Journal of Multivariate Analysis 110 (2012) 106–120

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes

V. Radu Craiu *, Avideh Sabeti

Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario M5S 3G3, Canada

ARTICLE INFO

Article history:
Available online 2 April 2012

AMS subject classifications:
62H05
62F15
62C40

Keywords:
Adaptive Markov chain Monte Carlo
Bayesian inference
Conditional copula model
Cubic spline
Deviance information criterion
Mixed outcomes

ABSTRACT

Conditional copula models are flexible tools for modelling complex dependence structures in regression settings. We construct Bayesian inference for the conditional copula model adapted to regression settings in which the bivariate outcome is continuous or mixed. The dependence between the copula parameter and the covariate is modelled using cubic splines. The proposed joint Bayesian inference is carried out using adaptive Markov chain Monte Carlo sampling. The deviance information criterion (DIC) is used for selecting the copula family that best approximates the data and for choosing the calibration function. The performances of the estimation and model selection methods are investigated using simulations.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Central to modern statistical analysis is modelling and understanding the dependence between random variables. However, the number of options at the statistician's disposal is limited due to the sparsity of available multivariate distributions. Copulas represent a flexible alternative in which one can bypass the use of multivariate distributions by using Sklar's theorem [48] to model separately the marginal distributions and the joint dependence structure.

In this paper we consider the bivariate case, as the extension to more than two responses requires a considerably more complex statistical machinery. Let Y_1 and Y_2 be continuous random variables of interest with joint distribution function H and marginal distributions F_1 and F_2, respectively. Sklar's theorem ensures the existence of a unique copula $C : [0, 1]^2 \to [0, 1]$, which satisfies $H(y_1, y_2) = C\{F_1(y_1), F_2(y_2)\}$, for all $(y_1, y_2) \in \mathbb{R}^2$. Often, the function C assumes a parametric form indexed by a copula parameter θ. A nice introduction to copulas and their properties is [38] and the connections between various copulas and dependence concepts are discussed in detail in [31].

It should be noted that copula models have been mostly used for dependence between continuous random variables (e.g., [9,10,14,17,28,46,59]) but there is growing interest in the study and applications of copula models for mixed (discrete and continuous) data [21,13,42,52,53,55]. Recent work by [21,49] has shown that extra care is needed in performing and interpreting statistical inference for copula models when some of the marginals are discrete.

While frequentist methods have been used predominantly in the copula literature (see, for instance, [20,22]), the availability of powerful computers and "off-the-shelf" algorithms have recently led to a few Bayesian methods for copula estimation, model selection and goodness-of-fit (see, for instance, [29,34,37,42,47,49]).

A natural extension of the classical copula model allows the copula parameter to vary with covariate values as in [35]. This idea, formalized by the conditional copula model of [40], allows for realistic copula modelling in regression settings

* Corresponding author.
E-mail addresses: craiu@utstat.toronto.edu (V.R. Craiu), sabeti@utstat.toronto.edu (A. Sabeti).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.03.010

[3,23]. For instance, if, in addition to Y_1 and Y_2, we have information on a covariate X, then the influence of X on the dependence between Y_1 and Y_2 can be modelled by the conditional copula $C(\cdot \mid X)$, which is the joint distribution function of $U_1 \equiv F_{1|X}(Y_1 \mid x)$ and $U_2 \equiv F_{2|X}(Y_2 \mid x)$ given $X = x$, where $Y_i \mid X = x$ has cdf $F_{i|X}(\cdot \mid x)$, $i = 1, 2$. For each x in the support of X, it was showed in [40] that for continuous random variables the joint conditional distribution is uniquely defined by

$$H_X(y_1, y_2 \mid x) = C\{F_{1|X}(y_1 \mid x), F_{2|X}(y_2 \mid x) \mid x\}, \quad \text{for all } (y_1, y_2) \in \mathbb{R}^2.$$

In general, it is not known how the copula parameter varies with the covariate X. Moreover, visualizing the dependence patterns in the data is not possible as the available samples are not identically distributed (usually there is no replication at a given value of X). Therefore, one needs to use flexible methods to model the *calibration function* that characterizes the relationship between the copula parameter and the covariates. This naturally leads to the use of semiparametric [3] and nonparametric inferential tools [39,58] in the case of continuous marginals. Conditional copulas have been used mostly in the context of financial time series [5,32,40,41] but have recently been used in classical regression models too [3].

In this paper we propose joint Bayesian inference for a conditional copula model in which of interest, besides the marginal models, is the dependence between the outcomes. Our approach can handle the case of continuous and mixed (binary and continuous) outcomes. The Bayesian approach proposed here offers a number of advantages. First, it is a principled method to produce full likelihood-based inference as recommended by [21,49] in the case of copula models with some of the margins discrete. Second, the Bayesian approach offers a natural solution to the "propagation of errors" challenge in which the variance of the marginal model estimators must be accounted for when assessing the variance of the copula estimators. Within the Bayesian paradigm, the posterior distribution gives a full representation of the uncertainty in the *whole* data and prior. An added bonus is that the simultaneous estimation of both marginal distributions' parameters and copula parameters results in better understanding of the parameter dependences and leads to better performance of model selection criteria, as discussed in [47].

The shape of the calibration function is difficult to model in general and we thus need to use flexible models to capture its structure. In this paper we consider a Bayesian cubic spline model in which the choice for the position of the knots is data-driven. Sampling from the posterior distribution is performed using an adaptive MCMC algorithm following the principles developed in [26]. For more details regarding the theory and implementation of adaptive MCMC, we refer to [11,25,43].

In the next section we detail the statistical model along with the prior specification. Section 3 contains the computational algorithm used to sample from the posterior and the criterion used for model selection. The simulation study is summarized in Section 4. The paper closes with a discussion in which future directions are outlined.

2. The model

In this section we describe the conditional copula model for the dependent outcome data. We separate the model formulations used for continuous and for mixed outcomes (binary and continuous), as the two situations require different computational algorithms to sample from the posterior distribution and the resulting estimators exhibit different efficiency.

2.1. Continuous outcomes case

The data consist of bivariate continuous random variables (V_1, V_2) and covariate X measured for n samples. Marginally, V_i and X are related through $V_i \sim \mathcal{N}(X\beta_i, \sigma_i^2)$, for $i = 1, 2$. The conditional dependence between V_1 and V_2 is defined via a conditional copula model with joint density

$$f(V_1, V_2|X) = \prod_{i=1}^{2} \frac{1}{\sigma_i}\phi\left(\frac{V_i - X\beta_i}{\sigma_i}\right) \times c^{(1,1)}\left\{\Phi\left(\frac{V_1 - X\beta_1}{\sigma_1}\right), \Phi\left(\frac{V_2 - X\beta_2}{\sigma_2}\right)\middle| \theta(X)\right\},$$

where $c^{(a,b)}(u, v|\theta) = \partial^{a+b}C(u, v|\theta)/\partial u^a \partial v^b$, for all $0 \le a, b \le 1$.

An important part of the model is the specification of $\theta(X)$. In this paper we assume that $X \in \mathbb{R}$ and follow [3,4] by assuming that $g(\theta) = \eta(X)$ where g is a known function that maps the support of the copula parameter onto the real line and $\eta : \mathbb{R} \to \mathbb{R}$ is the unknown *calibration function* we want to estimate. We adopt the flexible cubic spline model suggested by [50] in which

$$\eta(z) = \sum_{j=0}^{3} \alpha_j z^j + \sum_{k=1}^{K} \psi_k (z - \gamma_k)_+^3, \tag{1}$$

where $a_+ = \max(0, a)$. It is well known that the performance of spline-based estimators are influenced by the location of the knots $\gamma_1, \ldots, \gamma_K$. In our model this choice is automatic and data-driven.

2.2. Mixed outcomes case

In the mixed outcome case, the response consist of one binary and one continuous random variable, denoted Q and W, respectively. We are interested in statistical inference for the marginal logistic and linear regression models but also the estimation of the calibration function.

For the response (Q, W) with covariate X we assume the marginal models

$$\Pr(Q = 1|X) = \frac{\exp(X\beta_1)}{1+\exp(X\beta_1)},$$
$$W|X \sim \mathcal{N}(X\beta_2, {\sigma_2}^2).$$

The dependence between Q and W is defined using a parametric conditional copula $C\{\cdot,\cdot|\theta(X)\}$. Using the derivations in [53] we obtain the contribution of the jth sample to the likelihood

$$\Pr(Q_j = a, W_j|X_j, \omega) = \frac{1}{\sigma_2}\phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right) \times \left[c^{(0,1)}\left\{\frac{\exp(aX_j\beta_1)}{1+\exp(X_j\beta_1)}, \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\}\right]^{1-a} \times \left[1 - c^{(0,1)}\left\{\frac{\exp\{(1-a)X_j\beta_1\}}{1+\exp(X_j\beta_1)}, \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\}\right]^{a}, \tag{2}$$

where $a \in \{0, 1\}$ and ω represents the vector of all the parameters involved in the model.

