Stat

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

Additive models for conditional copulas

Avideh Sabeti, Mian Wei and Radu V. Craiu*

Received 30 July 2014; Accepted 12 September 2014

Conditional copulas are flexible statistical tools that couple joint conditional and marginal conditional distributions. In a linear regression setting with more than one covariate and two dependent outcomes, we consider additive models for studying the dependence between covariates and the copula parameter. We examine the computation and model selection tools needed for Bayesian inference. The method is illustrated using simulations and a real example. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: additive models; Bayesian inference; conditional copulas; cross-validated marginal likelihood; cubic splines; Markov chain Monte Carlo

1 Introduction

Starting with the seminal paper of Sklar (1959), copulas have developed into an important tool used for qualitative and quantitative evaluations of dependence in statistical models. If $Y_1, Y_2, \ldots, Y_k$ are continuous random variables with joint distribution function H and marginal distributions $F_1, F_2, \ldots, F_k$, the unique copula $C : [0,1]^k \to [0,1]$ "couples" the joint and the marginal distributions via $H(y_1, \ldots, y_k) = C\{F_1(y_1), \ldots F_k(y_k)\}$, for all $(y_1, \ldots, y_k) \in \mathbb{R}^k$. Therefore, in order to define H, we need to know the marginals F_i and the copula C. This can be convenient in situations in which one has a good grasp on the marginal distributions.

As a natural extension, conditional copulas couple joint conditional and marginal conditional distributions (Lambert & Vandenhende, 2002; Patton, 2006). Specifically, if $X \in \mathbb{R}^p$ is a covariate vector, then

$$H_X(y_1, \ldots, y_k \mid X) = C\{F_{1|X}(y_1 \mid X), \ldots, F_{k|X}(y_k \mid X) \mid X\}, \quad \text{for all } (y_1, \ldots, y_k) \in \mathbb{R}^k. \tag{1}$$

Conditional copulas models play an essential part in modelling high-dimensional data. For instance, consider a random vector $Y = (Y_1, \ldots, Y_4) \in \mathbb{R}^4$. Using the decomposition used by Acar et al. (2012, equation (3), p. 75), we can show that its four-dimensional continuous density $f(y) := f(y_1, y_2, y_3, y_4)$ can be decomposed as

$$\begin{aligned} f(y_1, y_2, y_3, y_4) = {} & f_1(y_1) f_2(y_2) f_3(y_3) f_4(y_4) \\ & \times c_{12}\{F_1(y_1), F_2(y_2)\} c_{23}\{F_2(y_2), F_3(y_3)\} c_{14}\{F_1(y_1), F_4(y_4)\} \\ & \times c_{13|2}\{F_{1|2}(y_1|y_2), F_{3|2}(y_3|y_2)\} c_{24|1}\{F_{2|1}(y_2|y_1), F_{4|1}(y_4|y_1)\} \\ & \times c_{43|12}\{F_{4|12}(y_4 \mid y_1, y_2), F_{3|12}(y_3 \mid y_1, y_2)\}, \end{aligned} \tag{2}$$

where, if $\mathcal{A}, \mathcal{B} \subset \{1, 2, 3, 4\}$ are a set of indices, then we have used the following notations: $f_{\mathcal{A}}$ and $F_{\mathcal{A}}$ are, respectively, the joint density and distribution functions of $\{Y_j : j \in \mathcal{A}\}$; $f_{\mathcal{A}|\mathcal{B}}$ and $F_{\mathcal{A}|\mathcal{B}}$ are the conditional density and distribution

Department of Statistics, University of Toronto, Toronto, Ontario M5S3G3, Canada
***Email: craiu@utstat.toronto.edu**

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

functions of $\{Y_j : j \in \mathcal{A}\}$ given $\{Y_h : h \in \mathcal{B}\}$; $c_{\mathcal{A}}$ and $c_{\mathcal{A}|\mathcal{B}}$ denote, respectively, the copula density for $\{Y_j : j \in \mathcal{A}\}$ and the conditional copula density of $\{Y_j : j \in \mathcal{A}\}$ given $\{Y_h : h \in \mathcal{B}\}$. Not surprisingly, increasing the dimension of Y will result in a decomposition like (2) where copula densities are conditioned on more than two random variables. Acar et al. (2012) have shown that when replacing the conditional copulas with unconditional ones in (2), we are likely to incur inferential losses in terms of both bias and efficiency.

The conditional copula can also be a useful modelling tool in regression settings in which we observe outcomes $Y_1, \ldots, Y_k$ along with covariate vector $X \in \mathbb{R}^p$, and of interest is not only the effect of the covariate on each response but also the effect of X on the dependence structure between the responses. Throughout the paper, we consider parametric copula families in which the function C assumes a parametric form indexed by a copula parameter θ. In many applications one can reasonably assume that θ will vary with X. However, it is generally difficult to guess the functional relationship between θ and the covariate vector X, so its estimation requires flexible models that can capture a wide variety of patterns. This naturally leads to the use of semiparametric (Acar et al., 2011; Craiu & Sabeti, 2012) and non-parametric inferential tools (Omelka et al., 2009; Veraverbeke et al., 2011; Abegaz et al., 2012). However, as the dimension p of the covariate vector X increases, the volume of data required to keep the error within reasonable bounds increases very quickly (Abegaz et al., 2012). The generic motivational examples discussed earlier prompt our search for practical inferential procedures for conditional copula models when $p > 1$. The paper is developed for situations in which the parameter θ is a scalar and there are two (i.e. $k = 2$) continuous outcomes of interest, Y_1 and Y_2, that are marginally linked to the vector of covariates via linear regression models.

