Analysis of paediatric visual acuity using Bayesian copula models with sinh-arcsinh marginal densities

Julian Stander*[1], Luciana Dalla Valle[1], Charlotte Taglioni[2], Brunero Liseo[3], Angie Wade[4] and Mario Cortina-Borja[4]

[1]School of Computing, Electronics and Mathematics, University of Plymouth, UK
[2]Department of Statistical Sciences, University of Padova, Italy
[3]Department of Methods and Models for Territories, Economy and Finance, University of Rome La Sapienza, Italy
[4]Great Ormond Street Institute of Child Health, University College London, UK

Accepted for Publication in *Statistics in Medicine*, 20th March 2019

Abstract

We analyse paediatric ophthalmic data from a large sample of children aged between 3 and 8 years. We modify the Bayesian additive conditional bivariate copula regression model of Klein and Kneib [27] by using sinh-arcsinh marginal densities with location, scale and shape parameters that depend smoothly on a covariate. We perform Bayesian inference about the unknown quantities of our model using a specially tailored Markov chain Monte Carlo algorithm. We gain new insights about the processes which determine transformations in visual acuity with respect to age, including the nature of joint changes in both eyes as modelled with the age-related copula dependence parameter. We analyse posterior predictive distributions to identify children with unusual sight characteristics, distinguishing those who are bivariate, but not univariate outliers. In this way we provide an innovative tool that enables clinicians to identify children with unusual sight who may otherwise be missed. We compare our simultaneous Bayesian method with the two-step frequentist generalized additive modelling approach of Vatter and Chavez-Demoulin [56].

Keywords: Bayesian dependence modelling, Conditional copulas, Generalized additive models for location, scale and shape (gamlss), Sinh-arcsinh distribution, Visual acuity

*J.Stander@plymouth.ac.uk

1 Introduction

An important aspect of paediatric epidemiology is the construction of age-related standards referring to the proportion of children who have reached specific developmental milestones at particular ages, for example smiling for the first time, taking a first step, and menarche. Such standards are often used as part of screening programmes, whose aim is to identify children at sufficient risk of a specific disorder who may benefit from further investigation or direct preventive action to avoid death or disability and to improve their quality of life [42]. In ophthalmic epidemiology, one is interested in, for instance, changes in visual capacity defined as the resolving power or the ability to separate objects of the eye. Visual acuity reflects the spatial resolution of the visual system and is defined by the smallest retinal image which can be appreciated at a given distance [26].

Monitoring and promptly detecting changes in visual acuity at an early age is particularly important for the long-term visual health of individuals. Studies of age-related changes in children developing normally have shown that visual acuity evolves considerably in the first ten years of life. Visual acuity, as assessed by Snellen optotype charts, is measured on an ordinal scale, and outcomes of 3/3, 3/4–5, 3/6, 3/9, 3/12, 3/18, 3/24, 3/36 or 3/60 are based on the assumption that the resolving power, in a healthy adult, is 1 minute of arc. These quotients are termed Minimum Angle of Resolution (MAR). The numerator of the acuity outcome refers to the test distance in meters (usually 6 for adults, 3 for children); the denominator is the distance in meters at which that size of letter would subtend 1 minute of arc. Testing of children takes place from a distance of 3 meters, so an outcome of 3/3 refers to normal visual acuity and outcome of 3/4–5 or worse indicates that the person has impaired vision. A Snellen score of 6/6 meters, or 20/20 as commonly expressed in feet, means that the subject can resolve details as small as 1 minute of visual angle. Usually, visual acuity is measured on the $\log 10$ scale of MAR and a logMAR value of 0 means adequate vision in this context. For instance, logMAR values of -0.3 and 0.3 mean that details as small as 1/2 and 2 minutes of visual angle can be resolved. Thus, negative values of logMAR indicate better visual resolution capacity than positive ones.

Due to the changes in visual acuity in young children, clinicians and researchers advocate the development of age criteria with monitoring and diagnostic purposes. Sonksen and colleagues [51] measured linear visual acuity with the Sonksen logMAR test on children from a population-based normative sample in order to construct age-related standards for children from 2 years 9 months to 8 years. This test is recognised [13] as superior for measuring visual acuity as it is better able than the Snellen scale to fit the rapid changes in visual acuity in children, and is now considered the gold standard [20].

Visual acuity matures through childhood up to about age 7 or 8 years. It is important that a child who is visually impaired is identified as early as possible in order to start corrective treatment [61]. The aim of this paper is to model bivariately the age-related changes in paediatric visual acuity in both eyes together. The only existing reference ranges are for each eye separately (monocular) or overall (binocular) vision. By contrast, bivariate references can identify children whose vision, both monocular and binocular, may appear to be normal, but for whom the difference between the binocular assessment for each eye may indicate a clinical problem and so may allow particular conditions to be identified. Therefore, our approach allows children with unusual vision for their age, in either one or both eyes, to be identified.

We considered logMar [47] measurements from each eye of children from the Cambridge city health district who were tested within the place of recruitment, for example their school, between January 2000 and June 2001; full details of the data collection process are given in Wade and colleagues [62]. In particular, we worked with a sample of 2721 children with ages that range from 3.01 to 7.99 years. The data are shown in Figure 1 with colour being used to represent age group. Figure 1 was produced using the `ggplot2` package [64] running in version 3.5.1 of R [43]. The `YlOrRd` sequential colour palette from the `RColorBrewer` package [38] was chosen so that the differences between children of different ages are visible. The expected pattern of lower logMar values (improved visual acuity) in older age groups can be seen in Figure 1.

In order to gain flexibility, we modify the additive conditional bivariate copula regression model of Klein and Kneib [27] by using sinh-arcsinh marginal densities with location, scale and shape parameters that depend smoothly on a covariate. A bivariate copula [22, 37] is a joint probability density function with uniform marginals on $[0, 1]$. An example of such a copula

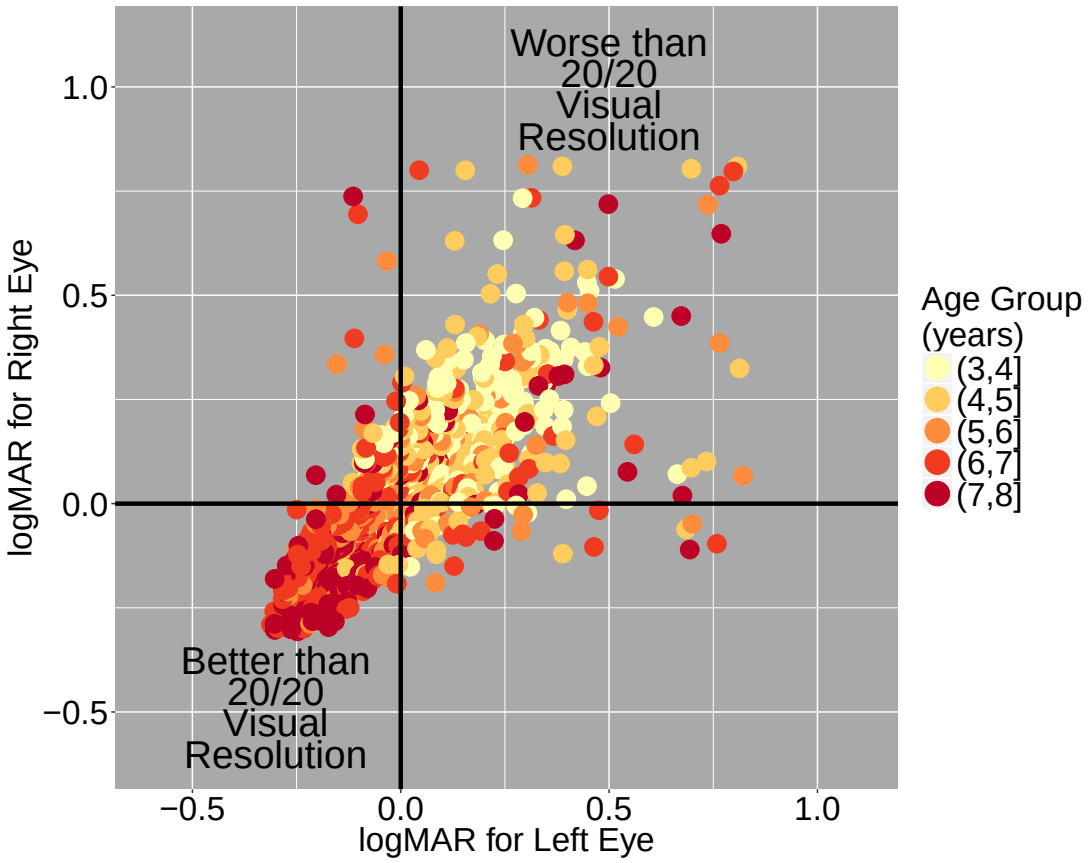

Figure 1: The values of logMar for 2721 children. A sequential colour palette is used to indicate age group.

probability density $c(u_1, u_2; \zeta)$, the shape of which is controlled by a parameter ζ, is shown in the bottom left panel of Figure 2. Copulas allow us to generate a very broad family of bivariate distributions by separating the dependence structure from the marginal models as illustrated in the whole of Figure 2. In particular, flexible bivariate probability density functions for the pair of random variables X_1 and X_2 can be created by combining c with marginal densities $f_1(x_1)$ and $f_2(x_2)$ (bottom right, top left of Figure 2) to produce a joint density

$$f(x_1, x_2; \zeta) = c\left(F_1(x_1), F_2(x_2); \zeta\right) \times f_1(x_1) \times f_2(x_2), \tag{1}$$

as shown in the top right of Figure 2, in which F_1 and F_2 are the cumulative distribution functions of X_1 and X_2. Sklar's theorem [50] essentially says that almost any joint density can be constructed in this way.

Copulas have been the modelling tool of many applications in medical statistics and biostatistics over the last few years. Recent applications include Nikoloulopoulos [39] who used copula mixed models to capture the negative association between the number of true positives and true negatives in a bivariate meta-analysis of diagnostic test accuracy studies. Zhong and Cook [67] analyzed the heritability of chronic diseases, such as psoriatic arthritis, by modelling the nature and extent of within-family associations through a copula. A Gaussian copula is applied by Conlon and colleagues [9] to model the joint distribution of variables of interest given surrogate outcomes in a colorectal cancer clinical trial.

Conditional copula models allow the dependence between variables to vary according to the values of a covariate. Examples of recent applications of conditional copula models in the frequentist framework include Acar and colleagues [2], who developed methodology to analyze the dependence structures between the birth weights of twins arranged by birth order at different gestational ages. Gijbels and colleagues [18] applied the methodology to two different problems: understanding the effect of the gross domestic product per capita on the relationship between life expectancy at birth of males and females and modelling microbial characteristics of soil samples in terms of the amount of metals contained in the soil. Acar and colleagues [3]

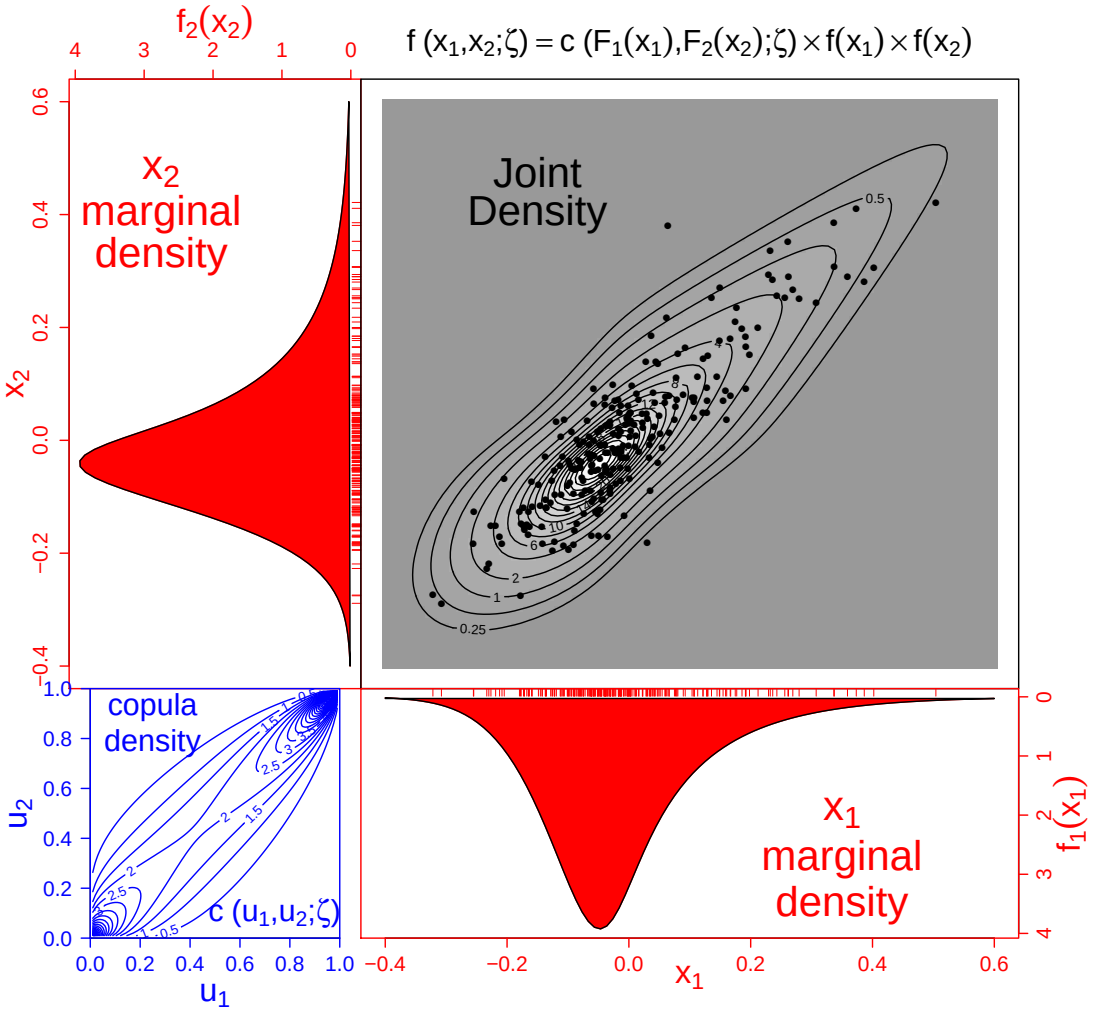

Figure 2: Bottom left: a bivariate copula probability density function $c(u_1, u_2; \zeta)$, in this case a basic Gumbel copula, the shape of which is controlled by the parameter ζ. The associated marginal distributions are uniform on $[0, 1]$. Bottom right: the marginal probability density function f_1 of a random variable X_1. Top left: the marginal probability density function f_2 of a random variable X_2. Top right: the marginal probability densities f_1 and f_2 are combined with the copula model c to form the joint probability density $f(x_1, x_2; \zeta)$ of (X_1, X_2).

developed inference methodology in the conditional copula context to test for an effect of body mass index on the dependence between pulse pressures measured at two different examination periods. Bayesian inference in conditional copula models was first proposed by Craiu and Sabeti [10], who investigated the performance of their approach using simulated data. Sabeti and colleagues [46] used this methodology to explore how the birth weights of twins depend on gestational and maternal ages.