The computational challenges encountered when performing Bayesian inference for the logistic model are alleviated if one uses a latent variable formulation in which $Y \sim \text{Logistic}(X\beta_1)$ and $Q = \mathbf{1}(Y \geq 0)$, where Logistic(m) has density $f_L(y|m) = \exp(m-y)/\{1+\exp(m-y)\}^2$ for any $y \in \mathbb{R}$. We extend the above formulation to the conditional copula model (2). Specifically, we assume that the dependence between the latent variable Y and W is characterized by the same conditional copula $C\{\cdot,\cdot|\theta(X)\}$ as in (2). If observed, the contribution of the jth sample to the complete data likelihood would be

$$f(Y_j, W_j|X_j, \omega) = f_L\left(Y_j|X_j\beta_1\right)\frac{1}{\sigma_2}\phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right) \times c^{(1,1)}\left\{F_L\left(Y_j|X_j\beta_1\right), \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\}. \tag{3}$$

It is not hard to see that (2) can be obtained from (3) by averaging over the latent variables. For instance, since $Q_j = \mathbf{1}(Y_j \geq 0)$,

$$\begin{aligned}
&\Pr(Q_j = 0, W_j|X_j, \omega) = \Pr(Y_j < 0, W_j|X_j, \omega)\\
&= \int_{-\infty}^{0} f_L\left(Y_j|X_j\beta_1\right) \times c^{(1,1)}\left\{F_L\left(Y_j|X_j\beta_1\right), \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\} \mathrm{d}Y_j \times \frac{1}{\sigma_2}\phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\\
&= \int_{-\infty}^{0} \frac{\partial}{\partial Y_j} c^{(0,1)}\left\{\frac{1}{1+\exp(X_j\beta_1 - Y_j)}, \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\} \mathrm{d}Y_j \times \frac{1}{\sigma_2}\phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\\
&= c^{(0,1)}\left\{\frac{1}{1+\exp(X_j\beta_1)}, \Phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right)\middle|\theta(X_j)\right\}\frac{1}{\sigma_2}\phi\left(\frac{W_j - X_j\beta_2}{\sigma_2}\right),
\end{aligned}$$

and the last expression is exactly (2) in which $a = 0$. Although the two models (2) and (3) are equivalent, the latent variable formulation leads to the efficient data augmentation algorithm that will be discussed in detail in Section 3.

The usual dependence measures can be used in the bivariate continuous and mixed outcome models. However, in the case in which at least one variable is discrete, the dependence parameters are functions of all the parameters in the model, as discussed by [21,49]. For instance, conditional on the covariate X, the population value of the Kendall's tau between Q and W is

$$\tau(\omega|X) = 4\mathrm{E}\{H(Q, W|X)|X\} - 1 = \frac{3 + 2\exp(\beta_1 X) + 3\exp(2\beta_1 X)}{\{1+\exp(\beta_1 X)\}^2} - 4\int_{\mathbb{R}}\left[\frac{1}{\sigma_2}\phi\left(\frac{w - X\beta_2}{\sigma_2}\right) \times C\left\{\frac{1}{1+\exp(X\beta_1)}, \Phi\left(\frac{w - X\beta_2}{\sigma_2}\right)\middle|\theta(X)\right\}\right]\mathrm{d}w, \tag{4}$$

where $H(\cdot,\cdot|X)$ is the conditional joint cdf of (Q, W) given X. The integral in (4) is generally intractable but can be approximated via Monte Carlo integration. Moreover, the parameters $\theta(x)$ and $\tau(x)$ are no longer in a one-to-one correspondence, as is the case for continuous outcomes. Implicitly, one has to decide whether the model is parametrized using θ or τ and in this paper we choose to work with the former.

2.3. Prior specification

While the prior distributions used in a normal regression model are well understood (e.g., see [18]), recent work of [6,15] has shown that the same vague priors have a very different effect on the coefficients of a logistic regression model. Therefore, the priors for $(\beta_1, \beta_2, \sigma_2)$ are different for the continuous and mixed outcome models. In our simulations we use only one covariate. For the continuous outcome model we have specified the following priors

$$\beta_1 \sim \mathcal{N}(0, 10),$$
$$\beta_2 \sim \mathcal{N}(0, 10),$$
$$\sigma_2 \sim \text{InvGam}(0.1, 0.1).$$

In the mixed outcome case, we choose the prior distributions for the marginal logistic model so that the a priori marginal distribution of $p_Q = \Pr(Q = 1)$ is approximately Unif(0, 1), as recommended by [6].

$$\beta_1 \sim \mathcal{N}(0, \tau^2),$$
$$\beta_2 \sim \mathcal{N}(0, 10),$$
$$\sigma_2 \sim \text{InvGam}(0.1, 0.1),$$

where the choice for τ depends on the range of the covariate X. In our simulations, $-1 \le X \le 1$ and $\tau = 3$.

For the prior specification of the parameters involved in the cubic spline we follow [16] and fix the maximum number of knots that one allows in the model, denoted by $K_{\max}$. The range spanned by the observed values of covariate X is divided into $K_{\max}$ intervals of equal length, $I_1, \ldots, I_{K_{\max}}$, and we assume that each interval I_k contains at most one knot. In order to complete the model specification, we introduce the auxiliary random variables $\{\zeta_k : 1 \le k \le K_{\max}\}$, where for all $k \in \{1, \ldots, K_{\max}\}$

$$\zeta_k = \begin{cases} 1 & \text{if there is a knot } \gamma_k \in I_k, \\ 0 & \text{otherwise.} \end{cases}$$

The model (1) becomes then

$$\eta(z) = \sum_{j=0}^{3} \alpha_j z^j + \sum_{k=1}^{K_{\max}} \zeta_k \psi_k (z - \gamma_k)_+^3 \tag{5}$$

and one can see that the number of non-zero terms in the sum depends on the values of $\zeta_1, \ldots, \zeta_{K_{\max}}$. We expand the model to include the latent variables ζ_k and define their sampling distributions. Specifically, if we let $|\zeta| = \sum_{k=1}^{K_{\max}} \zeta_k$ be the number of knots that are used in the model then

$$p(|\zeta| \mid \lambda) \propto \frac{\lambda^{|\zeta|}}{|\zeta|!} \mathbf{1}_{\{|\zeta| \le K_{\max}\}},$$

i.e., $|\zeta|$ follows the right truncated Poisson distribution with parameter λ, and maximum value $K_{\max}$. In addition,

$$p(\zeta \mid |\zeta|) = \binom{K_{\max}}{|\zeta|}^{-1},$$
$$p(\zeta|\lambda) = p(\zeta \mid |\zeta|) p(|\zeta| \mid \lambda).$$

The form of $p(\zeta \mid |\zeta|)$ implies that, given a number of knots for the model, all configurations of intervals containing a knot are equally likely. The priors for all the parameters involved in the spline model for η are chosen regardless of the type of outcome as

$$\lambda \sim \text{Bin}(K_{\max}, p = 0.5),$$
$$\alpha_j \sim \mathcal{N}(0, 10), \quad \forall 0 \le j \le 3$$
$$\psi_k \sim \mathcal{N}(0, 10), \quad \forall 1 \le k \le K_{\max}$$
$$\gamma_k \sim \text{Unif}[I_k], \quad \forall 1 \le k \le K_{\max}.$$

Unlike [16] who pre-select and fix the value for λ, we let the choice be data-driven. Similar hierarchical models used in Bayesian regression have been shown to correct for multiplicity; see [12,45]. This was motivated by simulations showing that an unsuitable value for λ can result in poor fits. One can see that the number of parameters grows linearly with $K_{\max}$ and we have used throughout our simulations $K_{\max} = 4$. Our simulation studies have showed that using a large $K_{\max}$ does not improve the fit, but slows down significantly the computational algorithm we use to sample from the posterior distribution.

3. Estimation and model selection

The Bayesian paradigm dictates that inference must be based on π, the posterior distribution of the parameter vector. Given the complexity of both the continuous and mixed outcome models, it is no surprise that π is analytically intractable and its properties can be studied only via Monte Carlo methods. In the next section we describe the Markov chain Monte Carlo (MCMC) algorithms [7,24] used to sample from π.

3.1. Markov chain Monte Carlo sampling from the posterior distribution

The MCMC algorithms for the continuous and mixed outcomes are very similar. We detail below the sampler for the mixed outcome and the section ends with a remark concerning the changes required for the model with continuous outcomes.

