 (0.16)	−0.52 (0.28)	−0.95 (0.13)
Variance components					
$Var(b_1)$	16.74 (5.21)	16.01 (4.08)	14.96 (3.82)	15.91 (3.92)	14.15 (3.27)
$Var(b_2)$	16.61 (5.82)	14.21 (4.30)	12.52 (4.41)	15.13 (4.35)	12.11 (4.22)
$\sigma^2_{e^1}$	1.97 (0.027)	1.93 (0.026)	1.94 (0.027)	1.92 (0.025)	1.91 (0.024)
$\sigma^2_{e^2}$	1.46 (0.35)	1.39 (0.23)	1.36 (0.16)	1.45 (0.22)	1.34 (0.14)
Model selection criterion					
AIC	7791	7403	7307	7665	7273
BIC	7845	7458	7363	7612	7326

the beginning of the study without any intervention may be significantly associated with the degree of patient improvement according to the VAS and ODI changes in the follow-up values. This result has been already addressed by previous researchers (Karaman et al., 2011).

As expected, our analysis shows that a significant positive relationship exists between

age and both ODI and VAS. It means that an increase in age leads to higher disability and pain severity. We can show that the risk of disabling back pain rises in older ages. Accordingly, it is highly recommended to find a desirable policy for LBP in elderly patients. Another finding of our study is the significant positive relationship between BMI and both ODI and VAS. It means that the pain severity and the risk of disabling back pain rise in the overweight category. We also observed a relationship between the disability of patients with chronic pain and a significant negative relationship between physical and mental health with both ODI and VAS. These show that there is a correlation between the reduction in pain and the improvement in disability simultaneously with an increase in the quality of life when excluding the effect of other factors. The negative correlation of quality of life with chronic low back pain is in concordance with other studies (e.g., Di Iorio et al. (2007)). As expected, decreased disability also had an impact on the physical and mental components score of the quality of life given the bilateral relationship. However, we omitted it in our study and only investigated the effect of quality of life components on the pain severity and disability of patients.

8 Discussion

The basic requirement of the analysis of multiple responses in mixed-effects modeling is to construct a multivariate distribution from desired marginal distributions with a given dependence structure. A flexible tool is a copula model, which extends the multivariate linear mixed-effects models in a way that the dependence structure between multiple correlated responses is not necessarily limited to be linear. Furthermore,

the proposed methodology is useful when the marginal distributions of responses are non-normal. Besides, it is convenient to model heterogeneous data with some unobserved subpopulations. The strategy was to specify the separate LME models with random intercepts with each response distributed as the DG. Then, we used a copula function to join the random intercepts of responses. Since the copula eliminates the effect of univariate margins from their dependence structure, the strategy causes greater flexibility in designing mixed-effects models that are applicable in real empirical applications. It is also helpful when several peaks exist in joint or each one of the marginal distributions of responses but is not directly detectable. An interesting extension is to allow controlling of the unobserved subject heterogeneity by letting some regression coefficients being heterogeneous across subjects and consequently fit random slopes models. This is a topic of our future research.

We should mention that our proposed strategy to jointly model clustered data differs in methodology in comparison to other tools in the literature. In the analysis of binary and continuous responses, Gueorguieva and Agresti (2001) propose a correlated probity model without using the copula approach. Lambert and Vandenhende (2002) propose an adaptable way of modeling the dependence between the components of non-normal multivariate longitudinal-data by using the copula but without any notice on the multimodal structure of data.

Also, to relax the normality assumption in the multivariate longitudinal settings Nai Ruscone and Osmetti (2017) introduce the implementation of the D-vine copula function. Our proposed strategy uses familiar copulas to illustrate how two types of dependence between variables and over time appear in the multivariate-clustered longitudinal data framework.

Our proposed strategy can be used suitably as an attractive alternative to the multivariate mixture modeling since it can cover multimodality via a fewer number of parameters without employing any selection method for the number of mixture components. It is not however applicable to research studies with aims concentrated only on classification, clustering, or discrimination of population under investigation. Although our simulation studies show that the proposed strategy is convenient for the analysis of multimodal correlated data, further research is required to illustrate the strengths and weaknesses of the strategy when comparing with finite mixture models.

We employed a numerical integration technique using the Gauss-Hermite quadrature to carry out statistical inference through the maximum likelihood approach. Although most commonly available software packages, such as SAS or R, are useful to implement the technique, in our experience fitting multiple mixed-effects models with several responses is somehow complicated. The optimization algorithms may terminate due to non-convergence. The analyzer requires some carefully selected initial values. Thus, for future work, we suggest performing other estimation approaches based on Bayesian computation. They can be easily implemented inaccessible software packages, such as OpenBUGS, STAN, or JAGS in R.