In this paper we work with a class of bivariate copula additive models aimed at providing a full understanding of how left and right eye logMar distributions change with age. As our marginal model we use the four parameter sinh-arcsinh distribution proposed by Jones [24], which has the normal distribution as a special case. We allow the location, scale and two shape parameters of the sinh-arcsinh distribution to be smooth functions of age. In this way our marginal models are examples of generalized additive models for location, scale and shape proposed by Rigby and Stasinopoulos [45]. As is now commonplace in statistical inference, we learn about the unknown parameters of our model in the Bayesian framework using a specially tailored Markov chain Monte Carlo algorithm (MCMC) [8, 41, 54].

Vatter and Chavez-Demoulin [56] developed a frequentist generalized additive model framework for conditional bivariate dependence structures. They adopt a two-step estimation procedure, in which the marginal structures are estimated before the dependence relationships, and derive a range of valuable asymptotic results about this procedure. In their paper, they only allow marginal distributions to depend on covariates through a single parameter, although they state that this is done for the sake of simplicity. We use four-parameter sinh-arcsinh models to provide rich marginal modelling. In [56] the dependence relationship is estimated by means of an iteratively reweighted ridge regression algorithm. Vatter and Chavez-Demoulin [56] estimate smoothing parameters using a generalized cross-validation technique, whilst our method provides inference about these parameters in the form of their posterior distribution.

An advantage of a two-step estimation procedure, as adopted by Vatter and Chavez-Demoulin [56] and others, is that it offers a considerable amount of generality. It is possible, for example, to use the R packages `gamlss` [45] or `mgcv` [65] to fit a very broad class of marginal models, after which the `gamCopula` package [36] associated with [56] can be employed for joint modelling. The `gamCopula` package takes full advantage of the modelling flexibility of the `mgcv` package. Because of these considerations, the two-step estimation procedure is more flexible and easier to use than a simultaneous estimation approach such as ours. A disadvantage of the two-step estimation procedure is that the uncertainty associated with marginal estimation does not properly propagate to the estimation of the copula parameters. Bayesian simultaneous estimation approaches do not suffer from this drawback, although they introduce an extra layer of computational complexity with which the user has to deal. We will briefly discuss results obtained from the two-step approach of [56] in Section 4.

Our model is based on the additive conditional bivariate copula regression modelling framework of Klein and Kneib [27]. Klein and Kneib [27] work with a broad range of predictor functions, some defined using B-spline basis functions, penalized in a variety of ways. This allows them to consider spatial and random effects, as well as nonlinear effects of continuous covariates. In our work, we modify the approach of [27] in two ways. First, Klein and Kneib [27] use normal or three-parameter Dagum distributions as marginal models, whilst we use four-parameter sinh-arcsinh distributions. Secondly, we adopt a different way of constructing the MCMC proposal densities. This is based on numerically approximating full conditional probability density functions by a Gaussian proposal density that assigns considerable probability mass to relevant areas of the parameter space. Klein and Kneib [27] also present a brief discussion about problems with interval coverage rates associated with the two-step approach. Marra and Radice [32] have recently developed a similar modelling approach, again in the frequentist framework, but they consider two- and three-parameter marginal distributions. Although they estimate a global likelihood, they impose smoothing through a penalty term that can have a Bayesian interpretation. They base some of their inference on a Bayesian large sample approximation which allows them to obtain reliable point-wise confidence intervals for linear and non-linear functions of the model coefficients. Although this approach does not directly account for smoothing parameter uncertainty, Marra and Wood [33], for example, demonstrate using a careful simulation study that disregarding such uncertainty may not be problematic and that the Bayesian large sample approximation works well in practice. However, we prefer to work with an approach that takes full account of

all parameter uncertainty. Our Bayesian model with its sinh-arcsinh marginal distributions allows us to gain new insights about the dependence of visual acuity on age and to identify children with unusual sight for their age who may otherwise be missed in a way that quantifies estimation uncertainty.

The rest of this paper is structured as follows. In Section 2 we present in detail our bivariate copula additive models for location, scale, shape and dependence, and we introduce a range of copula families. We describe inference for our model in the Bayesian framework in Section 3. In Section 4 we discuss how the Watanabe-Akaike Information Criterion (WAIC) can be used for copula family selection and present a simulation study that confirms the utility of WAIC for this. We also describe the results of applying our model to the paediatric ophthalmic data shown in Figure 1, discussing the use of the posterior predictive distribution to identify children with unusual vision for their age, in either one or both eyes. We also compare our simultaneous Bayesian method with the two-step frequentist generalized additive modelling approach of Vatter and Chavez-Demoulin [56]. Finally, in Section 5 we present a short discussion.

2 A bivariate copula additive model for location, scale, shape and dependence

In this section we present a bivariate copula additive model aimed at providing an understanding of how left and right logMar distributions change with age. Let $(y_i^{\mathrm{L}}, y_i^{\mathrm{R}})$ stand for

$$(\text{Left logMar}_i, \text{Right logMar}_i)$$

for child $i = 1, \ldots, n$, where the sample size $n = 2721$, and let $w_i = \text{age}_i$. We assume that $(y_i^{\mathrm{L}}, y_i^{\mathrm{R}})$ given w_i, $i = 1, \ldots, n$, are independent. We saw from Figure 1 that older children tend to have lower logMar scores (and hence better visual acuity) compared to their younger counterparts. This effect becomes clearer when we consider marginal scatter plots of left and right logMar against age, as shown in Figure 3. We can see from the estimate of the location parameter of the marginal distributions of logMar at each age (middle curve), obtained using the methodology that we present in this paper, that logMar tends to reduce with age. Here we model underlying relationships using smooth curves formulated as natural cubic splines, as will be discussed in Sections 2.1 and 2.3.

As explained in Section 1, a copula-based approach separates marginal from dependence modelling. As Figure 3 suggests that the distribution of logMar given age has a heavy upper tail, we adopt a sinh-arcsinh family of distributions [24] as marginal models. This sinh-arcsinh model has excellent numerical properties and can represent a wide variety of distributional shapes, as we will see in Section 2.2. We will then discuss marginal modelling in Section 2.3, dependence modelling and model extensions in Section 2.4, and the copula families that we use in Section 2.5.

2.1 Natural cubic splines

We shall make use of natural cubic splines in our model, as shown in Figure 3. A natural cubic spline curve g is defined in terms of K knots $\omega_1 < \cdots < \omega_K$. It is cubic between and linear beyond the knots, with continuous second derivative g''. It can be shown that g is uniquely defined by its knot values $\mathbf{g} = (g_1, \ldots, g_K)^T$ [19, 30, 65]. Green and Silverman [19] discussed Bayesian inference for natural cubic splines, while Thompson and colleagues [55] extended their approach to the quantile regression setting. We further discuss natural cubic splines and their use in Sections 3 and 4.

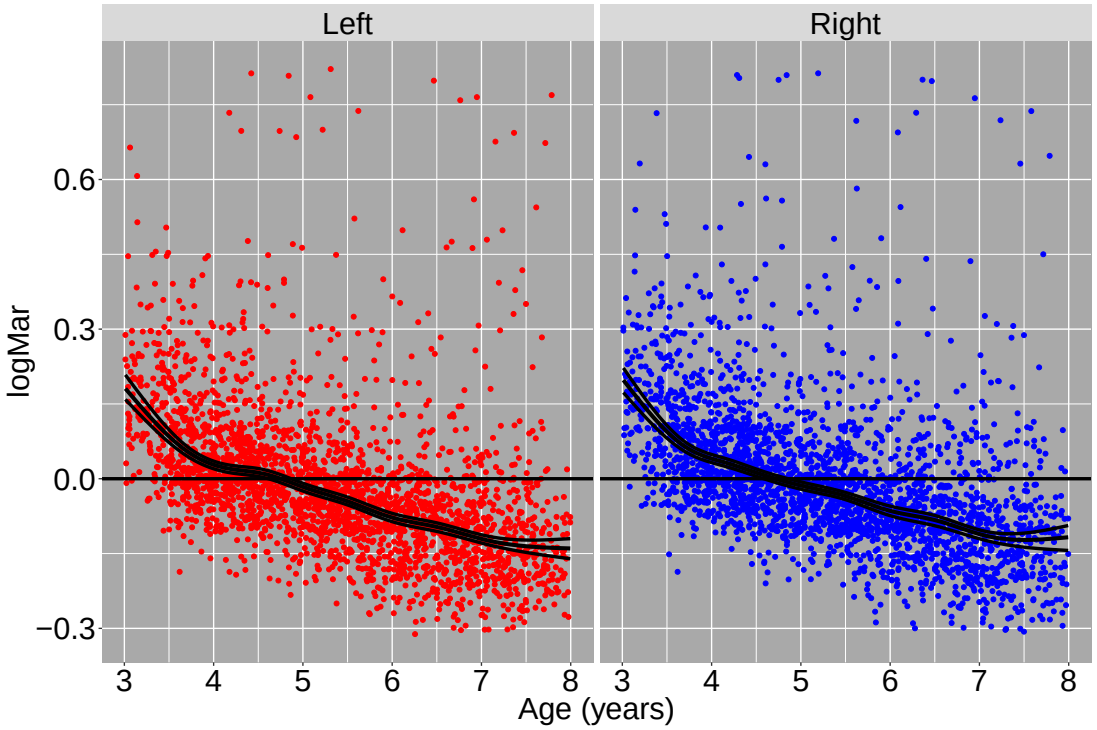

Figure 3: Scatter plots of the marginal dependence of logMar on age. Data from the left and right eyes are shown separately. The middle smooth curve, which was obtained using the methodology of this paper, is an estimate of the location parameter of the marginal distributions of logMar at each age given by the posterior mean. It is a natural cubic spline. The outer curves provide 95% credible intervals for this relationship.

2.2 Sinh-arcsinh distributions

Our marginal modelling will be based on a sinh-arcsinh family of distributions [24]. This family can assume a wide variety of distributional shapes including normal, asymmetric and kurtotic, the last two of which are particularly relevant when modelling logMar. We parameterize the associated probability density function using location μ, scale $\sigma > 0$, and lower $l > 0$ and upper $u > 0$ tail weight parameters as:

$$f(y \,|\, \mu, \sigma, l, u) = \frac{1}{\sqrt{2\pi}\sigma} \frac{Z_1}{\sqrt{1+z^2}} \exp\left(-\frac{Z_2^2}{2}\right),$$

in which

$$Z_1 = \frac{1}{2}\left\{u \exp\left(u \sinh^{-1}(z)\right) + l \exp\left(-l \sinh^{-1}(z)\right)\right\}$$

and

$$Z_2 = \frac{1}{2}\left\{\exp\left(u \sinh^{-1}(z)\right) - \exp\left(-l \sinh^{-1}(z)\right)\right\},$$

where $z = (y - \mu)/\sigma$. This sinh-arcsinh family of distributions was discussed in greater depth by Jones and Pewsey [25] in which a different parameterization was used, and is implemented in the R `gamlss` package under the acronym SHASH [45]. Values of l and u less (greater) than 1 correspond to tails heavier (lighter) than normal. It is natural to work with $\log(l)$ and $\log(u)$ so that 0 divides tails that are heavier from those that are lighter than normal on this log scale. Figure 4 presents examples of SHASH probability density functions f, with $\mu = 0$ and $\sigma = 1$, for different values of $\log(l)$ and $\log(u)$. The standard normal distribution, corresponding to $\log(l) = \log(u) = 0$, is shown in grey. Negative (positive) values of $\log(l)$ or $\log(u)$ correspond to a lower or upper tail that is heavier (lighter) than that of the standard normal. It is worth mentioning that the parameterization of the SHASH distribution that we use is natural and easy to interpret. Moreover, the adoption of generating distributions that are different from normal [25] means that the SHASH distribution itself can be extended to have even richer tail behaviour. We shall use the notation $F(y \,|\, \mu, \sigma, l, u)$ for the corresponding SHASH cumulative distribution function.

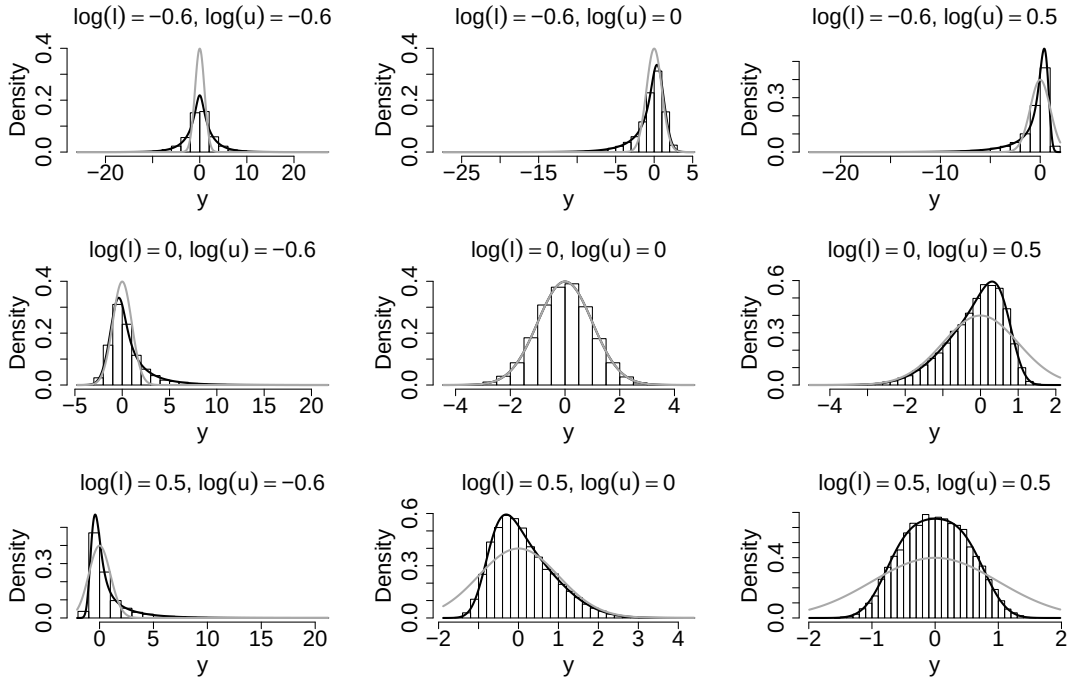

Figure 4: Examples of sinh-arcsinh probability density functions $f(y \,|\, \mu = 0, \sigma = 1, l, u)$ (black curve) calculated using the `dSHASH` function of the R `gamlss` package [45] for different values of $\log(l)$ and $\log(u)$, together with the standard normal probability density function (grey curve). Histograms of $10,000$ random variates generated using the `rSHASH` function are also shown. The central panel corresponds to the standard normal distribution with $\log(l) = \log(u) = 0$. In the first and third rows $\log(l) = -0.6$ and 0.5, respectively, while in the first and third columns $\log(u) = -0.6$ and 0.5, respectively.