3.1.1. Mixed outcome case

The general form of the algorithm follows the data augmentation (DA) principle [56] based on which we build an ergodic Markov chain with values in the parameter space and whose stationary distribution is π. The DA chain is built by sampling alternatively from: (a) the conditional distribution of the latent (or missing data), $\mathcal{D}_{\text{mis}} = (\vec{Y}, \vec{\zeta})$ given the parameters and the observed data $\mathcal{D}_{\text{obs}} = (\vec{Q}, \vec{W}, \vec{X})$, and (b) the conditional distribution of the parameter ω given the complete data, $\mathcal{D}_{\text{com}} = (\mathcal{D}_{\text{mis}}, \mathcal{D}_{\text{obs}})$. Again, the conditional distributions needed for (a) and (b) do not have canonical forms, so we use either the Random Walk Metropolis (RWM) or the Independent Metropolis (IM) algorithm to update each component in a systematic-scan Gibbs-like scheme.

Denote $(\omega_t, \vec{Y}_t, \vec{\zeta}_t)$ the state of the Markov chain at time t, where $\omega = (\beta_1, \beta_2, \sigma_2, \vec{\gamma}, \vec{\alpha}, \vec{\psi}, \lambda)$, (we use $\vec{\gamma}$ to denote the vector of all the knots used, and $\vec{\alpha}$, $\vec{\psi}$ are defined similarly). We detail below the update at time $t+1$ of each parameter and missing data value:

β's: For each $\beta_i, i = 1, 2$ we use a RWM-within-Gibbs to sample from the conditional distribution $\pi(\beta_i|\omega_t \setminus \beta_i, \mathcal{D}_{\text{mis}}, \mathcal{D}_{\text{obs}}) \propto \pi(\omega|\mathcal{D}_{\text{mis}}, \mathcal{D}_{\text{obs}})$. The proposal in this case is $\mathcal{N}(\beta_{i,t}, s_t^2)$. During an initialization period of 1000 iterations, the standard deviation of the proposal is kept fixed and afterwards it is adapted as $s_t = \sqrt{SV(\beta_{i,1:t}) + \epsilon}$, where $SV(\beta_{i,1:t})$ is the sample variance of the samples $\beta_{i,1}, \ldots, \beta_{i,t}$ and $\epsilon = 0.01$ is introduced in order to safeguard against degenerate situations where SV may be zero. Note that the normalizing constant for the conditional distribution $\pi(\beta_i|\omega_t \setminus \beta_i, \mathcal{D}_{\text{mis}}, \mathcal{D}_{\text{obs}})$ does not influence the acceptance probability needed for the RWM transition kernel. This is generally valid for all the conditional distributions described below and explains why the algorithm can be implemented even when the conditional distributions have intractable normalizing constants.

σ_2: We prefer to work with unrestricted state spaces, so we parametrize the model in terms of $\xi_2 = \ln \sigma_2$ and run the MCMC chain for ξ_2. After the sampling is completed, we transform the ξ_2-samples back to the σ_2 scale. Because the range of ξ_2 is $\mathbb{R}$, we can use the RWM-within-Gibbs transitions to sample from the conditional distribution $\pi(\xi_2|\omega_t \setminus \xi_2) \propto \pi(\omega|\mathcal{D}_{\text{mis}}, \mathcal{D}_{\text{obs}})$. The same adaptive procedure is used as for the β updates.

Y's: The latent variables Y_i are updated using an IM-within-Gibbs. Note that the auxiliary variables $Y_1, \ldots, Y_n$ are different from the ones used by [42,49]. Without the copula structure in the model, the conditional distribution of $Y_i|(\omega_t, \mathcal{D}_{\text{mis}} \setminus Y_i, \mathcal{D}_{\text{obs}})$ is truncated logistic (truncated to $(0, \infty)$ if $Q_i = 1$ and to $(-\infty, 0)$ if $Q_i = 0$) and can be sampled using the inverse cdf method. We use the truncated logistic as the independent proposal distribution in each update. The observed acceptance rates are above 80%.

α's: There is no range restriction for each α_i and no direct sampling strategy is possible, so we use the RWM-within-Gibbs with the same adaptive regime as the one described for β.

ψ's: If $\zeta_{k,t} = 1$ we use the RWM-within-Gibbs strategy to update ψ_k using the same adaptive regime as the one described for β. If $\zeta_{k,t} = 0$, ψ_k is not updated and thus $\psi_{k,t+1} = \psi_{k,t}$.

ζ's: The updates are performed using the Metropolis-within-Gibbs strategy for the *entire* latent variable vector $\vec{\zeta} = (\zeta_1, \ldots, \zeta_{K_{\max}})$. For updating $\vec{\zeta}$ we use two type of moves: we either add/delete a component (i.e. transforming a zero component into a one or vice versa) or swap two components. In our applications, we choose with probability $1/2$ to add/delete a component chosen at random and with probability $1/2$ to permute two components of $\vec{\zeta}$ that are selected at random.

γ's: If $\zeta_{k,t} = 1$ we use an IM update for γ_k using as proposal the prior distribution of γ_k. If $\zeta_{t,k} = 0$ then $\gamma_{t+1,k}$ is not updated.

λ: For λ we use an IM update with proposal distribution equal to the prior, i.e. $\text{Bin}(0.5, K_{\max})$.

This automatic approach to tuning the sampling algorithm leads to a more balanced comparison between the models that are fit under various copula families. This is desirable because Bayesian model comparison results depend crucially on the ability of sampling equally well from the posterior distributions produced by all the models considered.

Remark. Samples from the posterior distribution in the continuous case are obtained using an algorithm very similar to the one we just described. All the parameters and the auxiliary data $\zeta_1, \ldots, \zeta_{K_{\max}}$ are updated using the same strategies. However, an important difference is that we do not need to introduce the auxiliary variables $Y_1, \ldots, Y_n$, so we obtain in this case a faster MCMC algorithm.

3.2. Model selection

We distinguish two model selection problems associated with the conditional copula models discussed here. The first one involves the choice of the parametric copula family that is suitable for the observed data. The second problem is deciding whether a simple parametric specification of the calibration function is suitable for the data at hand. While the spline formulation is flexible enough, one should be aware that the increased complexity of the spline model comes at a computational price. The posterior distribution has support of significantly higher dimension than, say, a model with a constant calibration function.

Among the Bayesian model selection techniques developed for copula-based inference we mention the work of [47] who examined the performance of the deviance information criterion (DIC; [54]) for copula selection, [44] who discussed the performance of a number of model selection tools including Bayes factors [33], predictive functions and the CPO statistic

(see [30,29]) who chose the model with the largest posterior probability. However, some of these methods cannot be applied directly to the conditional copula model considered here. For instance, a meaningful comparison of the posterior probabilities obtained for different models relies on stable estimates of the Bayes factors which is often impossible when the corresponding posterior distributions have supports of different dimensions. This is the case when we compare the models with different specifications of the calibration functions. Our investigations have indeed shown that the estimates of Bayes factors in this case are too volatile to be practically useful. In the case of copula selection (when the copula parameter is constant), [29] use a common prior for the Kendall's τ across different families. For the models presented here, we could not adapt a similar approach as the prior specification of τ is done indirectly through the specification of priors for the parameters defining the cubic spline model. When taking into account the fact that the range of τ is different for different copula families, one realizes the difficulty to set up the spline model priors so that the models are a priori on equal footing.

We explore here the use of DIC for selecting the copula family and choosing between the general spline and a simple parametric form for η. The DIC proposed by [54] is defined as

$$\mathrm{DIC}(\mathcal{M}) = 2\overline{D(\omega)} - D(\bar{\omega}) \tag{6}$$

where the model deviance $D(\omega) = -2 \ln p(\mathcal{D}_{\mathrm{obs}}|\omega, \mathcal{M})$ is twice the negative observed-data log-likelihood, and the notation $\overline{A}$ represents the expectation of A with respect to the posterior distribution. Models with the lowest DIC value are preferred. The theoretical properties of DIC have been studied by [8,36,54]. All the required expectations in (6) are generally intractable analytically so they are computed via Monte Carlo integration.

4. Simulation study

We performed a large number of simulations to study: (1) the spline model's flexibility in capturing non-linear trends in the calibration function and (2) the power to select the correct copula model and to determine the form of the calibration function. Although the model can be implemented for continuous outcomes, we found that the more challenging setup, in terms of estimation and model selection, occurs when the outcomes are mixed, discrete and continuous. We therefore focus our simulation study solely on the mixed outcome scenario.

We have generated data under the Clayton copula family for two sample sizes, $n = 150, 450$. Three dependence patterns are created using the following calibration functions:

C1: $\eta(z) = \ln(3)$;
C2: $\eta(z) = \ln\{0.07z^6 - 0.37(z+1)(z-0.5) + 0.3\}$;
C3: $\eta(z) = \ln\{4.5 - 1.5\sin(\pi z)\}$.