Appendix

Note that the variable 'lastid' is set to 1 for the last record of the same patient and to zero otherwise.

```
proc nlmixed data=pain qpoints=30;
parms ...;
```

```
bounds s2a1>0,s2a2>0,s2e1>0,s2e2>0;
if var=1 then do; mu=b01+b11*female + ...+ a1; s2e=s2e1; end;
else if var=2 then do; mu=b02+b12*female + ...+ a2; s2e=s2e2; end;
\lnlike=-0.5*\ln(s2e)-(y-mu)**2/(2*s2e);
z1=a1/sqrt(s2a1);
p1=0.5/(sqrt(s2a1)*gamma(alpha))*abs(z1)**(alpha-1)*exp(-(abs(z1)));
F1=0.5+0.5*sign(z1)*CDF('gamma',abs(z1),alpha,1); if F1>0.9999 then F1=0.9999;
p2=pdf('Normal', a2, 0, s2a2);
F2=cdf('Normal', a2, 0, s2a2); if F2>0.9999 then F2=0.9999;
\lnclaytonden=\ln(theta+1)+\ln(p1)+\ln(p2) -(theta+1)*(\ln(F1)+\ln(F2))
  -(1/theta+2)*\ln(F1**(-theta)+F2**(-theta)-1);
\lnnormalden=-a1**2/2-a2**2/2;
if lastid=1 then \lnlike=\lnlike+\lnclaytonden-\lnnormalden;
model y ~ general(\lnlike);
random a1 a2 ~ normal([0,0],[1,0,1]) subject=patient; run;
```

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation.

Beurskens, A., De Vet, H., and Koke, A. (1996). Responsiveness of functional status in lbp: comparison of different instruments. *Pain*, **95**, 71–76.

Böhning, D. (1999). *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others*, volume 81. CRC press, London.

Butler, S. M. and Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine*, **11**(14-15), 1981–2000.

Davis, P. J. and Rabinowitz, P. (2007). *Methods of numerical integration*. Courier Corporation.

Di Iorio, A., Abate, M., Guralnik, J. M., Bandinelli, S., Cecchi, F., Cherubini, A., Corsonello, A., Foschini, N., Guglielmi, M., Lauretani, F., et al. (2007). From chronic low back pain to disability, a multifactorial mediated pathway: The inchianti study. *Spine*, **32**(26), E809.

Fang, K. T., Kotz, S., and NG, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, New York.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, **62**, 424–431.

Frees, E. W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, **2**(1), 1–25.

Golub, G. H. (1973). Some modified matrix eigenvalue problems. *Siam Review*, **15** (2), 318–334.

Gómez, E., Gomez-Viilegas, M., and Marín, J. (1998). A multivariate generalization of the power exponential family of distributions. *Communications in Statistics-Theory and Methods*, **27**(3), 589–600.

Gueorguieva, R. V. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, **1**, 177–193.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association*, **96**(455), 1102–1112.

Hennig, C. (2000). Identifiablity of models for clusterwise linear regression. *Journal of Classification*, **17**, 273–296.

Jensen, M. P., Karoly, P., and S., B. (1986). The measurement of clinical pain intensity: a comparison of six methods. *Pain*, **27**, 117–126.

Karaman, H., Tüfek, A., Kavak, G., Kaya, S., Yildirim, Z. B., Uysal, E., and Çelik, F. (2011). 6-month results of transdiscal biacuplasty on patients with discogenic low back pain: preliminary findings. *International Journal of Medical Sciences*, **8** (1), 1.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Lambert, P. and Vandenhende, F. (2002). A copula-based model for multivariate non-normal longitudinal data: analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine*, **21**(21), 3197–3217.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **50**(3), 325–335.

Lin, T. I. (2010). Robust mixture modeling using multivariate skew t distributions. *Statistics and Computing*, **20**(3), 343–356.

Liu, L. and Yu, Z. (2007). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, **27**(16), 3105–3124.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, **26** (3), 388–402.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, linear and mixed models*. Wiley, New York.

Nai Ruscone, M. and Osmetti, S. (2017). Modelling the dependence in multivariate longitudinal data by pair copula decomposition. *In: Ferraro M. et al. (eds) Soft Methods for Data Science. SMPS 2016. Advances in Intelligent Systems and Computing, vol 456*, pages 373–380.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer Science & Business Media.

Park, J., Shin, J., Choi, Y., Youn, Y., Lee, S., Kwon, S., Lee, H., Kang, M., Ha, I., and Shin, I. (2010). Integrative package for low back pain with leg pain in korea: a prospective cohort study. *Complementary Therapies in Medicine*, **18**, 78–86.

Pinheiro, J. C., Liu, C., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**(2), 249–276.

Pumi, G. and Lopes, S. (2012). Parameterization of copulas and covariance decay of stochastic processes with applications. *arXiv preprint arXiv*, **1**, 1204.3339.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, **8**, 229–231.

Thiébaut, R., Jacqmin-Gadda, H., Chêne, G., Leport, C., and Commenges, D. (2002). Bivariate linear mixed models using sas proc mixed. *Computer Methods and Programs in Biomedicine*, **69**(3), 249–256.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217–221.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer-Verlag, New York.

Ware, J. E., Kosinski, M., Bayliss, M. S., McHorney, C. A., Rogers, W. H., and Raczek, A. (1995). Comparison of methods for the scoring and statistical analysis of sf-36 health profile and summary measures: summary of results from the medical outcomes study. *Medical Care*, **33**, 264–279.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**(3), 795–802.