2.3 Marginal modelling

As just mentioned, we assume that y_i^{L} and y_i^{R} follow SHASH distributions:

$$\begin{aligned} y_i^{\mathrm{L}} &\sim f\left(\cdot \,|\, \mu_i^{\mathrm{L}}, \sigma_i^{\mathrm{L}}, l_i^{\mathrm{L}}, u_i^{\mathrm{L}}\right) \\ y_i^{\mathrm{R}} &\sim f\left(\cdot \,|\, \mu_i^{\mathrm{R}}, \sigma_i^{\mathrm{R}}, l_i^{\mathrm{R}}, u_i^{\mathrm{R}}\right), \; i = 1, \ldots, n, \end{aligned}$$

and we model the parameters of these SHASH distributions as smooth functions of age. In particular, we adopt the following forms for μ_i^{L} and μ_i^{R}:

$$\begin{aligned} \mu_i^{\mathrm{L}} &= \mu^{\mathrm{L}} + g_\mu^{\mathrm{L}}(w_i) \\ \mu_i^{\mathrm{R}} &= \mu^{\mathrm{R}} + g_\mu^{\mathrm{R}}(w_i), \end{aligned} \tag{2}$$

where μ^{L} (μ^{R}) are the overall left (right) logMar locations across all ages, and g_μ^{L} (g_μ^{R}) are natural cubic splines uniquely defined by their values $g_{\mu,1}^{\mathrm{L}}, \ldots, g_{\mu,K}^{\mathrm{L}}$ ($g_{\mu,1}^{\mathrm{R}}, \ldots, g_{\mu,K}^{\mathrm{R}}$) at K knots $\omega_1, \ldots, \omega_K$. For identifiability we impose sum-to-zero constraints

$$\sum_{j=1}^{K} g_{\mu,j}^{\mathrm{L}} = \sum_{j=1}^{K} g_{\mu,j}^{\mathrm{R}} = 0, \tag{3}$$

so that the curves g_μ^{L} and g_μ^{R} are centered around zero. This means that it is not possible to define the model in an equivalent way by adding a constant to μ^{L} or μ^{R} and subtract it from g_μ^{L} or g_μ^{R}.

To place the K knots we sort the w_i values and select those with equally spaced subscripts in $1, \ldots, n$. We use $K = 10$ in the results that we present in this paper.

Because we perform inference in the Bayesian framework, we follow the usual approach of working with positive precisions $p_i^{\mathrm{L}} = 1/{\sigma_i^{\mathrm{L}}}^2$ and $p_i^{\mathrm{R}} = 1/{\sigma_i^{\mathrm{R}}}^2$ instead of scale parameters σ_i^{L}

and σ_i^{R}. These precisions are modelled on the log scale as

$$\begin{aligned} \log\left(p_i^{\mathrm{L}}\right) &= \mu_p^{\mathrm{L}} + g_p^{\mathrm{L}}(w_i) \\ \log\left(p_i^{\mathrm{R}}\right) &= \mu_p^{\mathrm{R}} + g_p^{\mathrm{R}}(w_i), \end{aligned} \tag{4}$$

where μ_p^{L} (μ_p^{R}) are overall log precision levels for left (right) logMar across all ages, and g_p^{L} (g_p^{R}) are natural cubic splines subject to sum-to-zero identifiability constraints similar to equation (3). Because our inference methodology is simulation-based, it is straightforward to learn about scale parameters from simulated values of precisions using the simple transformation scale parameter $= 1/\sqrt{\text{precision}}$.

We model the lower and upper tail weights l_i^{L}, l_i^{R}, u_i^{L} and u_i^{R} in a similar way:

$$\begin{aligned} \log\left(l_i^{\mathrm{L}}\right) &= \mu_l^{\mathrm{L}} + g_l^{\mathrm{L}}(w_i), \qquad \log\left(l_i^{\mathrm{R}}\right) = \mu_l^{\mathrm{R}} + g_l^{\mathrm{R}}(w_i), \\ \log\left(u_i^{\mathrm{L}}\right) &= \mu_u^{\mathrm{L}} + g_u^{\mathrm{L}}(w_i) \text{ and } \log\left(u_i^{\mathrm{R}}\right) = \mu_u^{\mathrm{R}} + g_u^{\mathrm{R}}(w_i). \end{aligned} \tag{5}$$

2.4 Dependence modelling and model extensions

We now define the joint probability density function of y_i^{L} and y_i^{R} based on (1) as:

$$\begin{aligned} &f\left(y_i^{\mathrm{L}}, y_i^{\mathrm{R}}; \zeta_i\right) = \\ &\quad c\left(F(y_i^{\mathrm{L}} \,|\, \mu_i^{\mathrm{L}}, \sigma_i^{\mathrm{L}}, l_i^{\mathrm{L}}, u_i^{\mathrm{L}}), F(y_i^{\mathrm{R}} \,|\, \mu_i^{\mathrm{R}}, \sigma_i^{\mathrm{R}}, l_i^{\mathrm{R}}, u_i^{\mathrm{R}}) \,|\, \zeta_i\right) \times \\ &\quad f\left(y_i^{\mathrm{L}} \,|\, \mu_i^{\mathrm{L}}, \sigma_i^{\mathrm{L}}, l_i^{\mathrm{L}}, u_i^{\mathrm{L}}\right) \times f\left(y_i^{\mathrm{R}} \,|\, \mu_i^{\mathrm{R}}, \sigma_i^{\mathrm{R}}, l_i^{\mathrm{R}}, u_i^{\mathrm{R}}\right), \end{aligned}$$

in which c is a bivariate copula probability density function whose dependence parameter ζ_i is allowed to depend on age w_i as:

$$\zeta_i = \mu + g(w_i), \tag{6}$$

where g is a natural cubic spline constrained as in (3). We will discuss the copula families that we consider for c in Section 2.5 together with the parameter ζ.

Because of the form of (2) and (4) to (6), we think of our model as a bivariate copula additive model for location, scale, shape and dependence.

Rahi and colleagues [44] discuss bivariate data that comprise spherical equivalence measurements on both eyes of adults, together with birth weight $w_i^{(1)}$, maternal age $w_i^{(2)}$ and gender $w_i^{(3)} \in \{0, 1\}$. It is not hard to extend the model defined through (2) and (4) to (6) to deal with their data:

$$\begin{aligned} \mu_i^{\mathrm{L}} &= \mu^{\mathrm{L}} + g_{\mu}^{\mathrm{L}^{(1)}}(w_i^{(1)}) + g_{\mu}^{\mathrm{L}^{(2)}}(w_i^{(2)}) + \beta^{\mathrm{L}} w_i^{(3)} \\ \mu_i^{\mathrm{R}} &= \mu^{\mathrm{R}} + g_{\mu}^{\mathrm{R}^{(1)}}(w_i^{(1)}) + g_{\mu}^{\mathrm{R}^{(2)}}(w_i^{(2)}) + \beta^{\mathrm{R}} w_i^{(3)} \end{aligned}$$

$$\begin{aligned} \log\left(p_i^{\mathrm{L}}\right) &= \mu_p^{\mathrm{L}} + g_p^{\mathrm{L}^{(1)}}(w_i^{(1)}) + g_p^{\mathrm{L}^{(2)}}(w_i^{(2)}) + \beta_p^{\mathrm{L}} w_i^{(3)} \\ \log\left(p_i^{\mathrm{R}}\right) &= \mu_p^{\mathrm{R}} + g_p^{\mathrm{R}^{(1)}}(w_i^{(1)}) + g_p^{\mathrm{R}^{(2)}}(w_i^{(2)}) + \beta_p^{\mathrm{R}} w_i^{(3)} \end{aligned}$$

$$\begin{aligned} \log\left(l_i^{\mathrm{L}}\right) &= \mu_l^{\mathrm{L}} + g_l^{\mathrm{L}^{(1)}}(w_i^{(1)}) + g_l^{\mathrm{L}^{(2)}}(w_i^{(2)}) + \beta_l^{\mathrm{L}} w_i^{(3)} \\ \log\left(l_i^{\mathrm{R}}\right) &= \mu_l^{\mathrm{R}} + g_l^{\mathrm{R}^{(1)}}(w_i^{(1)}) + g_l^{\mathrm{R}^{(2)}}(w_i^{(2)}) + \beta_l^{\mathrm{R}} w_i^{(3)} \\ \log\left(u_i^{\mathrm{L}}\right) &= \mu_u^{\mathrm{L}} + g_u^{\mathrm{L}^{(1)}}(w_i^{(1)}) + g_u^{\mathrm{L}^{(2)}}(w_i^{(2)}) + \beta_u^{\mathrm{L}} w_i^{(3)} \\ \log\left(u_i^{\mathrm{R}}\right) &= \mu_u^{\mathrm{R}} + g_u^{\mathrm{R}^{(1)}}(w_i^{(1)}) + g_u^{\mathrm{R}^{(2)}}(w_i^{(2)}) + \beta_u^{\mathrm{R}} w_i^{(3)} \end{aligned}$$

$$\zeta_i = \mu + g^{(1)}(w_i^{(1)}) + g^{(2)}(w_i^{(2)}) + \beta w_i^{(3)},$$

in which all the μ and β terms are constant, and all the g terms are natural cubic splines subject to sum-to-zero constraints. In this way our modelling framework can deal with data with a range of covariates.

2.5 Copula families

There are many different copula families allowing very flexible dependence modelling [22, 37]. We will work with the commonly used Gaussian, t, Clayton, Gumbel and Frank copula models [37, 66]. This set of copula families defines a broad range of bivariate distributions for use in modelling. When working with the t copula, we fix the number ν of degrees of freedom, as it is sometimes difficult to perform inference about ν. We denote the resulting copula t-ν, and refer to [1, 34, 35, 59] for further discussion. The basic Clayton and Gumbel copulas allow for positive correlation only. For general modelling we extend these copulas to 'double copulas' by including rotated versions, which allow modelling of negative dependence. So the Gumbel-0-90 double copula, for example, comprises the basic Gumbel copula shown in Figure 2 together with its rotation by 90°. The Gumbel-0-90 double copula cumulative distribution function with dependence parameter θ can be written as

$$C_{\text{Gumbel-0-90}}(u_1, u_2; \theta) = \begin{cases} C_{\text{Gumbel-0}}(u_1, u_2; \theta) & \text{if} \quad \theta \in [1, \infty) \\ C_{\text{Gumbel-90}}(u_1, u_2; \theta) & \text{if} \quad \theta \in (-\infty, -1] \end{cases}$$

where $C_{\text{Gumbel-0}}(u_1, u_2; \theta)$ is the basic Gumbel copula cumulative distribution function and $C_{\text{Gumbel-90}}(u_1, u_2; \theta)$ is the cumulative distribution function of the Gumbel copula rotated by 90°. The Gumbel-0-90 double copula cumulative distribution function therefore takes the form

$$C_{\text{Gumbel-0-90}}(u_1, u_2; \theta) = \begin{cases} \exp\left[-\left\{(-\log(u_1))^\theta + (-\log(u_2))^\theta\right\}^{1/\theta}\right] & \text{if} \quad \theta \in [1, \infty) \\ u_2 - \exp\left[-\left\{(-\log(1-u_1))^\theta + (-\log(u_2))^\theta\right\}^{1/\theta}\right] & \text{if} \quad \theta \in (-\infty, -1]. \end{cases}$$

The Gumbel-180-270 double copula with dependence parameter θ is a 180° rotation of the Gumbel-0-90 copula and has cumulative distribution function

$$C_{\text{Gumbel-180-270}}(u_1, u_2; \theta) = \begin{cases} C_{\text{Gumbel-180}}(u_1, u_2; \theta) & \text{if} \quad \theta \in [1, \infty) \\ C_{\text{Gumbel-270}}(u_1, u_2; \theta) & \text{if} \quad \theta \in (-\infty, -1] \end{cases}$$

where $C_{\text{Gumbel-180}}(u_1, u_2; \theta)$ is the cumulative distribution function of the Gumbel copula rotated by 180°, often called the 'survival copula', and $C_{\text{Gumbel-270}}(u_1, u_2; \theta)$ is the cumulative distribution function of the Gumbel copula rotated by 270°. The Gumbel-180-270 double copula cumulative distribution function can be expressed as

$$C_{\text{Gumbel-180-270}}(u_1, u_2; \theta) = \begin{cases} u_1 + u_2 - 1 + \exp\left[-\left\{(-\log(1-u_1))^\theta + (-\log(1-u_2))^\theta\right\}^{1/\theta}\right] & \text{if} \quad \theta \in [1, \infty) \\ u_1 - \exp\left[-\left\{(-\log(u_1))^\theta + (-\log(1-u_2))^\theta\right\}^{1/\theta}\right] & \text{if} \quad \theta \in (-\infty, -1]. \end{cases}$$

Since dependence parameters from distinct copula families often have different domains of definition and interpretations, we parameterized them all in terms of Kendall's correlation coefficient $\tau \in (-1, 1)$, to allow easy comparison. There exists a one-to-one relationship between the parameter of each copula family and Kendall's τ. For example, for $C_{\text{Gumbel-0}}(u_1, u_2; \theta)$ and $C_{\text{Gumbel-180}}(u_1, u_2; \theta)$, $\tau = 1 - 1/\theta$, while for $C_{\text{Gumbel-90}}(u_1, u_2; \theta)$ and $C_{\text{Gumbel-270}}(u_1, u_2; \theta)$, $\tau = -1 - 1/\theta$. For the corresponding relationships for other copula families, see [6].

In order to map Kendall's τ into the real line, following Zhong and Cook [67], Almeida and colleagues [5] and many other authors, we apply Fisher's transformation $\mathcal{F}$:

$$\zeta = \mathcal{F}\{\tau(\theta)\} = \frac{1}{2}\log\left\{\frac{1+\tau(\theta)}{1-\tau(\theta)}\right\} = \tanh^{-1}\{\tau(\theta)\} \in \mathbb{R}.$$

For example, for $C_{\text{Gumbel-0}}(u_1, u_2; \theta)$ and $C_{\text{Gumbel-180}}(u_1, u_2; \theta)$, $\zeta = \log(2\theta - 1)/2$, while for $C_{\text{Gumbel-90}}(u_1, u_2; \theta)$ and $C_{\text{Gumbel-270}}(u_1, u_2; \theta)$, $\zeta = \log\{-1/(2\theta + 1)\}/2$. Then, ζ is allowed to depend on the covariate as described in equation (6).