The constant calibration **C1** is useful in studying the performance of the spline model under the classical copula setting, while **C2** and **C3** represent situations of weak and strong dependence, respectively. Since the focus of our analysis is on the estimation of the calibration function, we keep the marginal parameters fixed at $\beta_1 = \beta_2 = 5, \sigma_2 = \exp(0.25) \approx 1.3$ and throughout the simulation study $X \sim \mathrm{Unif}(-1, 1)$. Our simulations have shown that the goodness-of-fit improvement is small, but the computational load is growing fast for large values of $K_{\max}$. In all our simulations, we have kept $K_{\max} = 4$, although similar inference is obtained for all values in the set $\{4, \ldots, 8\}$. In general, one can use a Bayesian cross-validation procedure (see [57]) to find a suitable value of $K_{\max}$, but such an approach will significantly increase the computational load.

The data is analysed using three copula families: Clayton, Frank and Gumbel with the corresponding link functions $g_C(x) = \ln(1+x)$, $g_F(x) = x$ and $g_G(x) = \ln(x-1)$, respectively.

4.1. Performance of the MCMC algorithm

In all scenarios we have used 4 parallel chains to produce independent Monte Carlo samples that are used for the Gelman–Rubin convergence diagnostics (see [19] for details). More precisely, each chain is used to draw 70,000 samples from the posterior distribution and, after discarding the first 15,000 samples as burn-in, we retain every 20th sample for computing the parameter estimates. While we could have used all the samples, the computation time of each estimate and DIC value would have been much longer, but without much of a gain in efficiency, due to the high autocorrelation. The burn-in and run time have been decided based on the Gelman–Rubin diagnostics. We have also used a much longer run period of 200,000 samples for each chain in one scenario as a benchmark. The small differences observed (reported in the Appendix) suggest that there is little Monte Carlo efficiency to be gained from running the chains longer. Different users who would like to implement this approach to different data should run their own diagnostic tests and, based on them, decide on the number of Monte Carlo samples. The initialization values for each chain are sampled at random using the initial distributions: $\beta_i \sim \mathrm{Unif}(-2, 2)$, $i = 1, 2$, $\xi_2 \sim \mathrm{Unif}(0, 2)$, $\alpha_j \sim \mathrm{Unif}(0, 0.1)$ for $i = 0, 1, 2, 3$, $\lambda \sim \mathrm{Unif}\{1, 2, 3\}$ and for each $k \in \{1, \ldots, K_{\max}\}$, $\gamma_k \sim \mathrm{Unif}(I_k)$, $\zeta_k \sim \mathrm{Bernoulli}(0.5)$ and $\psi_k \sim \mathrm{Unif}(0, 0.1)$. It takes about 8 min and 25 min to generate 100,000 draws from the posterior on a Dual Xeon X5680 CPU (3.33 GHz) for $n = 150$ and $n = 450$, respectively.

The posterior mean spline function $\hat{\eta}(z)$ is the Monte Carlo average of all $\eta_i(z)$ that are obtained by plugging in the i-th sampled values of $\vec{\alpha}$, $\vec{\psi}$, $\vec{\gamma}$ and $\vec{\zeta}$ in (5). The algorithm's performance is quite robust to the choice of prior distribution for λ, as we have registered very similar performance and estimates when we changed the prior to uniform or a different binomial.

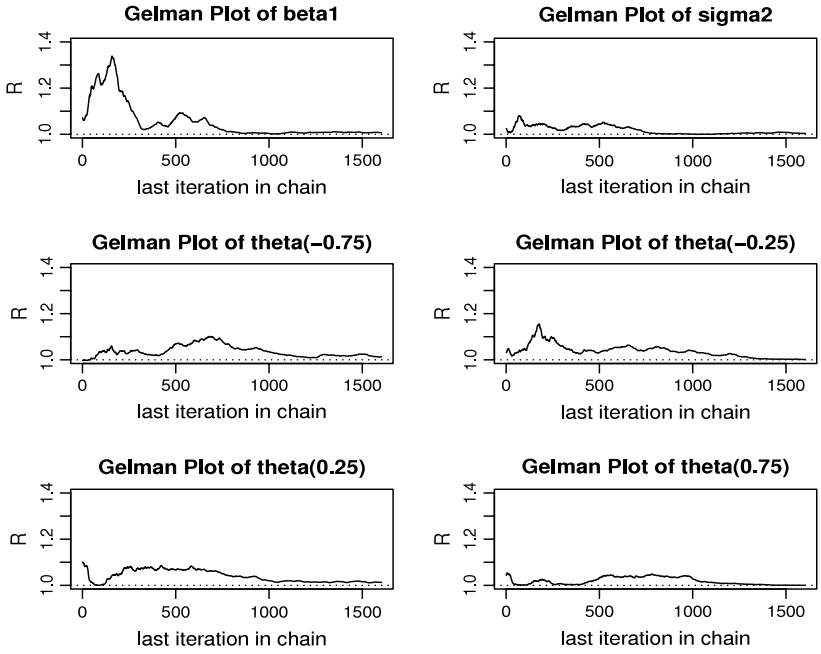

Fig. 1. Values of the Gelman–Rubin when data of size $n = 150$ is generated using a Clayton copula under scenario **C3**.

The influence of ζ and (γ, ψ) samples on the posterior estimations of η is complex. For instance, due to the form of (5), the knots γ_k situated in the leftmost intervals I_k have a diminished influence on the final estimate. During simulations, we have observed that switching a knot from say, I_1 to I_2 has little influence on the posterior mean of $\eta(x)$ but certainly affects the convergence diagnostic for the corresponding γ and ψ components. For this reason, we choose to produce Gelman–Rubin diagnostic plots for the θ function itself. We illustrate these at four values of the covariate X, $\{-0.75, -0.25, 0.25, 0.75\}$. Since the performance of the β_2 component of the Markov chain is superior to that of β_1, we have decided to present only the latter one as it represents the "worst case scenario" for convergence. Here we present some of the simulation diagnostics and plots under Scenario **C3**. The corresponding figures for Scenarios **C1** and **C2** are included in the Appendix. The simulation diagnostics do not change for the two sample sizes used so we report here the case $n = 150$. In Fig. 1 we present the Gelman–Rubin diagnostic plots for β_1, σ_2, $\theta(-0.75)$, $\theta(-0.25)$, $\theta(0.25)$, and $\theta(0.75)$. One can see that even immediately after burn-in, the $\mathbb{R}$ statistic is close to 1 which indicates that the chains mix well. The acceptance rates are between 10% and 50%. The lower acceptance rates, around 10%, are registered for components of γ situated in the leftmost intervals (lower values of the covariate) and the highest acceptance rates, around 50% are for the components of gamma situated in the rightmost intervals (largest values of the covariate values). The transition kernels for the remaining parameters and auxiliary variables exhibit acceptance rates between 20% and 35%.

Fig. 2 shows the autocorrelation plots for the same parameter values. The Appendix contains additional simulations indicating that weaker dependence between the responses (under scenario **C2**) leads to smaller within-chain correlations.

The trace plots obtained from samples produced by one of the parallel chains are shown in Fig. 3.

In practice, one can combine the samples obtained from all the parallel chains to produce estimates for the parameters of interest. In Fig. 4 we present the histogram of samples obtained for the parameters considered. Fig. 5 shows the Bayesian posterior mean of $\theta(z)$ (dashed line) against the true function (solid line) and the 95% pointwise credible bands (dotted lines).

4.2. Estimation

The plots presented show typical realizations of a single analysis. In order to get a more general view, we have analysed 100 data sets replicated under each scenario, **C1**, **C2** and **C3**. The results are summarized in Table 1 where we show the integrated squared bias (IBias^2) the integrated variance (IVar) and the integrated mean squared error (IMSE) incurred when estimating $\theta(z)$. The sample size affects significantly the efficiency of the estimator, a fact observed also in [3]. Not surprisingly, using the correct copula to fit the data results in more precise estimates of $\theta(z)$, thus making the copula choice an essential ingredient of the methodology proposed here. In the next section we illustrate the performance of the DIC criterion for copula selection.