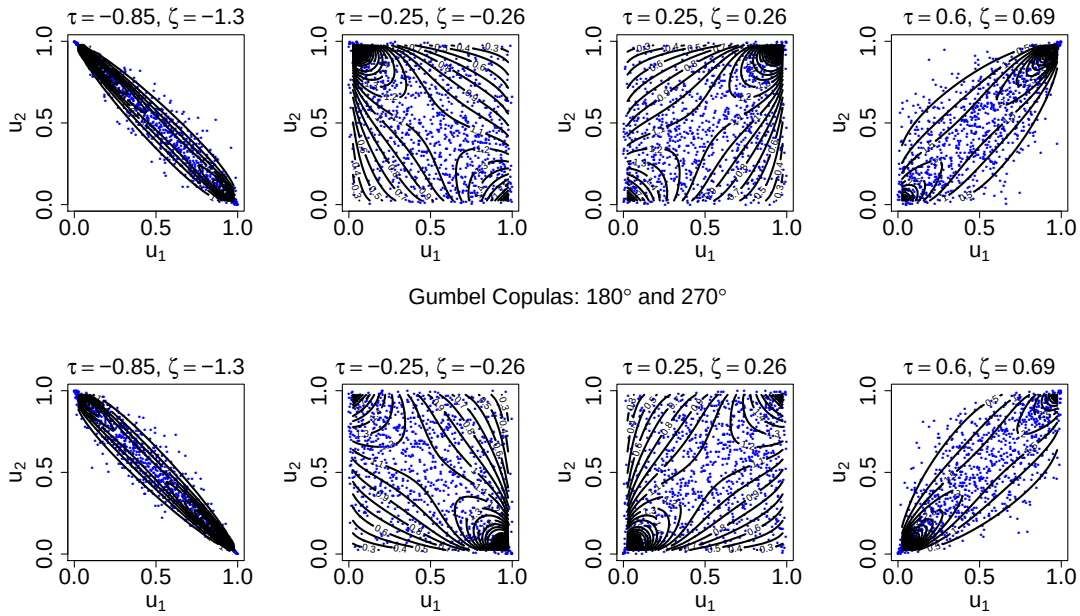

Figure 5: Top row: examples of Gumbel-0-90 double copulas. Bottom row: examples of Gumbel-180-270 double copulas.The third and fourth copula probability density functions c on the top row are examples of the basic Gumbel copula, while the first and second density functions are examples of basic Gumbel copulas rotated by 90° in order to allow for negative correlation. The values of Kendall's τ and of $\zeta = \mathcal{F}(\tau)$, where $\mathcal{F}$ is Fisher's transformation, are shown.

Examples of both the Gumbel-0-90 and the Gumbel-180-270 double copulas are shown in Figure 5. It can be seen, for example, that, when the correlation is large and positive, the Gumbel-0-90 (Gumbel-180-270) double copula has a wide (thin) lower and a thin (wide) upper tail. Similarly, the Clayton-0-90 (Clayton-180-270) double copula has a thin (wide) lower and a wide (thin) upper tail, while the tails of the Frank copula are identical.

3 Prior distributions and Bayesian inference

We discuss prior distributions for the unknown parameters of our model in Section 3.1. We outline how we simulated values of all the parameters in Section 3.2. In particular, we describe a way of constructing the MCMC proposal densities for the spline knot values that is related to, but different from Klein and Kneib [27].

3.1 Prior distributions

As we are performing inference in the Bayesian framework, we need to assign prior distributions to the unknown quantities about which we wish to make inference. We assume that, before seeing any data, all parameters in $\mathcal{M} = \left\{\mu^{\mathrm{L}}, \mu^{\mathrm{R}}, \mu_p^{\mathrm{L}}, \mu_p^{\mathrm{R}}, \mu_l^{\mathrm{L}}, \mu_l^{\mathrm{R}}, \mu_u^{\mathrm{L}}, \mu_u^{\mathrm{R}}, \mu\right\}$ are normally distributed with zero means and low precisions (for example, 0.01). We define the prior distributions on each element of $\mathcal{G} = \left\{g_\mu^{\mathrm{L}}, g_\mu^{\mathrm{R}}, g_p^{\mathrm{L}}, g_p^{\mathrm{R}}, g_l^{\mathrm{L}}, g_l^{\mathrm{R}}, g_u^{\mathrm{L}}, g_u^{\mathrm{R}}, g\right\}$ as we now explain for g following [19, 55]. We believe that before seeing any data g should be smooth, so we model the prior probability density $\pi(g \,|\, \lambda)$ for g as

$$\begin{aligned} \pi(g \,|\, \lambda) &\propto \lambda^{(K-2)/2} \exp\left(\frac{1}{2}\,\lambda \text{ smoothness of } g\right) \\ &\propto \lambda^{(K-2)/2} \exp\left(-\frac{1}{2}\lambda \int g''(w)^2 dw\right), \end{aligned} \tag{7}$$

so that smooth gs are favoured. Larger values of the positive smoothing parameter λ result in more probability density being given to smoother curves g.

Wood [65] presents a very detailed practical and mathematical discussion of smoothers in his Chapter 5. He explains how natural cubic splines with knots at all n covariate values are smoothers with somewhat ideal properties (his page 199). In practice, however, the basis independent roughness term $\lambda \int g''(w)^2 dw$, which leads to our prior probability density $\pi(g \,|\, \lambda)$ through (7), yields functions that are much smoother than the n degrees of freedoms suggest, meaning that natural cubic splines with n knots are wasteful. Our natural cubic splines are defined at only $K = 10$ knot values, meaning that the functions in $\mathcal{G}$ have an economical representation, while maintaining good properties. Using $K = 20$ knot values makes very little difference to the results, but of course leads to an increase in computational time. Employing a small number of knots has the consequence that extensions to more than one covariate, as discussed briefly in Section 2.3, are not computationally burdensome. Moreover, the parameterization in terms of knot values is easy to understand visually. For further discussion about natural cubic splines, see [21]. Recent examples of the use of natural cubic splines in the copula context are given in [56, 57]. Our modelling framework could be modified to work with different types of splines with prior densities defined using different roughness terms.

Green and Silverman [19] show that

$$\pi(g \,|\, \lambda) = \pi(\mathbf{g} \,|\, \lambda) \propto \lambda^{(K-2)/2} \exp\left(-\frac{1}{2}\lambda \, \mathbf{g}^T \mathcal{S} \mathbf{g}\right),$$

in which $\mathbf{g} = (g_1, \ldots, g_K)^T = (g(\omega_1), \ldots, g(\omega_K))^T$ are the values of g at the knots and $\mathcal{S}$ is a known $K \times K$ symmetric matrix. Finally, a gamma prior is adopted for the positive smoothing parameter λ such that the mean and standard deviation of λ are 20 and 15, respectively. This means that the prior on λ gives considerable probability density to a wide range of possible λ values and so has little influence on the posterior distribution of λ and of the corresponding g. We found that when the prior standard deviation of λ was considerably reduced, to 1 for example, the influence of this more concentrated prior on the posterior distribution of λ was quite strong. Let $\mathcal{L} = \left\{\lambda^{\mathrm{L}}, \lambda^{\mathrm{R}}, \lambda_p^{\mathrm{L}}, \lambda_p^{\mathrm{R}}, \lambda_l^{\mathrm{L}}, \lambda_l^{\mathrm{R}}, \lambda_u^{\mathrm{L}}, \lambda_u^{\mathrm{R}}, \lambda\right\}$. A similar prior construction applies for the remaining elements of $\mathcal{G}$ and their associated smoothing parameters $\mathcal{L}$. As is standard practice, we assume that, before seeing any data, the parameters in $\mathcal{M}$, $\mathcal{G}$ and $\mathcal{L}$ are independent both within and between these sets.

3.2 Bayesian inference

Inference is now based on the posterior distribution

$$\pi\left(\mathcal{M}, \mathcal{G}, \mathcal{L} \,|\, (y_i^{\mathrm{L}}, y_i^{\mathrm{R}}),\ i = 1, \ldots, n\right).$$

We wrote R code to simulate values of all the parameters according to this posterior distribution using a Metropolis-within-Gibbs algorithm [8, 41]. We used Gaussian proposals when updating each element of $\mathcal{M}$.

Klein and Kneib [27] discuss the construction of the proposal densities for each element of $\mathcal{G}$ in their Section 3.1. The basic idea of their approach is to approximate the log full conditional probability density function of $g = (g_1, \ldots, g_K)$, for example, by a Taylor expansion to achieve a normal proposal density with mean corresponding to the mode at the current state and covariance matrix corresponding to the curvature at this mode. In this way the proposal density puts probability mass on releavant areas of the sampling space. Their technique avoids the manual tuning of the standard deviation of a random walk proposal and is strongly related to the iterative weighted least squares algorithm of Gamerman [14] and to Fisher's scoring; further discussion is provided in Section 3.1 of [29]. It requires finding mathematical expressions for the first and second derivitaves of the associated log full conditional distribution. These can be complicated to find as Section 3.1.2 of Klein and Kneib [27] and Appendix C of Klein and colleagues (2013) [29] illustrate. Klein and colleagues (2015) [28] state that, if it is not possible to derive these derivatives analytically, appropriate approximations can be used to construct the proposal density. This is indeed what we do, as the derivations are far from straightforward for our model. In particular, we use R [43]'s function `optim` to find the maximizing value and curvature of the log full conditional numerically. Using such a numerical optimization approach means that the algorithm can be applied in the case of a large range of

models without the need for extensive differentiations. Optimization is, however, computationally expensive, if it is applied at each iteration of the MCMC algorithm. We found that very good results were obtained by applying numerical optimization only during the first five MCMC iterations and every three hundred iterations thereafter. We maintained the sum-to-zero constraint by thinking of the log full conditional probability density function of g, for example, as a function of only $g_1, \ldots, g_{K-1}$, with g_K set to $-(g_1 + \cdots + g_{K-1})$.

Because of conjugacy, each element of $\mathcal{L}$ can be sampled from a gamma distribution with appropriately defined parameters. It is for this reason that we kept the multiplicative $\lambda^{(K-2)/2}$ constant in (7). We ran the resulting MCMC algorithm for $20,000$ iterations, and discarded the first $5,000$ iterations as burn-in. We employed a thinning interval of 25 iterations.; see [17, 31, 60] for a discussion of thinning and its possible drawbacks. All our posterior summaries are based on thinned chain values. Increasing the number of iterations to $100,000$, for example, and using different thinning intervals or even no thinning led to very little change in the overall results.

4 Results

	Copula Used in Model													
Copula Used	t-4		Gaussian		Clayton-1		Clayton-2		Gumbel-1		Gumbel-2		Frank	
in Simulation	m	$\bar{W}$	m	$\bar{W}$	m	$\bar{W}$	m	$\bar{W}$	m	$\bar{W}$	m	$\bar{W}$	m	$\bar{W}$
t-4	99	**−276.8**	0	−201.6	0	−152.2	0	−162.2	1	−227.6	0	−226.8	0	−193.2
Gaussian	1	−183.3	93	**−215.4**	0	−128.5	0	−132.2	1	−185.7	4	−183.9	1	−189.2
Clayton-1	0	−212.5	0	−205.8	98	**−325.0**	0	−12.9	0	−111.0	2	−302.6	0	−194.9
Clayton-2	0	−218.6	0	−209.2	0	−10.6	89	**−324.6**	11	−306.5	0	−114.2	0	−197.7
Gumbel-1	0	−233.1	0	−222.1	0	−76.8	2	−251.5	98	**−274.4**	0	−166.5	0	−203.3
Gumbel-2	0	−236.2	0	−229.7	1	−254.3	0	−84.3	0	−171.1	99	**−278.0**	0	−208.2
Frank	0	−147.8	0	−176.9	0	−99.6	0	−102.7	0	−150.8	0	−145.9	100	**−207.9**

Table 1: Results of the simulation study for Bayesian copula model selection. m represents the number of times out of 100 that the copula used in the model that we propose had the lowest value of WAIC. $\bar{W}$ is the sample mean of the hundred WAIC values at each combination of copula families. Bold is used to indicate the minimum $\bar{W}$ value across the copulas used in the model (that is, across rows). Clayton-1, Clayton-2, Gumbel-1 and Gumbel-2 stand for the Clayton-0-90, Clayton-180-270, Gumbel-0-90 and Gumber-180-270 copula families.

In this section we discuss the results obtained from applying our Bayesian methodology to simulated and real data. In Section 4.1 we explore the use of two information criteria for model selection by means of a simulation study. We discuss for our real data how the location, scale, shape and dependence parameters of our model are related to the covariate age in Section 4.2. A brief comparison with the two-step frequentist approach of Vatter and Chavez-Demoulin [56] is provided in Section 4.3, in which model selection is again discussed. In Section 4.4 the posterior predictive distribution is used to identify children with unusual sight characteristics, distinguishing those who are bivariate, but not univariate outliers.

4.1 Bayesian copula model selection using WAIC and DIC

We used the Watanabe-Akaike Information Criterion (WAIC) [15, 63] and the Deviance Information Criterion (DIC) [52, 53] as goodness-of-fit type criteria to choose among Bayesian models that used the t-ν, $\nu \in \{2, \ldots, 10, 20, 100\}$, Gaussian, Clayton-0-90, Clayton-180-270, Gumbel-0-90, Gumbel-180-270 and Frank copulas.

DIC was proposed in 2002 by Spiegelhalter and colleagues [52] and is defined as

$$\begin{aligned} \text{DIC} \;=\; & -2 \text{ log-likelihood at parameter posterior means} \\ & +2 \text{ effective number of parameters.} \end{aligned}$$

The precise definition of the effective number of parameters is not important for our purposes, but is given in [53] where it is also explained how DIC may be calculated from the output of the MCMC algorithm. DIC is actually a penalized 'badness-of-fit' statistic and so models with a lower DIC should be preferred. A desirable feature of DIC is its invariance under monotone increasing transformations of the marginal distributions, which makes it suitable for copula modelling and selection [11].

WAIC was proposed in 2013 [63] and is now preferred over DIC because it provides a more fully Bayesian approach for estimating the out-of-sample predictive performance of a model [15]. It is defined as

$$\begin{aligned} \text{WAIC} \;=\; & -2 \text{ log-pointwise posterior predictive density} \\ & +2p_W, \end{aligned} \tag{8}$$

where the log pointwise posterior predictive density is defined, for a sample of independently and identically distributed observations $(z_1, \ldots, z_n)$ and a generic statistical model $p(\cdot|\theta)$, as

$$\sum_{i=1}^{n} \log\left(E\left[p(z_i|\theta)\right]\right). \tag{9}$$

The aim of the penalizing correction factor p_W is to avoid over-fitting. It can be computed in two different ways which can be viewed as approximations to cross-validation [15]. Here we consider

$$p_W = \sum_{i=1}^{n} \text{var}\left[\log\left(p(z_i|\theta)\right)\right]. \tag{10}$$

Both the expectation and the variance that appear in (9) and (10) are computed with respect to the posterior distribution of θ and can be easily calculated from the MCMC output. In fact, Watanabe's original definition of WAIC [63] differs from (8) in that it is divided by n and does not have the factor of 2. The definition (8) used here is on a scale that is comparable with DIC. As with DIC, models with a lower WAIC should be preferred.

For the models that we consider WAIC and DIC give similar values. For this reason and because WAIC is now preferred, when we present our simulation study results in Section 4.1.1, we concentrate on WAIC. When we discuss our real data results, we present the values of WAIC and DIC so that a comparison can be made.

4.1.1 Simulation study for Bayesian copula model selection

We confirmed the utility of WAIC for Bayesian copula model selection by a simulation study that we report briefly. For each of the seven copula families t-4, Gaussian, Clayton-0-90, Clayton-180-270, Gumbel-0-90, Gumbel-180-270 and Frank, we simulated 100 random samples of size $n = 1,000$ from the bivariate copula additive model for location, scale, shape and dependence that we describe in Section 2. We chose to work with the t-4 copula family because it performed best in the case of the real data, as we will see in Section 4.2. For each sample, the values of the covariates $w_1, \ldots, w_n$ were generated uniformly from the minimum to the maximum ages in the data set. In order to focus attention on copula selection, we set g_μ^{L}, g_μ^{R}, g_p^{L}, g_p^{R}, g_l^{L}, g_l^{R}, g_u^{L} and g_u^{R} to be zero. The covariate function $\zeta = \mu + g(w)$ was defined to grow linearly from -0.7 to 0.7 across the covariate range. This meant that our 700 simulated data sets are based on both negative and positive correlations, so illustrating the use of the double copula families described in Section 2.5. Bayesian inference based on $10,000$ MCMC iterations was then performed about the model using each of the seven copula families yielding $4,900$ WAIC values.

Figure 6 shows the $4,900$ values of WAIC from our simulation study. In eah panel the 100 randomly generated data sets have been ordered on the WAIC value for the copula model corresponding to the family used to simulate the data. It can be seen from the least variable (because of sorting) lower curves that the copula family employed to simulate the data is the one that generally has the lowest WAIC value when used in the model. Table 1 presents the number of times m out of 100 that the copula used in the model has the lowest value of WAIC. It can be seen by looking at the diagonal elements of the table that the correct copula family is generally selected. When this does not occur, Figure 6 shows that the correct copula family almost always yields only a slightly higher value of WAIC than the copula family with the lowest WAIC value, meaning that the correct family is only just missed. Table 1 also reports the sample mean $\bar{W}$ of the hundred WAIC values at each combination of copula families. The correct combination of copula families, corresponding to the diagonal of the table, always has the lowest sample mean. All these results from our simulation study confirm the utility of WAIC for copula selection.

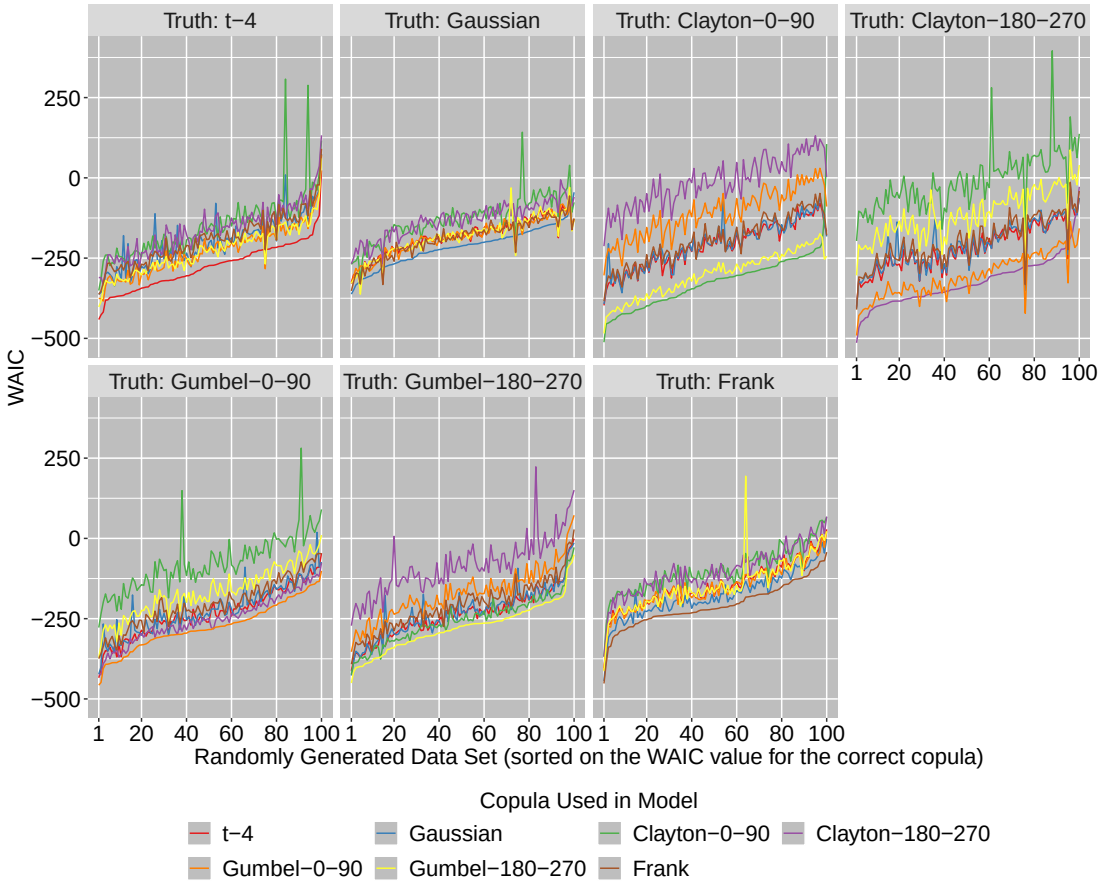

Figure 6: The 4,900 values of WAIC from our Bayesian copula model simulation study. One hundred data sets were simulated from the proposed model based on each of the seven copula families. Bayesian inference was then performed about the model using each of the seven copula families.

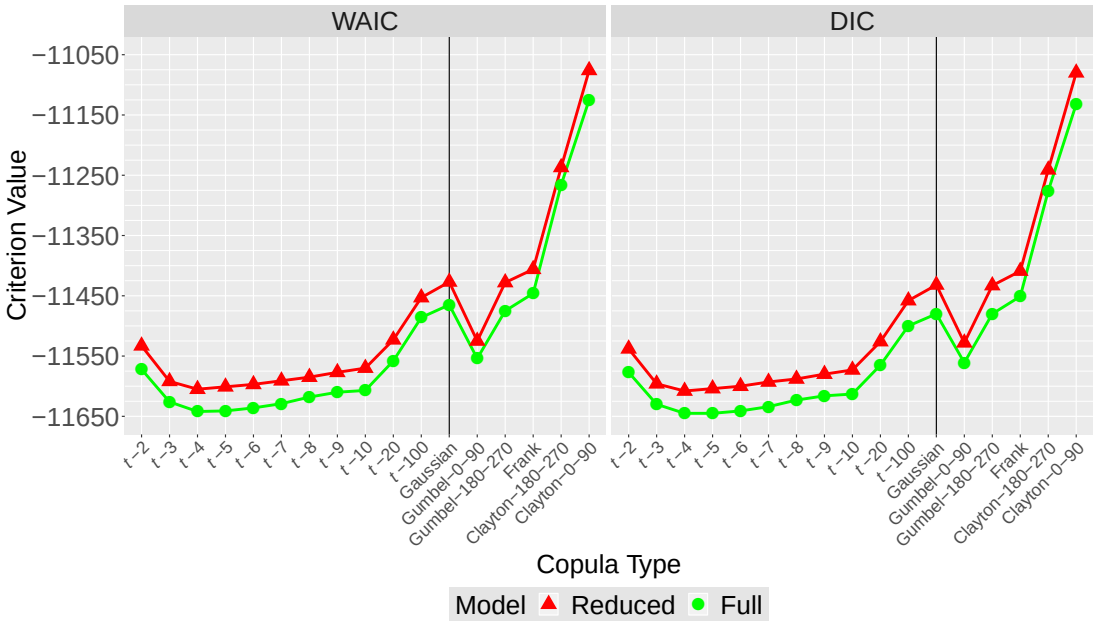

Figure 7: The values of WAIC and of DIC for the t-ν, $\nu \in \{2, \ldots, 10, 20, 100\}$, Gaussian, Clayton-0-90, Clayton-180-270, Gumbel-0-90, Gumbel-180-270 and Frank copula families. Results are shown for both the full and reduced models. After the Gaussian, the copulas have been arranged in increasing order of WAIC for the full model.

4.2 Real Data Results

Figure 7 presents the values of WAIC and of DIC for the t-ν, $\nu \in \{2, \ldots, 10, 20, 100\}$, Gaussian, Clayton-0-90, Clayton-180-270, Gumbel-0-90, Gumbel-180-270 and Frank copula families when our model, which we shall refer to here as the 'full model', was applied to the logMar data shown in Figure 1. It can be seen from Figure 7 that the t-4 is the preferred copula family for the full model, as it has the lowest value of WAIC. The performance of the t-ν copulas becomes less good as ν increases from 4, in which case the copulas themselves converge to the Gaussian. The Gumbel-0-90 copula is the best among the other copula families. We also performed inference about a 'reduced model' defined by setting g_p^{L}, g_p^{R}, g_l^{L}, g_l^{R}, g_u^{L}, g_u^{R} and g to be zero. Figure 7 also presents the values of WAIC for this reduced model, with the t-4 copula family again having the lowest value of WAIC. For all copula families, the value of WAIC for the reduced model is higher than for the full model, indicating that the full model is to be preferred. Corresponding values of DIC are also shown in Figure 7. These are similar to the values of WAIC, and so using DIC instead of WAIC would lead to very similar conclusions in this case.

All our code is written entirely in R [43]. The MCMC sampler described above runs in around 38 minutes for the full model and 15 minutes for the reduced model applied to our real data on an iMac with operating system High Sierra Version 10.13.5, a 4.2 GHz Intel Core i7 processor and 64 GB 2400 MHz DDR4 memory. The two-step frequentist approach of Vatter and Chavez-Demoulin [56], implemented through the `gamlss` [45] and `gamCopula` [36] R packages, is considerably faster.

4.2.1 Posterior inference for location, scale and shape

We used the Gelman-Rubin [16] diagnostic statistic to assess MCMC convergence. The Gelman-Rubin diagnostic involves checking the convergence of the algorithm with an ANOVA type approach using two or more samples generated in parallel. It yields a shrink factor $\hat{R}$; see also [41]. Values of $\hat{R}$ close to 1 indicate convergence. We used four samples generated from over dispersed starting points and monitored two quantities for illustration: μ and g_8. We obtained $\hat{R} = 1.004$ (upper 95% confidence limit 1.014) for μ and $\hat{R} = 1.000$ (1.001) for g_8. Brooks and Gelman's [7] multivariate version of the Gelman-Rubin diagnostic took the value 1.11. Figure 8 presents Gelman-Rubin-Brooks plots for these parameters and shows the evolution of $\hat{R}$ as the number of iterations of the thinned chains increases. Because of their proximity to 1, none of these $\hat{R}$s provides us with cause for concern.

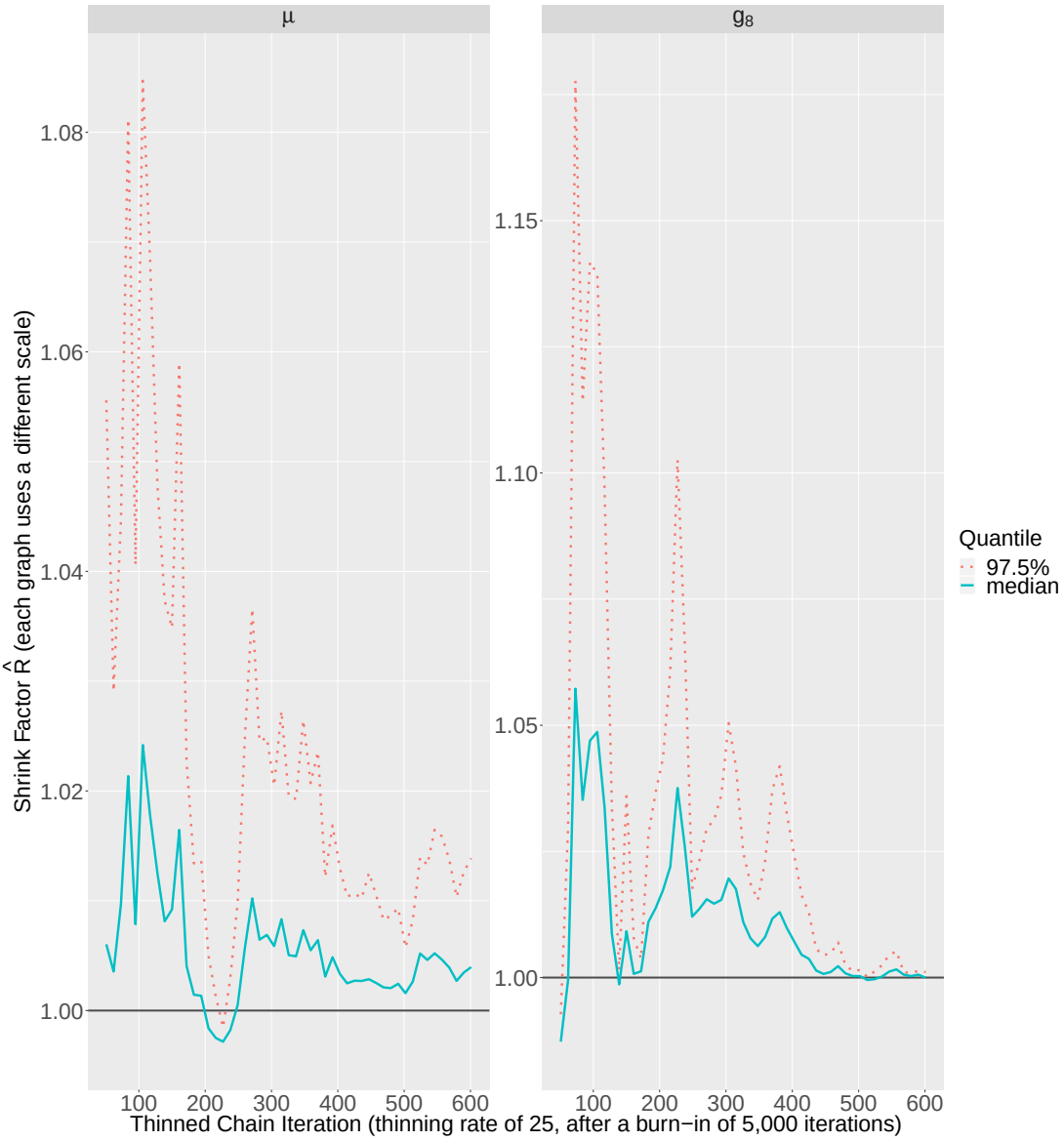

Figure 8: Gelman-Rubin-Brooks plots of the shrink factors $\hat{R}$ for μ and g_8

.