Table 2 reports the squared bias, variance and mean squared errors observed for the posterior means of β_1, β_2, σ_2 in the case with $n = 150$ and $n = 450$. We can see from Table 2 that inference for the marginal parameters is less sensitive to copula selection. While the largest statistical efficiency is obtained under the correct copula, the ranking of the competing models

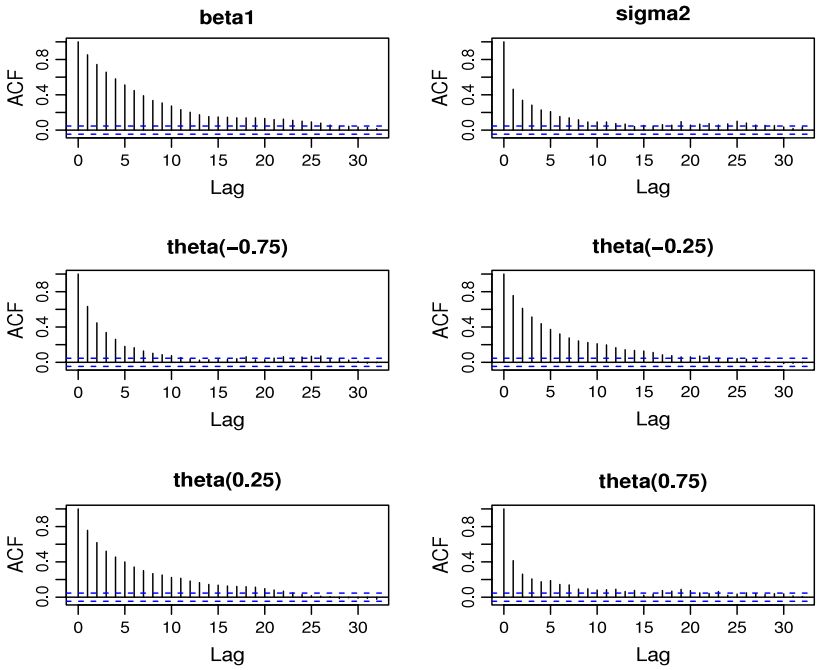

Fig. 2. ACF plots when data is generated using a Clayton copula under scenario **C3**.

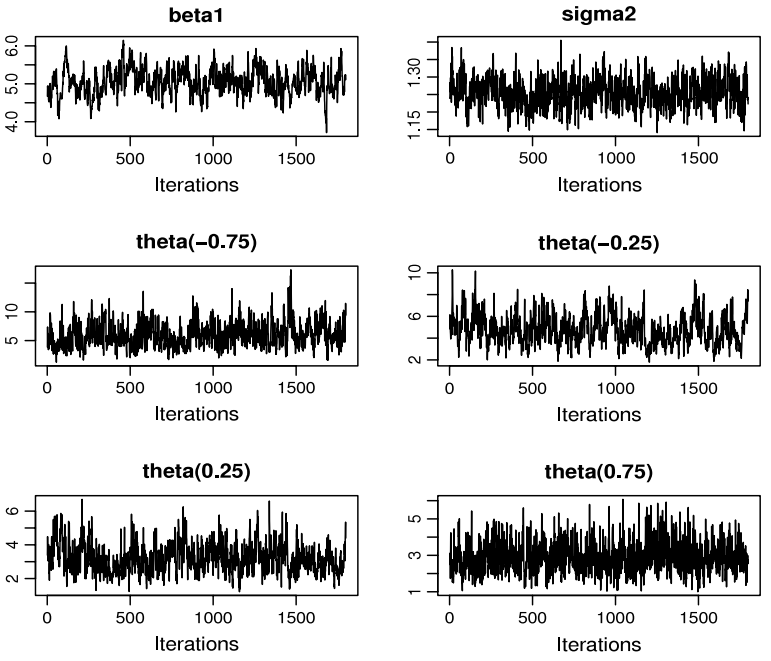

Fig. 3. Trace plots for samples obtained using one of the four parallel MCMC chains when data is generated using a Clayton copula under scenario **C3**.

depends on the calibration function. For instance, under **C2**, Clayton and Frank copulas produce similar performances, while under **C1** and **C3** Clayton and Gumbel lead to better efficiency than the Frank family. We also note the significantly larger MSE for β_1 than β_2, as the discrete response is less informative about the regression parameter.

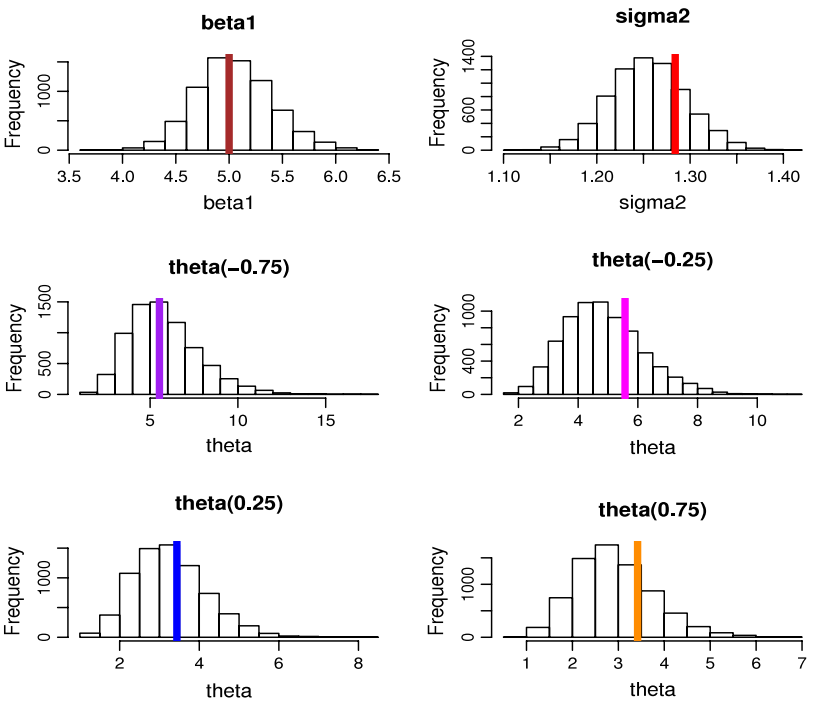

Fig. 4. Histograms of posterior samples obtained from all the parallel MCMC chains when data is generated using a Clayton copula under scenario **C3**. The solid line marks the true parameter value.

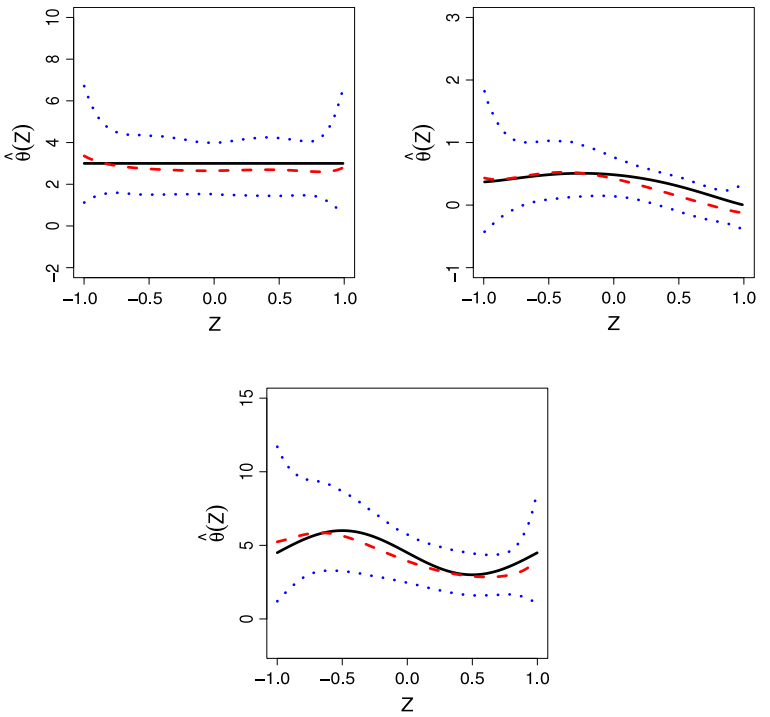

Fig. 5. Posterior of $\theta(z)$ (dashed line) against the true function (solid line) and the 95% pointwise credible bands (dotted lines) when data is generated using a Clayton copula under scenarios **C1** (top row, left panel), **C2** (top row, right panel) and **C3** (bottom row). The fit is based on samples obtained from all the parallel MCMC chains.

Table 1
Integrated squared bias, variance and mean squared error computed from 100 replicated data sets generated using the Clayton copula for the three scenarios **C1**, **C2** and **C3**. The figures shown in bold are obtained under the correct copula family.

Copula	$n = 150$			$n = 450$		
	IBias^2	IVar	IMSE	IBias^2	IVar	IMSE
	Scenario **C1**					
Clayton	**8.72**	**27.77**	**36.49**	**3.00**	**16.80**	**19.80**
Frank	395.45	62.46	457.91	352.04	59.77	411.81
Gumbel	2657.07	2172.18	4829.25	1100.38	315.37	1415.75
	Scenario **C2**					
Clayton	**1.79**	**12.72**	**14.51**	**0.03**	**0.27**	**0.30**
Frank	6.85	34.07	40.92	3.47	2.00	5.47
Gumbel	721.09	1550.19	2271.28	298.78	492.47	791.25
	Scenario **C3**					
Clayton	**14.96**	**65.38**	**80.34**	**8.43**	**48.66**	**57.09**
Frank	426.93	64.86	491.80	378.11	69.41	447.52
Gumbel	1862.32	1890.08	3752.4	1299.28	303.66	1602.94

Table 2
Squared bias, variance and mean squared error ($\times 100$) computed from 100 replicated data sets generated using the Clayton copula for the three scenarios **C1**, **C2** and **C3**.

Copula	n	Bias^2			Var			MSE		
		β_1	β_2	σ_2	β_1	β_2	σ_2	β_1	β_2	σ_2^2
		Scenario **C1**								
Clayton	150	**1.00**	**0.10**	**0.003**	**46.80**	**2.35**	**0.48**	**47.80**	**2.45**	**0.48**
	450	**0.2**	**0.01**	**0.001**	**17.6**	**1.23**	**0.17**	**17.6**	**1.24**	**0.16**
Frank	150	47.30	7.33	1.24	250.74	100.63	18.28	298.05	107.97	19.52
	450	9.1	1.74	0.27	107.96	45.85	9.26	117.06	47.59	9.53
Gumbel	150	9.50	0.59	0.02	97.88	5.73	1.08	107.39	6.33	1.09
	450	0.12	0.78	0.02	49.22	40.41	0.38	49.34	41.18	0.40
		Scenario **C2**								
Clayton	150	**0.56**	**0.04**	**0.09**	**42.00**	**2.85**	**0.72**	**42.56**	**2.89**	**0.81**
	450	**0.04**	**0.03**	**0.01**	**18.60**	**1.08**	**0.16**	**18.63**	**1.10**	**0.17**
Frank	150	0.78	0.03	0.04	42.10	3.02	0.73	42.87	3.04	0.77
	450	0.06	0.02	0.01	18.46	1.10	0.16	18.52	1.12	0.17
Gumbel	150	0.05	0.002	0.03	72.05	4.48	0.95	72.10	4.48	0.98
	450	0.07	0.05	0.01	29.83	1.37	0.22	29.91	1.42	0.23
		Scenario **C3**								
Clayton	150	**0.01**	**0.53**	**0.01**	**63.80**	**41.91**	**0.52**	**63.82**	**42.44**	**0.53**
	450	**0.07**	**0.001**	**0.001**	**11.79**	**1.27**	**0.18**	**11.86**	**1.27**	**0.18**
Frank	150	45.45	11.54	1.49	2.21	97.91	15.89	266.78	109.45	17.39
	450	5.29	1.37	0.11	73.49	32.88	3.19	78.78	34.25	3.30
Gumbel	150	11.33	0.37	0.29	114.81	10.51	1.49	126.14	10.88	1.78
	450	6.19	0.17	0.03	36.44	2.72	0.35	42.63	2.89	0.38

4.3. Copula selection and hypothesis testing

In this section we investigate the power of the DIC criterion to select among a number of copula families the one that fits best the data at hand. In order to assess the performance of DIC we have simulated 100 independent data sets of size $n = 150, 450$ under scenarios **C1**, **C2** and **C3**.

4.3.1. Copula selection

Each generated data set has been fitted using the Clayton, Frank and Gumbel copula families. In the left and right panels of Fig. 6 we plot the DIC values obtained when using the full model based on Frank and Clayton, and Gumbel and Clayton families, respectively, for $n = 450$. One can see that the DIC criterion does a good job at separating the three families and the Clayton family is chosen in over 90% of the cases over its two competitors. The similar plot for $n = 150$, shown in Fig. 7, indicates that with smaller sample sizes it is harder to select the correct copula. This is in agreement with [2] where it was also noticed a 5%–10% decrease in the probability of selecting the correct copula when the sample size is decreased from $n = 500$ to $n = 200$.

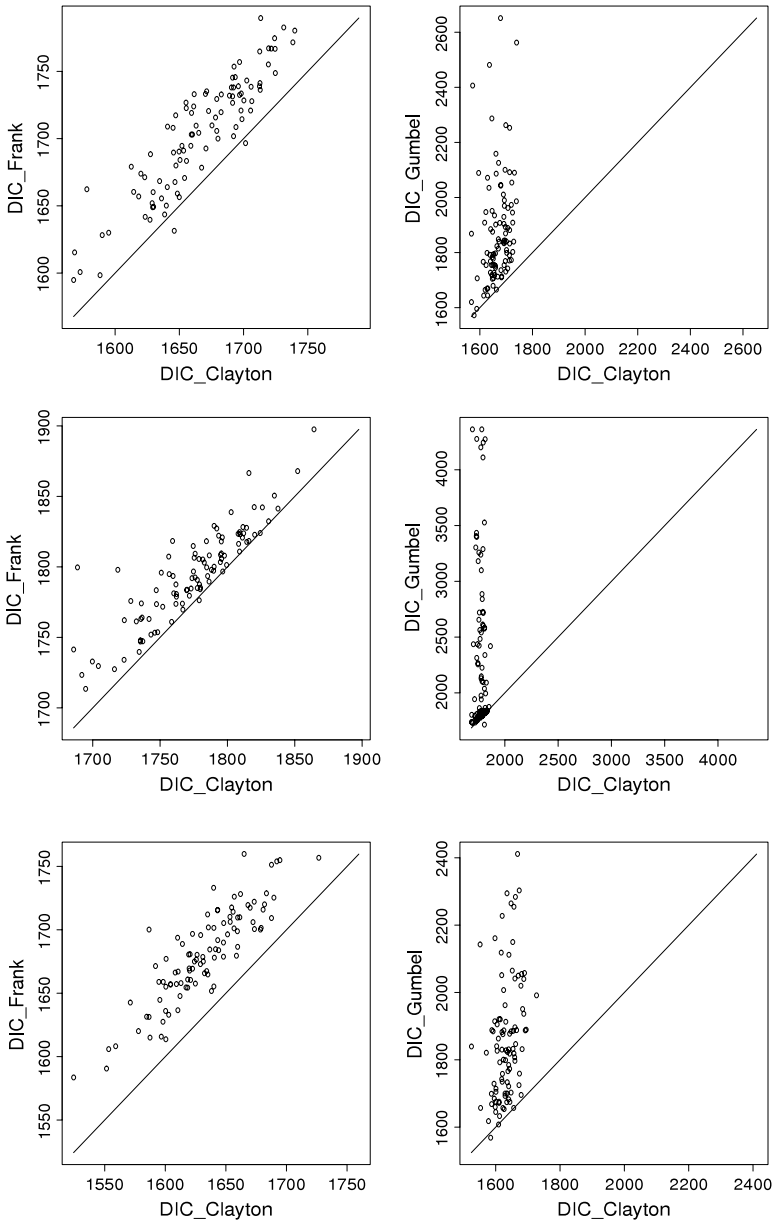

Fig. 6. DIC values for the three copula families considered under scenario **C1** (top row), **C2** (middle row) and **C3** (bottom row) for $n = 450$. Shown are plots of the DIC values for Frank family (left panel) and Gumbel family (right panel) against DIC values for Clayton family.

4.3.2. Hypothesis testing for $\theta(z)$

We have also evaluated the DIC's power to detect whether a parametric form for the calibration function is suitable. Data were simulated under scenarios **C1**, **C2** and **C3** and the DIC was computed for the full model using a spline specification of $\eta(z)$ and a constant or quadratic calibration function. Fig. 8 shows plots of the DIC computed under the two competing models for 100 replicated data points under each scenario. The DIC chooses the constant calibration model about 50% of the times under **C1**, as can be seen in the top panel of Fig. 8 and detects departure from the constant model in 75%–80% of replicated data sets under scenarios **C2** and **C3**. One may interpret the difficulty of choosing the constant calibration function over the spline as confirmation of the latter's good performance, since the constant is a degenerate spline with all coefficients equal to zero except for α_0. The same simulations performed under $n = 150$ sample size (corresponding plots are included in the Appendix) indicate that the power to separate the parametric and the spline calibration functions decreases when samples are smaller.