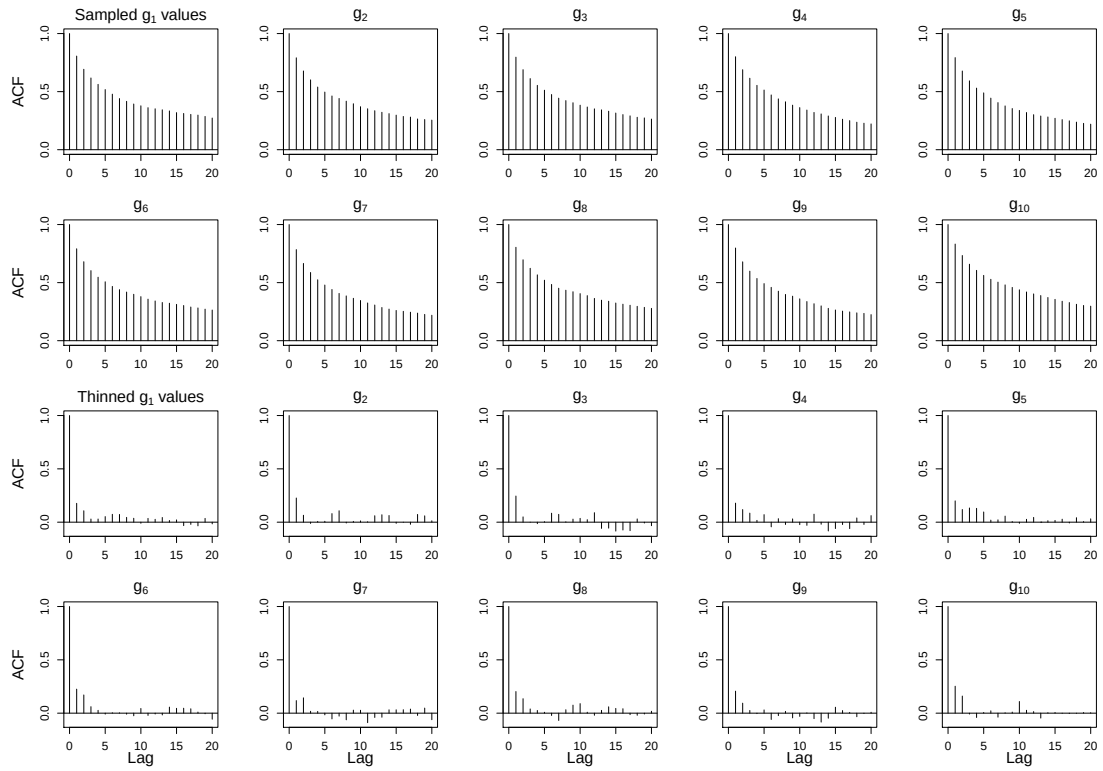

Figure 9: Autocorrelation functions for $g_1, \ldots, g_{10}$. The top two rows are for the MCMC sampled values, after burn-in. The bottom two rows are for those values after thinning has been applied.

As a further check of the performance of the algorithm, we looked at autocorrelation functions. Examples for the values $g_1, \ldots, g_{10}$ are shown in Figure 9. The autocorrelation functions for the sampled values themselves die off quite slowly. However, the autocorrelation decay is extremely rapid after thinning has been applied. Again we have no cause for concern.

Figure 3 shows the mean of the posterior distribution and associated 95% credible interval for the marginal location relationships $\mu^{\mathrm{L}} + g_{\mu}^{\mathrm{L}}(w)$ and $\mu^{\mathrm{R}} + g_{\mu}^{\mathrm{R}}(w)$ when the t-4 copula is used as our choice for c. It can be seen that the posterior mean of these logMar location quantities drops from around 0.15 for children aged 3 years to just above -0.15 for those aged 8 years, with the marginal behaviour in both eyes being rather similar.

Figure 10 shows how the scale parameters $\sigma^{\mathrm{L}} = 1/\sqrt{p^{\mathrm{L}}}$ and $\sigma^{\mathrm{R}} = 1/\sqrt{p^{\mathrm{R}}}$, in which $p^{\mathrm{L}} = \exp\left\{\mu_p^{\mathrm{L}} + g_p^{\mathrm{L}}(w)\right\}$ and $p^{\mathrm{R}} = \exp\left\{\mu_p^{\mathrm{R}} + g_p^{\mathrm{R}}(w)\right\}$, depend on age.

Figure 11 shows how the lower and upper log tail weights $\log(l^{\mathrm{L}})$, $\log(l^{\mathrm{R}})$, $\log(u^{\mathrm{L}})$ and $\log(u^{\mathrm{R}})$ depend on age. Values below (above) the horizontal 0 line correspond to tails of the SHASH distribution that are heavier (lighter) than those of the normal distribution. For both eyes the distribution of logMar given age has an upper tail that is much heavier than that of the normal distribution. This is consistent with what we saw in Figure 3.

4.2.2 Posterior inference for dependence

The posterior dependence of Kendall's $\tau = \mathcal{F}^{-1}(\zeta) = \mathcal{F}^{-1}\{\mu + g(w)\}$, where $\mathcal{F}^{-1}$ is the inverse of Fisher's transformation, on age can be seen in the left panel of Figure 12. The posterior mean of Kendall's τ tends to increase with age from around $\tau = 0.5$ for children aged 3 years to around 0.6 for children aged 8 years. This means that the dependence between the eyes becomes stronger over this age range.

Figure 10: The dependence of the scale parameters σ^{L} and σ^{R} on age. The middle smooth curve is the posterior mean of the scale parameter, while the outer curves provide 95% credible intervals. The t-4 copula was used as our choice for c.

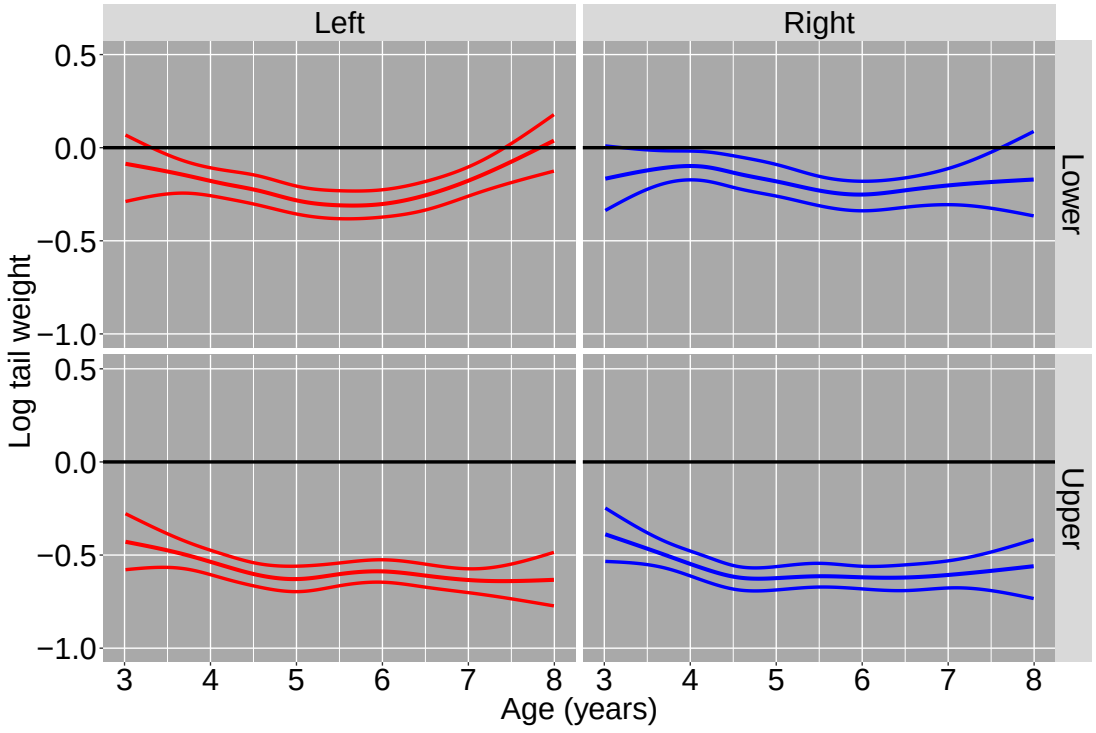

Figure 11: The dependence of the lower and upper log tail weights $\log(l^{\mathrm{L}})$, $\log(l^{\mathrm{R}})$, $\log(u^{\mathrm{L}})$ and $\log(u^{\mathrm{R}})$ on age. Values below (above) the horizontal 0 line correspond to tails of the SHASH distribution that are heavier (lighter) than those of the normal distribution. The middle smooth curve is the posterior mean of these dependence quantities, while the outer curves provide 95% credible intervals.

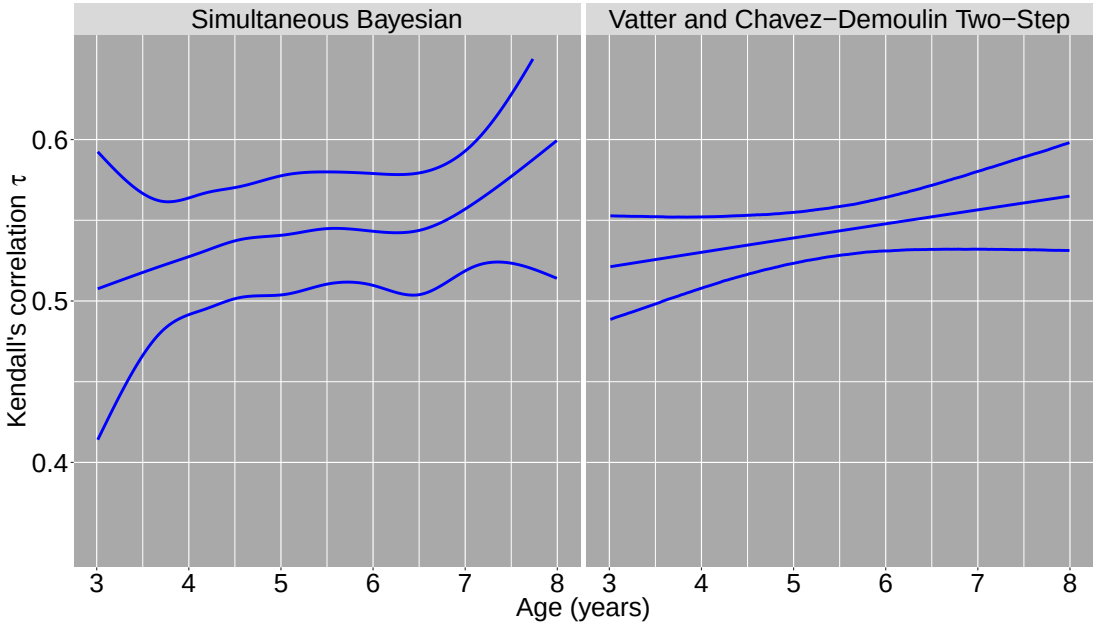

Figure 12: The dependence of Kendall's τ on age. Left panel: The middle curve is the posterior mean of τ and the outer curves are 95% credible intervals produced using the proposed simultaneous Bayesian approach. Right panel: The middle curve is the estimate of τ given by the two-step frequentist estimation procedure of Vatter and Chavez-Demoulin and the outer curves provide 95% confidence intervals.

4.3 Comparison with the Frequentist Estimation Procedure of Vatter and Chavez-Demoulin

For comparison we show the results of the two-step frequentist estimation procedure of Vatter and Chavez-Demoulin [56] in the right panel of Figure 12. We used the `gamlss` function of the `gamlss` [45] R package to fit SHASH marginal distributions, as discussed in Sections 2.2 and 2.3, to the left y_i^{L}, $i = 1, \ldots, n$, and right y_i^{R} data separately. We then transformed the associated left and right residuals to follow a $U[0, 1]$ distribution and modelled these jointly using the `gamBiCopFit` function from the `gamCopula` package [36]. Both functions `gamlss` and `gamBiCopFit` allow a broad class of models to be fitted and are fast. At the time of writing the `gamBiCopFit` function offers ten copula families, four of which – Gaussian, t, Clayton-0-90 and Gumbel-0-90 – overlap with the copula families discussed in Section 2.5 that we use. The `gamBiCopFit` function also estimates the number of degrees of freedom for the t copula, treating it as a nuisance parameter that is updated after each iteration. We selected the best copula using the value of AIC [4] provided by `gamBiCopFit`. Again, this was a t copula, with number of degrees of freedom estimated as 4.8. The frequentist procedure of Vatter and Chavez-Demoulin [56] estimated the dependence of τ on age in a similarly increasing way to our MCMC based Bayesian approach, although this estimate is considerably smoother. The 95% credible intervals associated with our simultaneous Bayesian approach are wider than 95% confidence intervals calculated as part of the second step of the Vatter and Chavez-Demoulin [56] procedure. In the former approach the uncertainty associated with marginal estimation and the smoothing parameters propagates to the estimation of the copula parameters.

We performed a simulation study similar to the one described in Section 4.1.1 to confirm the utility of AIC for frequentist copula model selection when using the Vatter and Chavez-Demoulin [56] approach. We considered four copula families t-4, Gaussian, Clayton-0-90 and Gumbel-0-90, and simulated data in the same was as in Section 4.1.1. Table 2 presents the number of times m out of 100 that the copula used in the model had the lowest AIC value. It can be seen that the Vatter and Chavez-Demoulin [56] method almost always selects the correct copula using AIC. When the t copula is chosen for data simulated from the Gaussian copula, the estimated number of degrees of freedom is over 30. A 95% confidence interval for the number of degrees of freedom of the t copula estimated from data

simulated from the t-4 copula is $(4.3, 4.6)$. Table 2 also reports the sample mean $\bar{A}$ of the hundred AIC values at each combination of copula families. The diagonal of the table always has the lowest sample mean. These results strongly confirm the utility of AIC for copula selection in the Vatter and Chavez-Demoulin [56] approach, and imply that this method has similar efficiency to ours in terms of chosing the right copula model.