One may wonder whether the additional effort required to compute the full spline model is justified. In Table 3 we summarize 100 replicated data to illustrate the absolute relative bias induced when estimating $\Pr(Q = 1|W, X)$ under the

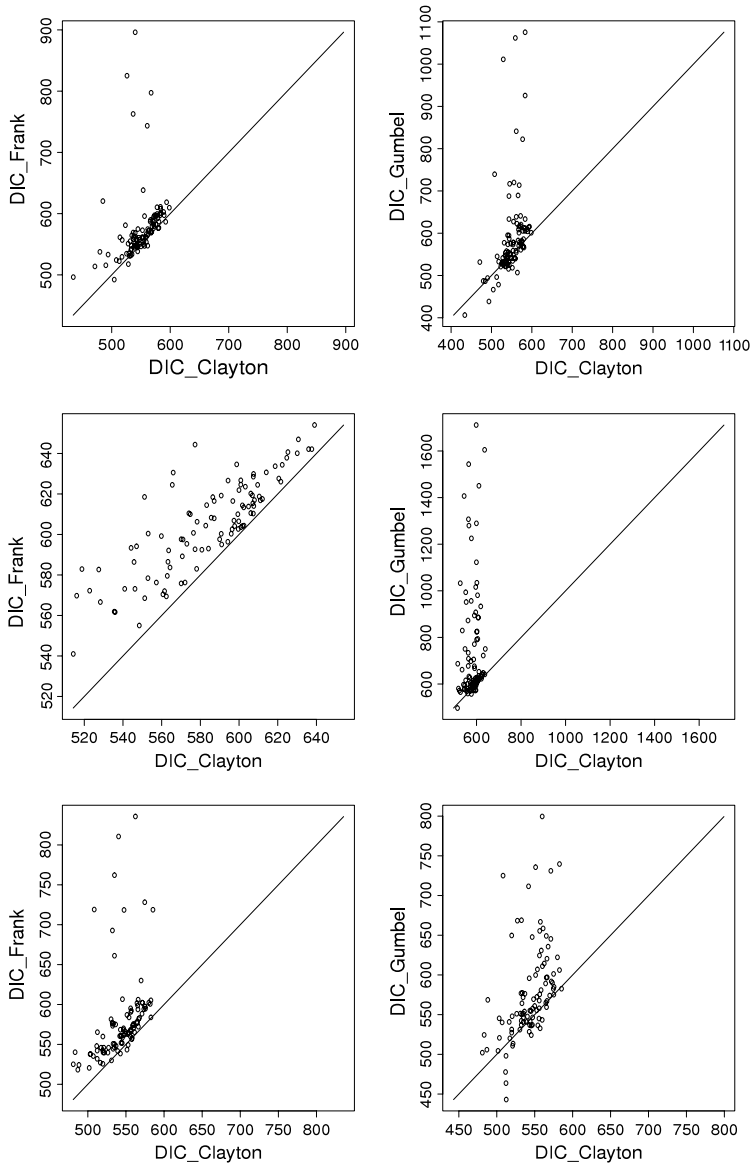

Fig. 7. DIC values for the three copula families considered under scenario **C1** (top row), **C2** (middle row) and **C3** (bottom row) for $n = 150$. Shown are plots of the DIC values for Frank family (left panel) and Gumbel family (right panel) against DIC values for Clayton family.

constant and spline copula models. One can see that, under scenario **C3**, for different values of X and $\beta = \beta_1 = \beta_2$, the relative bias of the predictions produced by the constant copula model can be significantly larger.

We have also tested DIC for its ability to select the spline model over a quadratic calibration function $\eta(z) = a_0 + a_1 z + a_2 z^2$. We have included in the Appendix the plots corresponding to those in Fig. 8. They suggest that DIC favours the spline model under scenario **C3**. This is not surprising as the calibration function under both **C1** and **C2** can be well approximated by a quadratic, albeit a degenerate one in the case of **C1**.

In general though, once the constant calibration function is deemed inappropriate for the data at hand, one would be hard put to decide on a particular polynomial degree to fit the data. The cubic spline model considered has the advantage of fitting a wide range of functions η in automatic fashion. With smaller sample sizes one has to find alternative ways to select the calibration function.

5. Conclusion and future work

We propose an approach for performing joint Bayesian inference for a conditional copula model for studying the dependence between continuous or mixed outcomes. The approach proposed in the paper can be extended to other marginal models (e.g. probit regression and linear regression with Student-distributed errors). However, as in any Bayesian analysis,

Table 3
Absolute relative bias estimates for $\theta(x)$ for the constant and spline copula models calculated under scenario **C3** from 100 independent data replicates. The values shown in bold are the true values of the parameters.

Model	$\theta(-0.75)$	$\theta(-0.25)$	$\theta(0.0)$	$\theta(0.25)$	$\theta(0.75)$
	$\beta_1 = \beta_2 = 5$				
True	**0.135**	**0.783**	**0.954**	**0**	**1**
Constant	0.452	0.170	0.0203	0	0
Spline	0.007	0.029	0.011	0	0
	$\beta_1 = \beta_2 = 2$				
True	**0.062**	**0.696**	**0**	**0.022**	**0.832**
Constant	0.483	0.010	0	3.400	0.130
Spline	0.069	0.005	0	0.222	0.013
	$\beta_1 = \beta_2 = 1$				
True	**0.137**	**0.028**	**0**	**0.066**	**0.184**
Constant	0.221	0.417	0	2.173	1.000
Spline	0.113	0.217	0	0.158	0.042

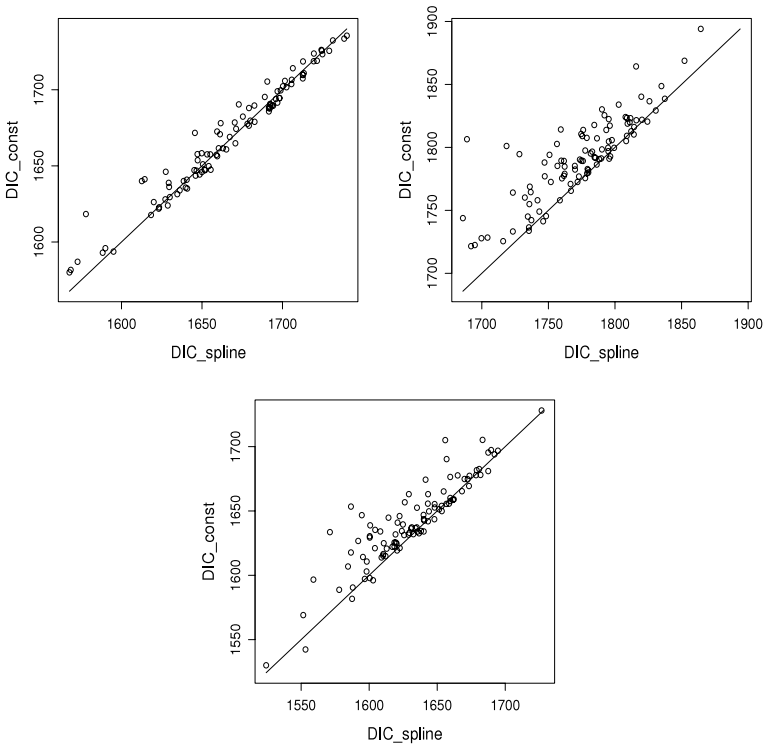

Fig. 8. Plot of DIC values for the model with a constant calibration function against values obtained for the model using a spline-specified calibration. The plots are obtained under scenario **C1** (top row, left panel), **C2** (top row, right panel) and **C3** (bottom row) with $n = 450$.

one must adapt to each new setting the computational algorithm required for posterior sampling. From this point of view our analysis is restricted to the logistic and Gaussian marginals. The functional dependence between the covariate and the copula parameter is modelled using a cubic spline model. In our simulations we have assumed that $\theta(x) \in \mathbb{R}$. We believe that the method can be applied to copulas with multivariate parameters, but due to space and time limitations we have not tackled such an extension. The DIC performs well when selecting the correct copula, but an alternative criterion is required to determine whether a parametric calibration function is suitable for the data at hand. We are currently working on extending this paper's framework to additive models [27] that would allow us to handle more than one covariate while keeping the computational burden within reasonable limits. We also investigate methods to expand the Bayesian approach to more than two response variables. A promising direction is offered by the vine copula representation that has been used successfully in constant copula models by [1,4,37,49,51]. However, for such complex structures, the model selection problem represents a serious obstacle which may require a different approach. We hope to report on these developments in a future communication.

Acknowledgements

We thank the Managing Guest Editor, Christian Genest, the Associate Editor and two anonymous referees for their careful, thorough and very helpful comments that have greatly improved the paper. We thank the organizers of the "Copula Models and Dependence" workshop in Montréal for the invitation to participate and to submit our work to this volume. We gratefully acknowledge the suggestions made by Elif Acar, Andrey Feuerverger, Johanna Nešlehová and Louis-Paul Rivest that have strengthened the paper. This research was supported by NSERC of Canada.

Appendix. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2012.03.010.

References

[1] K. Aas, C. Czado, A. Frigessi, H. Bakken, Pair-copula constructions of multiple dependence, Insurance: Mathematics & Economics 44 (2009) 182–198.
[2] E. Acar, Nonparametric estimation and inference for the copula parameter in conditional copulas, Ph.D. Thesis, University of Toronto, 2010.
[3] E. Acar, R.V. Craiu, F. Yao, Dependence calibration in conditional copulas: a nonparametric approach, Biometrics 67 (2011) 445–453.
[4] E. Acar, C. Genest, J. Nešlehová, Beyond simplified pair-copula constructions, Journal of Multivariate Analysis 110 (2012) 74–90.
[5] S. Bartram, S. Taylor, Y. Wang, The Euro and European financial market dependence, Journal of Banking and Finance 31 (2007) 1461–1481.
[6] Z. Baskurt, M. Evans, Inequalities for Bayes factors and relative belief ratios, Technical Report 1105, University of Toronto, 2011.
[7] S. Brooks, A. Gelman, G.L. Jones, X.L. Meng (Eds.), Handbook of Markov Chain Monte Carlo, Chapman & Hall, CRC, Boca Raton, FL, 2011.
[8] G. Celeux, F. Forbes, C.P. Robert, D.M. Titterington, Deviance information criteria for missing data models, Bayesian Analysis 1 (2006) 651–674.
[9] U. Cherubini, E. Luciano, W. Vecchiato, Copula Methods in Finance, John Wiley & Sons, Hoboken, NJ, 2004.
[10] D.G. Clayton, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, Biometrika 65 (1978) 141–151.
[11] R.V. Craiu, J.S. Rosenthal, C. Yang, Learn from thy neighbor: parallel-chain adaptive and regional MCMC, Journal of the American Statistical Association 104 (2009) 1454–1466.
[12] E. Cripps, C. Carter, R. Kohn, Variable selection and covariance selection in multivariate regression models, in: D.K. Dey, C.R. Rao (Eds.), Handbook of Statistics, vol. 25, Elsevier, North-Holland, Amsterdam, 2005.
[13] A.R. de Leon, B. Wu, Copula-based regression models for a bivariate mixed discrete and continuous outcome, Statistics in Medicine 30 (2011) 175–185.
[14] P. Embrechts, A. McNeil, D. Straumann, Correlation and dependence in risk management: properties and pitfalls, in: RISK Management: Value at Risk and Beyond, Cambridge University Press, 2002, pp. 176–223.
[15] M. Evans, G.H. Jang, Weak informativity and the information in one prior relative to another, Statistical Science 26 (2011) 423–439.
[16] Y. Fan, J.L. Dortet-Bernadet, S.A. Sisson, On Bayesian curve fitting via auxiliary variables, Journal of Computational and Graphical Statistics 19 (2010) 626–644.
[17] E.W. Frees, E.A. Valdez, Understanding relationships using copulas, North American Actuarial Journal 2 (1998) 1–25.
[18] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., in: Texts in Statistical Science Series, Chapman & Hall, CRC, Boca Raton, FL, 2004.
[19] A. Gelman, D.B. Rubin, Inference from iterative simulation using multiple sequences (with discussion), Statistical Science (1992) 457–511.
[20] C. Genest, I. Kojadinovic, J. Nešlehová, J. Yan, A goodness-of-fit test for bivariate extreme-value copulas, Bernoulli 17 (2011) 253–275.
[21] C. Genest, J. Nešlehová, A primer on copulas for count data, Astin Bulletin 37 (2007) 475–515.
[22] C. Genest, L.-P. Rivest, Statistical inference procedures for bivariate Archimedean copulas, Journal of the American Statistical Association 88 (1993) 1034–1043.
[23] I. Gijbels, N. Veraverbeke, M. Omelka, Conditional copulas, association measures and their applications, Computational Statistics & Data Analysis 55 (2011) 1919–1932.
[24] W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1996.
[25] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli 7 (2001) 223–242.
[26] H. Haario, E. Saksman, J. Tamminen, Componentwise adaptation for high dimensional MCMC, Computational Statistics 20 (2005) 265–273.
[27] T.J. Hastie, R.J. Tibshirani, Generalized Additive Models, Chapman & Hall, London, 1990.
[28] P. Hougaard, Analysis of Multivariate Survival Data, in: Statistics for Biology and Health, Springer-Verlag, New York, 2000.
[29] D. Huard, G. Évin, A.-C. Favre, Bayesian copula selection, Computational Statistics & Data Analysis 51 (2006) 809–822.
[30] J. Ibrahim, M.H. Chen, D. Sinha, Bayesian Survival Analysis, Springer-Verlag New York Inc., 2001.
[31] H. Joe, Multivariate Models and Dependence Concepts, in: Monographs on Statistics and Applied Probability, vol. 73, Chapman & Hall, London, 1997.
[32] E. Jondeau, M. Rockinger, The copula-GARCH model of conditional dependencies: an international stock market application, Journal of International Money and Finance 25 (2006) 827–853.
[33] R.E. Kass, A.E. Raftery, Bayes factors, Journal of the American Statistical Association 90 (1995) 773–795.
[34] P. Lambert, Archimedean copula estimation using Bayesian splines smoothing techniques, Computational Statistics & Data Analysis 51 (2007) 6307–6320.
[35] P. Lambert, F. Vandenhende, 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 (2002) 3197–3217.
[36] X.L. Meng, F. Vaida, What's missing for DIC with missing data? (comment on article by Celeux et al.), Bayesian Analysis 1 (2006) 687–698.
[37] A. Min, C. Czado, Bayesian inference for multivariate copulas using pair-copula constructions, Journal of Financial Econometrics 8 (2010) 511–546.
[38] R.B. Nelsen, An Introduction to Copulas, second ed., in: Springer Series in Statistics, Springer, New York, 2006.
[39] M. Omelka, I. Gijbels, N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing, Annals of Statistics 37 (2009) 3023–3058.
[40] A.J. Patton, Modelling asymmetric exchange rate dependence, International Economic Review 47 (2006) 527–556.
[41] A.J. Patton, Copula-based models for financial time series, in: T. Mikosch, J.P. Kreiß, R. Davis, T. Andersen (Eds.), Handbook of Financial Times Series, Springer, Berlin, Heidelberg, 2009.
[42] M. Pitt, D. Chan, R. Kohn, Efficient Bayesian inference for Gaussian copula regression models, Biometrika 93 (2006) 537–554.
[43] G. Roberts, J. Rosenthal, Coupling and ergodicity of adaptive MCMC, Journal of Applied Probability 5 (2007) 458–475.
[44] J.S. Romeo, N.I. Tanaka, A.C. Pedroso-de Lima, Bivariate survival modeling: a Bayesian approach based on copulas, Lifetime Data Analysis 12 (2006) 205–222.
[45] J.G. Scott, J.O. Berger, Bayes and empirical-Bayes multiplicity adjustment in the variable selection problem, Annals of Statistics 38 (2010) 2587–2619.
[46] J.H. Shih, T.A. Louis, Inferences on the association parameter in copula models for bivariate survival data, Biometrics 51 (1995) 1384–1399.
[47] R. Silva, H.F. Lopes, Copula, marginal distributions and model selection: a Bayesian note, Statistics and Computing 18 (2008) 313–320.
[48] A. Sklar, Fonctions de répartition à *n* dimensions et leurs marges, Publications de l'Institut de Statistique de l'Université de Paris 8 (1959) 229–231.

[49] M. Smith, M. Khaled, Estimation of copula models with discrete margins via Bayesian data augmentation, Journal of the American Statistical Association, in press (http://dx.doi.org/10.1080/01621459.2011.644501).
[50] M. Smith, R. Kohn, Nonparametric regression using Bayesian variable selection, Journal of Econometrics 75 (1996) 317–343.
[51] M. Smith, A. Min, C. Almeida, C. Czado, Modelling longitudinal data using a pair-copula decomposition of serial dependence, Journal of the American Statistical Association 105 (2010) 1467–1479.
[52] P.X.K. Song, Multivariate dispersion models generated from Gaussian copula, Scandinavian Journal of Statistics 27 (2000) 305–320.
[53] P.X.K. Song, M. Li, Y. Yuan, Joint regression analysis of correlated data using Gaussian copulas, Biometrics 65 (2009) 60–68.
[54] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, A. van der Linde, Bayesian measures of model complexity and fit (with discussion), Journal of the Royal Statistical Society, Series B 64 (2002) 583–639. (57).
[55] E. Spissu, A. Pinjari, R. Pendyala, C.R. Bhat, A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel, Transportation 36 (2009) 403–422.
[56] M.A. Tanner, Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, third ed., Springer, New York, 1996.
[57] A. Vehtari, J. Lampinen, Bayesian model assessment and comparison using cross-validated prediction densities, Neural Computation 14 (2002) 2349–2468.
[58] N. Veraverbeke, M. Omelka, I. Gijbels, Estimation of a conditional copula and association measures, Scandinavian Journal of Statistics 38 (2011) 766–780.
[59] W. Wang, M.T. Wells, Model selection and semiparametric inference for bivariate failure-time data, Journal of the American Statistical Association 95 (2000) 62–76.