	Copula Used in Model							
Copula Used	t		Gaussian		Clayton-0-90		Gumbel-0-90	
in Simulation	m	$\bar{A}$	m	$\bar{A}$	m	$\bar{A}$	m	$\bar{A}$
t-4	100	**−421.5**	0	−348.9	0	−294.8	0	−368.0
Gaussian	12	−367.3	88	**−370.0**	0	−262.8	0	−329.7
Clayton-0-90	0	−359.9	0	−342.5	100	**−440.5**	0	−220.3
Gumbel-0-90	0	−377.1	0	−363.0	0	−166.3	100	**−409.5**

Table 2: Results of the simulation study for copula model selection with the frequentist estimation procedure of Vatter and Chavez-Demoulin. m represents the number of times out of 100 that the copula used in the model had the lowest value of AIC. $\bar{A}$ is the sample mean of the hundred AIC values at each combination of copula families. Bold is used to indicate the minimum $\bar{A}$ value across the copulas used in the model (that is, across rows).

4.4 Using the Bayesian posterior predictive distribution to identify children with unusual vision

The Bayesian posterior predictive distribution can be used as a way of seeing how well our model corresponds to the data; see Chapter 10 of Ntzoufras [41] for a full discussion. The predictive distribution is

$$\pi\left((y^{\mathrm{L,\ New}}, y^{\mathrm{R,\ New}}) \,|\, (y_i^{\mathrm{L}}, y_i^{\mathrm{R}}),\ i = 1, \ldots, n; \text{age } w\right), \tag{11}$$

in which $(y^{\mathrm{L,\ New}}, y^{\mathrm{R,\ New}})$ is a data point from a new child of age w. We can simulate values from this distribution in a standard way by adding to our MCMC algorithm a step that samples a value from the data model defined in Section 2 for a fixed value of age w.

In Figure 13 we use coloured hexagons to present points simulated according to the predictive probability density function (11) at fixed age values $w = 3.5$, 4.5, 5.5, 6.5 and 7.5 years. These values correspond to the midpoints of the age ranges $(3, 4]$, $(4, 5]$, $(5, 6]$, $(6, 7]$ and $(7, 8]$ years used in Figure 1. The data points coming from children within these ages ranges are also shown using black points. We can see that points simulated from the predictive densities support most of the data well, although there are some outlying data points.

Figure 13 can therefore be used as a diagnostic aid enabling clinicians to identify children with unusual sight for their age. This can be quantified further as we now describe for the $(4, 5]$ age group.

We adopt a computationally inexpensive approach. First, using the sample from the predictive distribution (11) with age $w = 4.5$, we computed a univariate kernel density estimate [48, 49], $\mathrm{pred}^{\mathrm{L}}$ say, of the marginal distribution of $y^{\mathrm{L,\ New}}$; we found the marginal predictive density $\mathrm{pred}^{\mathrm{R}}$ based on $y^{\mathrm{R,\ New}}$ in a similar way. We also computed a bivariate kernel density estimate [58], $\mathrm{pred}^{\mathrm{L,\ R}}$ say, of the joint distribution of $(y^{\mathrm{L,\ New}}, y^{\mathrm{R,\ New}})$.

Next, we evaluated $\mathrm{pred}^{\mathrm{L}}$ at data values y_i^{L} corresponding to the 687 children in this age group and declared as 'unusual left eyes' those with associated density values that are below a threshold that cuts off the bottom 5%, say, of these density values. In this way we identified children with unusual left eyes in terms of the marginal predictive distribution of left eye logMar values. The data for these 35 children (red squares) tend to take very high or very low values on the horizontal 'logMar for Left Eye' axis of the top left panel of Figure 14.

Next, a similar approach based on $\mathrm{pred}^{\mathrm{R}}$ and y_i^{R} allowed us to identify children with unusual right eyes, this time in terms of the marginal predictive distribution of right eye

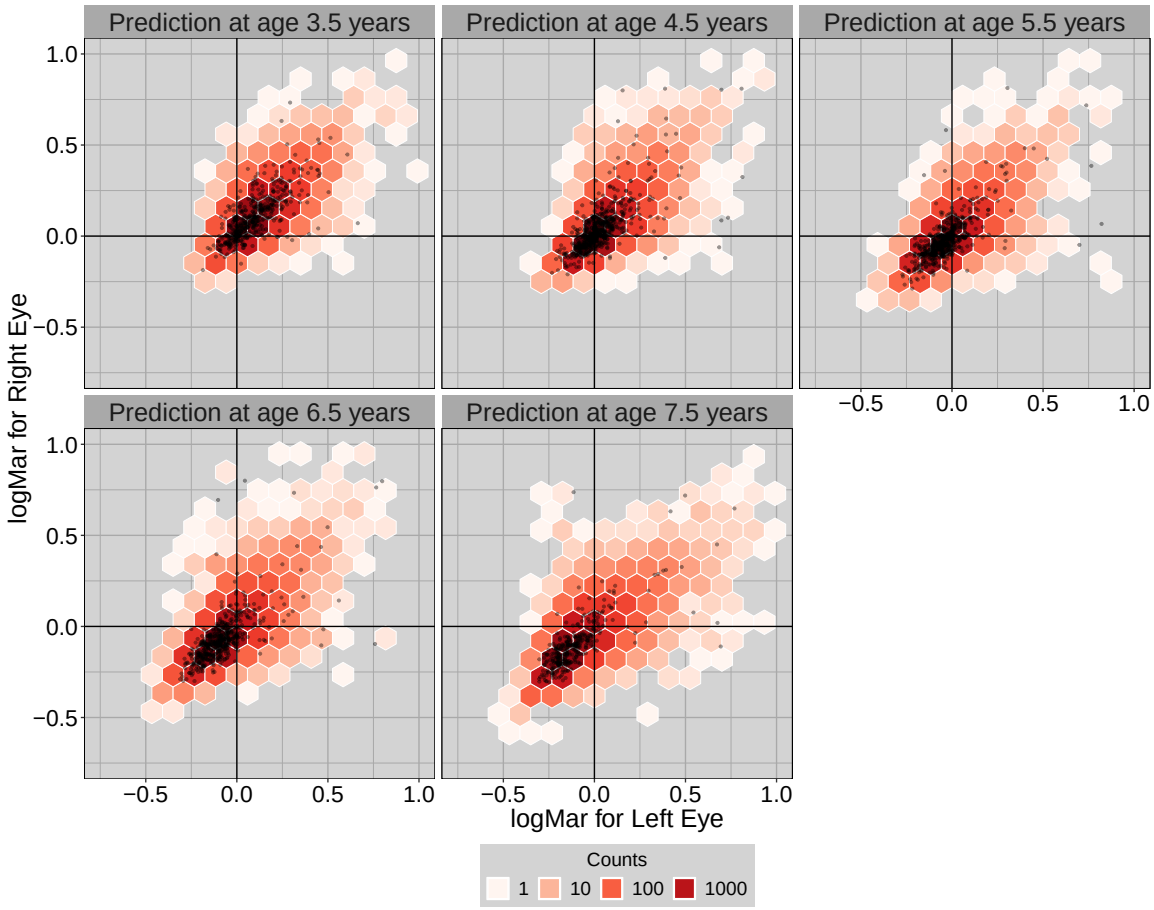

Figure 13: Points simulated from the posterior predictive probability density function (11) for values of $w \in \{3.5, 4.5, 5.5, 6.5, 7.5\}$ years are shown using coloured hexagons. These values of w are the midpoints of the age ranges $(3, 4]$, $(4, 5]$, $(5, 6]$, $(6, 7]$ and $(7, 8]$ years. Data points $(y_i^{\mathrm{L}}, y_i^{\mathrm{R}})$ coming from children with ages in these age ranges are shown in black.

logMar values. The data for these 35 children tend to take very high or very low values on the vertical 'logMar for Right Eye' axis of the top right panel of Figure 14.

Finally, eyes that are jointly unusual can be identified by evaluating the joint predictive density $\text{pred}^{\text{L, R}}$ at the data points $(y_i^{\text{L}}, y_i^{\text{R}})$ and ordering the resulting predictive density values. The data for these 35 children tend to take values that are away from the main body of data points in the bottom right panel of Figure 14.

In the bottom left panel of Figure 14 we highlight the union of the two sets of marginal outliers shown in the top row of Figure 14. These outlying points corresponds to a larger number of 56 children with unusual left or right eyes.

The 35 joint outliers shown in the bottom right panel of Figure 14 comprised 30 children who had already been identified as being unusual, together with 5 additional children (black triangles) who would not otherwise have been recognized as having potentially problematic vision. Of course, 26 of the 56 children already declared unusual are not included in the 35 children that are identified as joint outliers.

Future children could be judged against the thresholds obtained in this way. Choosing a value higher than 5% to define the thresholds would lead to more children being declared as outliers.

The fact that the bivariate copula approach that we have developed in this paper allows additional children to be identified as having unusual sight is an important feature of our methodology. Figure 15, which presents the equivalent bottom right panel of Figure 14 for all five age groups $(3, 4]$, $(4, 5]$, $(5, 6]$, $(6, 7]$ and $(7, 8]$ years, shows all the additional children in our data set that would be identified as unusual (black triangles). These children can then undergo further examinations to ascertain the exact nature of their unusual sight and can be treated accordingly.

5 Discussion

Statistical models of ophthalmic data should naturally consider the joint distribution of outcome variables on both eyes. In this paper we fitted bivariate copula regression models to logMar measurements from both eyes of 2721 children with dependence parameters conditional on their ages. In particular, we modelled the dependence parameter of the copula using an additive term of the child's age. We did this because the relation between logMar values from both eyes is an important aspect of the research question.

Our modelling is based on the additive conditional bivariate copula regression approach of Klein and Kneib [27], which itself offers considerable generality in the specification of the predictors. As the response is highly asymmetric, we modified their model by adopting sinh-arcsinh marginal distributions with location, scale and lower and upper tail shape parameters also modelled as additive terms of age. Marra and Radice [32] have recently developed similar frequentist models.

We performed inference about the unknown parameters of our model in the Bayesian framework, using a Metropolis-within-Gibbs MCMC algorithm. In this algorithm the proposal densities for the spline parameters were obtained by approximating the associated log full conditional probability density functions by normal densities. The means and covariances matrices of these normal densities were found using numerical optimization, so avoiding complicated mathematics. We worked with the t, Gaussian, Clayton-0-90, Clayton-180-270, Gumbel-0-90, Gumbel-180-270 and Frank bivariate copula families and performed a simulation study to confirm the utility of the Watanabe-Akaike Information Criterion as a goodness-of-fit type statistic to select the copula family.

Our results show that marginally for each eye the location of the logMAR distribution decreases as child age increases, and this allows us to quantify the rate at which visual acuity improves. We also found that the dependence between the eyes' visual resolution tends to become stronger across the age range. The predictive distribution associated with our model supports most of the data well and was used to highlight children with unusual sight characteristics, distinguishing those whose logMar values are bivariate, but not univariate outliers. This implies that considering bivariate outliers identifies children with unusual sight who would otherwise be undetected. These children might require further tests or corrective treatment.

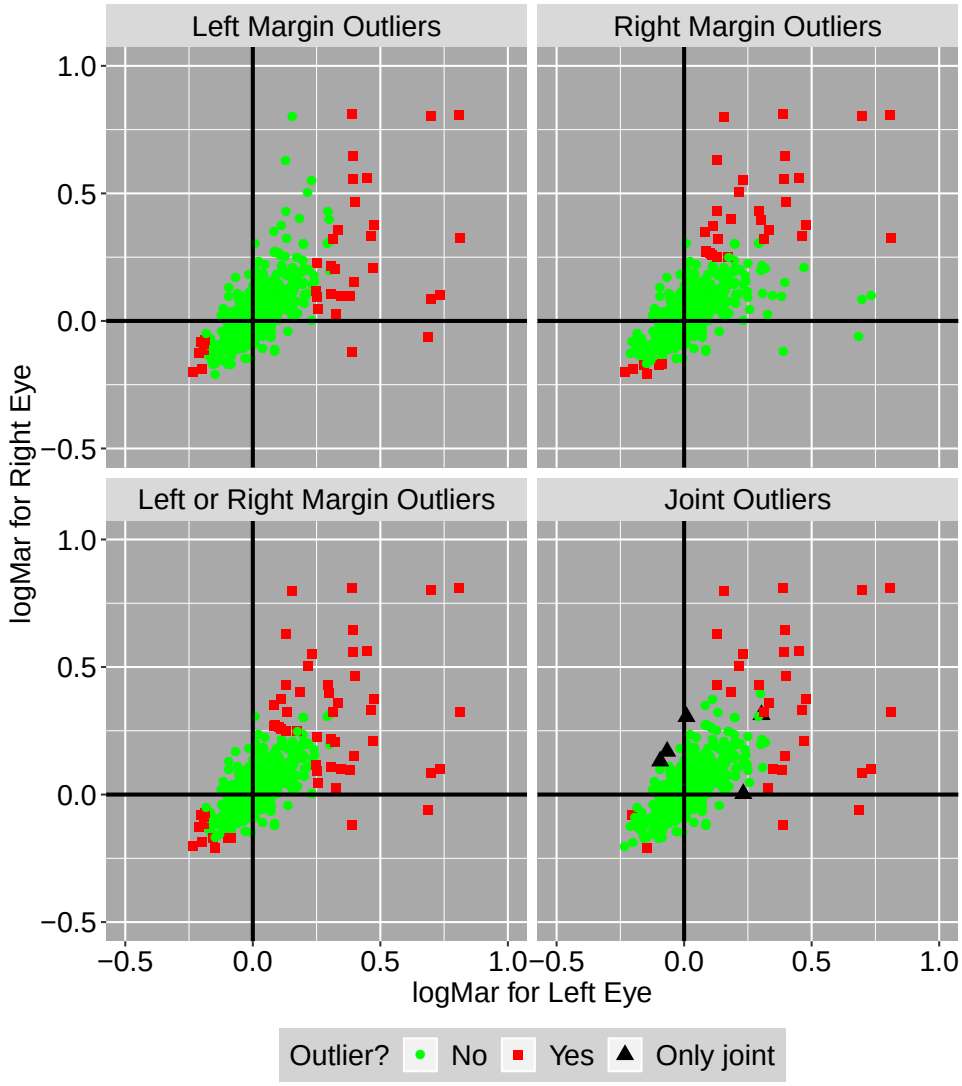

Figure 14: logMar values for children in the $(4, 5]$ years age group. Top row: unusual points based on the marginal predictive distributions pred^{L} and pred^{R} for left and for right logMar are shown using red squares. These points correspond to approximately 5% of children in this age group (35 points). Bottom left: the unusual points are the union of those shown in the top row (56 points). Bottom right: unusual points based on the joint predictive distribution $\text{pred}^{\text{L, R}}$ for left and right logMar. Again, these points comprise approximately 5% of those in this age group (35 points). Five points that are identified as outliers only when based on $\text{pred}^{\text{L, R}}$ are indicated using a triangle.

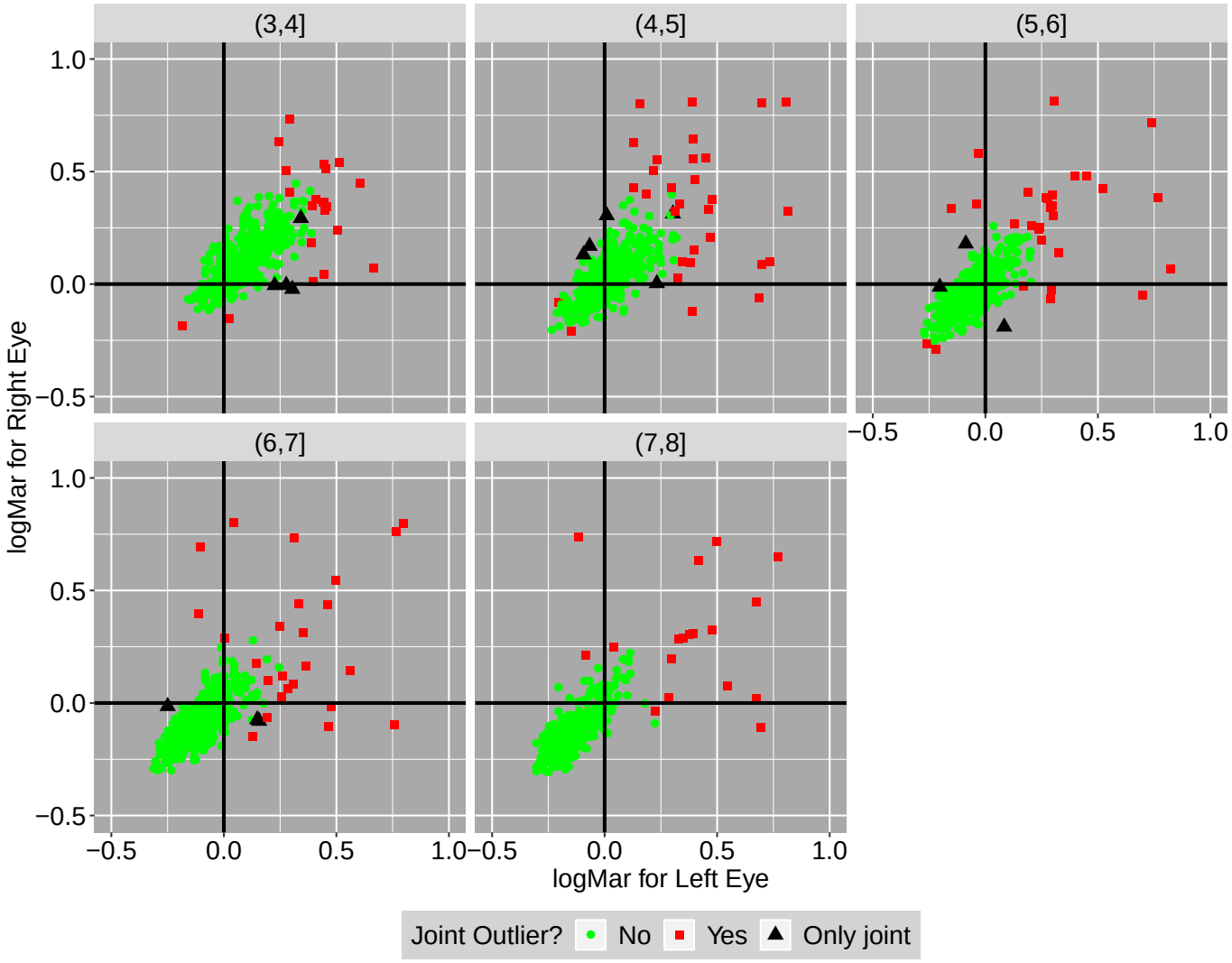

Figure 15: logMar values for children in the $(3, 4]$, $(4, 5]$, $(5, 6]$, $(6, 7]$ and $(7, 8]$ years age groups. Unusual points based on the joint predictive distribution $\text{pred}^{\text{L, R}}$ are shown. These points correspond to approximately 5% of children in each age group. The points that are identified as outliers only when based on $\text{pred}^{\text{L, R}}$ are indicated using a black triangle. These are the additional children that our bivariate copula approach identified as unusual.

Our Bayesian estimation method and the two-step frequentist procedure of Vatter and Chavez-Demoulin [56] both use additive models to fit copulas with rich marginal families and conditional dependence structure. Moreover, simulation studies confirmed similarly useful roles for AIC in the frequentist setting and WAIC in the Bayesian approach for copula selection. However, there are some important differences. First, in the Vatter and Chavez-Demoulin [56] procedure the marginal distributions are found in an initial step, after which the dependence relationship can be estimated using the R package `gamCopula` [36]. This gives it the advantages over our simultaneous procedure of simplicity of implementation, higher generality and computational speed. Secondly, the uncertainity associated with the estimation of the marginal distributions and the smoothing parameters does not fully propagate to the estimation of the copula parameters in the two-step approach. In addition, the Bayesian approach allows us to easily produce the posterior distribution and the interval estimates of any quantity of interest, which is a function of the parameters. For example one can be interested in the probability that a given child will have a visual acuity beyond a given threshold. This is something more difficult to perform in a frequentist set-up, and this is a bonus coming from the more complex computation.

Our future plans include extending the model to higher dimensions building on a vine-based approach [6, 12, 57]. We will also work with two-parameter bivariate copulas, such as BB type copulas [22, 23, 40] in order to control the behaviours of the two copula tails separately.

Acknowledgements

The first, second and fourth authors were supported by a Royal Society International Exchange Scheme grant for a project entitled 'High-Dimensional Bayesian Dependence Modelling with Conditional Copulas'. Some work for this study was performed at the Great Ormond Street Hospital, and the UCL Great Ormond Street Institute of Child Health, which received funding from the UK Department of Health's NIHR Biometical Research Centres funding scheme. No funders had any role in the design or analysis of the study or in the writing of this article.

We thank Professor Simon Wood for extremely valuable and rapid help concerning a computational aspect of our work, and Luca Bolognesi & Professor Claudia Czado for stimulating discussions. We are extremely grateful to the editor, Professor Louise Ryan, an associate editor and three reviewers for highly insightful comments that have led to very considerable improvements in this paper.

There are no conflicts of interest.

References

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

[2] E. Acar, R. Craiu, and F. Yao. Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453, 2011.

[3] E. Acar, R. Craiu, and F. Yao. Statistical testing of covariate effects in conditional copula models. *Electronic Journal of Statistics*, 7:2822–2850, 2013.

[4] H. Akaike. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AU-19:716–722, 1974.

[5] C. Almeida, C. Czado, and H. Manner. Modeling high-dimensional time-varying dependence using dynamic d-vine models. *Applied Stochastic Models in Business and Industry*, 32(5):621–638, 2016.

[6] E. Brechmann and U. Schepsmeier. Modeling dependence with c- and d-vine copulas: The R package cdvine. *Journal of Statistical Software*, 52:1–27, 2013.

[7] S. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[8] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, editors. *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, Boca Raton, 2011.

[9] A. Conlon, J. Taylor, and M. Elliott. Surrogacy assessment using principal stratification and a Gaussian copula model. *Statistical Methods in Medical Research*, 26(1):88–10, 2017.

[10] R. Craiu and A. Sabeti. In mixed company: Bayesian inference for conditional copulas models with discrete and continuous outcomes. *Journal of Multivariate Analysis*, 110:106–120, 2012.

[11] R. dos Santos Silva and H. Lopes. Copula, marginal distributions and model selection: a bayesian note. *Statistics and Computing*, 18(3):313–320, 2008.

[12] T. Erhardt, C. Czado, and U. Schepsmeier. R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71(2):323–332, 2015.

[13] F. Ferris and I. Bailey. Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the eye care technology forum. *Ophthalmology*, 103:181–182, 1996.

[14] D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7:57–68, 1997.

[15] A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.

[16] A. Gelman and D. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

[17] C. Geyer. Practical markov chain monte carlo. *Statistical Science*, 7:473–483, 1992.

[18] I. Gijbels, N. Veraverbeke, and M. Omelka. Conditional copulas, association measures and their applications. *Computational Statistics and Data Analysis*, 55:1919–1932, 2011.

[19] P. Green and B. Silverman. *Nonparametric Regression and Generalized Linear Models*. Chapman & Hall, London, 1994.

[20] E. Hartmann, V. Dobson, L. Hainline, W. Marsh-Tootle, G. Quinn, M. Ruttum, and Others. Pre–school vision screening: Summary of a task force report. On behalf of the maternal and child health bureau and the national eye institute task force on vision screening in the preschool child. *Pediatrics*, 106:1105–1116, 2000.

[21] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, New York, 2001.

[22] H. Joe. *Dependence Modelling with Copulas*. Chapman and Hall/CRC, Boca Raton, 2014.

[23] H. Joe and T. Hu. Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57:240–265, 1996.

[24] M. Jones. Contribution to the discussion of [45]. *Applied Statistics*, 54:546–547, 2005.

[25] M. Jones and A. Pewsey. Sinh-arcsinh distributions. *Biometrika*, 96:761–780, 2009.

[26] M. Kalloniatis and A. Johnston. Visual characteristics of low vision children. *Optometry and Vision Science*, 67:38–48, 1990.

[27] N. Klein and T. Kneib. Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. *Statistics and Computing*, 26:1–20, 2015.

[28] N. Klein, T. Kneib, S. Klasen, and S. Lang. Bayesian structured additive distributional regression for multivariate responses. *Applied Statistics*, 64:569–591, 2015.

[29] N. Klein, T. Kneib, and S. Lang. Bayesian structured additive distributional regression. *Working Papers in Economics and Statistics, University of Innsbruck*, 2013. `https://www.econstor.eu/handle/10419/101101`.

[30] P. Lancaster and K. Šalkauskas. *Curve and Surface Fitting: An Introduction*. Academic Press, London, 1986.

[31] S. MacEachern and M. Berliner. Subsampling the gibbs sampler. *The American Statistician*, 48:188–190, 1994.

[32] G. Marra and R. Radice. A Bivariate Copula Additive Model for Location, Scale and Shape. *Computational Statistics and Data Analysis*, 112:99–113, 2017.

[33] G. Marra and S. Wood. Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39:53–74, 2012.

[34] A. Min and C. Czado. Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, 8(4):511–546, 2010.

[35] A. Min and C. Czado. Bayesian model selection for multivariate copulas using pair-copula constructions. *The Canadian Journal of Statistics*, 39(2):239–258, 2011.

[36] T. Nagler and T. Vatter. *gamCopula: Generalized Additive Models for Bivariate Conditional Dependence Structures and Vine Copulas*, 2017. R package version 0.0-4.

[37] R. Nelsen. *An Introduction to Copula*. Springer, New York, second edition, 2006.

[38] E. Neuwirth. *RColorBrewer: ColorBrewer Palettes*, 2014. R package Version 11-2.

[39] A. Nikoloulopoulos. A mixed effect model for bivariate meta-analysis of diagnostic test accuracy studies using a copula representation of the random effects distribution. *Statistics in Medicine*, 34(29):3842–3865, 2015.

[40] A. Nikoloulopoulos, H. Joe, and H. Li. Vine copulas with asymmetric tail dependence and applications to financial return data. *Computational Statistics and Data Analysis*, 56:3659–3673, 2010.

[41] I. Ntzoufras. *Bayesian Modeling Using WinBUGS*. Wiley, Hoboken, 2009.

[42] C. Peckham and C. Dezateux. Issues underlying the evaluation of screening programmes. *British Medical Bulletin*, 54:767–778, 1998.

[43] C. T. R. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[44] J. Rahi, P. Cumberland, and C. Peckham. Myopia over the lifecourse: prevalence and early life influences in the 1958 British birth cohort. *Ophthalmology*, 118(5):797–804, 2011.

[45] R. Rigby and D. Stasinopoulos. Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, 54:507–554, 2005.

[46] A. Sabeti, M. Wei, and R. Craiu. Additive models for conditional copulas. *STAT: The ISI's Journal for the Rapid Dissemination of Statistics Research*, 3:300–312, 2014.

[47] A. Salt, A. Wade, R. Proffitt, S. Heavens, and P. Sonksen. The Sonksen logmar test of visual acuity: I. testability and reliability. *Journal of American Association of Pediatric Ophthalmology and Strabismus*, 11:589–596, 2007.

[48] S. Sheather and M. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 53:683–690, 1991.

[49] B. Silverman. *Density Estimation*. Chapman and Hall/CRC, Boca Raton, 1986.

[50] A. Sklar. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8:229–231, 1959.

[51] P. Sonksen, A. Wade, R. Proffitt, S. Heavensc, and A. Salt. The Sonksen logmar test of visual acuity: Ii. age norms from 2 years 9 months to 8 years. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 12(1):18–22, 2008.

[52] D. Spiegelhalter, N. Best, B. Carlin, and A. Van der Linde. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 64:583–639, 2002.

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

[54] J. Stander, L. Dalla Valle, and M. Cortina-Borja. A bayesian survival analysis of a historical dataset: How long do popes live? *The American Statistician*, 72(4):368–375, 2018.

[55] P. Thompson, Y. Cai, R. Moyeed, D. Reeve, and J. Stander. Bayesian nonparametric quantile regression using spline. *Computational Statistics and Data Analysis*, 54:1138–1150, 2010.

[56] T. Vatter and V. Chavez-Demoulin. Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141:147–167, 2015.

[57] T. Vatter and T. Nagler. Generalized Additive Models for Pair-Copula Constructions. *ArXiv e-prints*, Aug. 2017. `https://arxiv.org/abs/1608.01593v2` (visited: 2018-03-19).

[58] W. Venables and B. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002.

[59] C. Villa and F. Rubio. Objective priors for the number of degrees of freedom of a multivariate t distribution and the t-copula. *Computational Statistics & Data Analysis*, 124:197–219, 2018.

[60] L. WA and E. MJ. On thinning of chains in mcmc. *Methods in Ecology and Evolution*, 3:112–115, 2012.

[61] A. Wade, A. Ades, A. Salt, R. Jayatunga, and P. Sonksen. Age-related standards for ordinal data: modelling the changes in visual acuity from 2 to 9 years of age. *Statistics in Medicine*, 14:257–266, 1995.

[62] A. Wade, A. Salt, R. Proffitt, S. Heavens, and P. Sonksen. Likelihood-based modelling of age-related normal ranges for ordinal measurements: changes in visual acuity through early childhood. *Statistics in Medicine*, 23:3623–3640, 2004.

[63] S. Watanabe. A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, 14:867–897, 2013.

[64] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York, second edition, 2016.

[65] S. Wood. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, Boca Raton, second edition, 2017.

[66] J. Yan. Enjoy the joy of copulas: with a package `copula`. *Journal of Statistical Software*, 21(4):1–21, 2007.

[67] Y. Zhong and R. Cook. Augmented composite likelihood for copula modeling in family studies under biased sampling. *Biostatistics*, 17(3):437–452, 2016.