We propose here the use of additive models for studying the functional dependence between the covariate vector and the copula parameter. In this paper, we will improve on the statistical ingredients developed by Craiu & Sabeti (2012) in two directions. Most importantly, we will examine the performance of their Bayesian cubic spline estimator within an additive model framework. Secondly, we investigate the performance of the cross-validated marginal likelihood (CVML) criterion that adapts the seminal concept of cross-validation for marginal likelihood considered by Geisser & Eddy (1979) to the conditional copula setting.

In the next section, we introduce the statistical model and describe the computational algorithms needed for inference and the calculation of the CVML criterion. We will also introduce the additive model and the CVML criterion. Simulations and a real data analysis are discussed in Sections 3 and 4, respectively. The paper closes with a discussion of future research directions.

2 The model

In a regression setting, we consider the continuous bivariate outcome Y_1 and Y_2 along with covariate $X \in \mathbb{R}^p$. Marginally, each response Y_i, $i = 1, 2$, is modelled using a normal regression model. For a sample of size n, $\{(Y_{1j}, Y_{2j}, X_j) : 1 \le j \le n\}$, where $X_j = (X_{j1}, \ldots, X_{jp})^T$, we denote $\beta_i = (\beta_{i1}, \ldots, \beta_{ip})^T$ and assume that marginally

$$Y_{ij} \sim N(X_j^T \beta_i, \sigma_i^2), \ \forall 1 \le i \le 2, \ 1 \le j \le n, \tag{3}$$

leading to the conditional density of (Y_{1j}, Y_{2j}) given X_j

$$f(Y_{1j}, Y_{2j} \mid X_j) = \prod_{i=1}^{2} \frac{1}{\sigma_i} \phi\left(\frac{Y_{ij} - X_j^T \beta_i}{\sigma_i}\right) \times \tag{4}$$

$$\times\, c^{(1,1)}\left\{\Phi\left(\frac{Y_{1j} - X_j^T \beta_1}{\sigma_1}\right), \Phi\left(\frac{Y_{2j} - X_j^T \beta_2}{\sigma_2}\right) \,\middle|\, \theta(X_j)\right\}, \ \forall 1 \le j \le n, \tag{5}$$

 Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

The ISI's Journal for the Rapid Dissemination of Statistics Research

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

An important part of the model is the specification of $\theta(X)$. Many copula families have their parameter θ restricted to a subset of $\mathbb{R}$. We consider here a user-specified link function g that maps the support of the copula parameter onto the real line and set $g(\theta) = \eta(X)$, where $\eta : \mathbb{R}^p \to \mathbb{R}$ is the unknown *calibration function* we want to estimate. Given the one-to-one correspondence between the copula parameter $\theta(X)$ and the conditional Kendall's tau $\tau(X) = 4\mathrm{E}\{H(Y_1, Y_2 \mid X) \mid X\} - 1$, one can choose to parametrize the model on the τ or θ scale. In this paper, inference is performed directly on the copula parameter calibration function for computational convenience. However, when goodness-of-fit measures are reported across different copula families, one must use the τ scale, which is parametrization invariant (see also discussion in Acar et al., 2011).

We adopt an additive model (Hastie & Tibshirani, 1990) for $\eta(X)$

$$\eta(X) = \alpha_0 + \sum_{h=1}^{p} \eta_h(X_h), \tag{6}$$

where each $\eta_h : \mathbb{R} \to \mathbb{R}$ is specified using the flexible cubic spline model suggested by Smith & Kohn (1996) in which

$$\eta_h(X_h) = \sum_{j=1}^{3} \alpha_j^{(h)} X_h^j + \sum_{k=1}^{K^{(h)}} \psi_k^{(h)} (X_h - \gamma_k^{(h)})_+^3 \tag{7}$$

and $a_+ = \max(0,a)$. It is well known that the performance of spline-based estimators is influenced by the location of the knots $\gamma_k^{(h)}$. In our model, this choice is automatic and data-driven.

A general remark is that in our implementations, we assume the covariates are independent. In order to test this assumption when applying the method to real data, we have used tests based on the empirical copula process (Genest & Remillard, 2004; Kojadinovic & Holmes, 2009) and based on correlation of distances (Székely et al., 2007).

The priors assigned to the parameters involved in the marginal models are as follows:

$$\begin{aligned} \beta_i &\sim N(0, \sigma_i^2 \mathbf{I}_p), \ \forall i = 1,2, \\ \sigma_i^2 &\sim \mathrm{InvGam}(0.1, 0.1), \ \forall i = 1,2. \end{aligned}$$

For the parameters involved in the cubic spline , we follow the prior specifications used by Craiu & Sabeti (2012). For each covariate X_h, we select a fixed value for the maximum number of knots, $K_{\max}^{(h)}$. The range spanned by the observed values of covariate X_h is divided into $K_{\max}^{(h)}$ intervals of equal length, $I_1^{(h)}, \ldots, I_{K_{\max}^{(h)}}^{(h)}$, and we assume that each interval $I_k^{(h)}$ contains at most one knot. In order to complete the model specification, we introduce additional parameters $\{\zeta_k^{(h)} : 1 \leq k \leq K_{\max}^{(h)}\}$, where for all $k \in \{1, \ldots, K_{\max}^{(h)}\}$

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

The model (7) becomes then

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

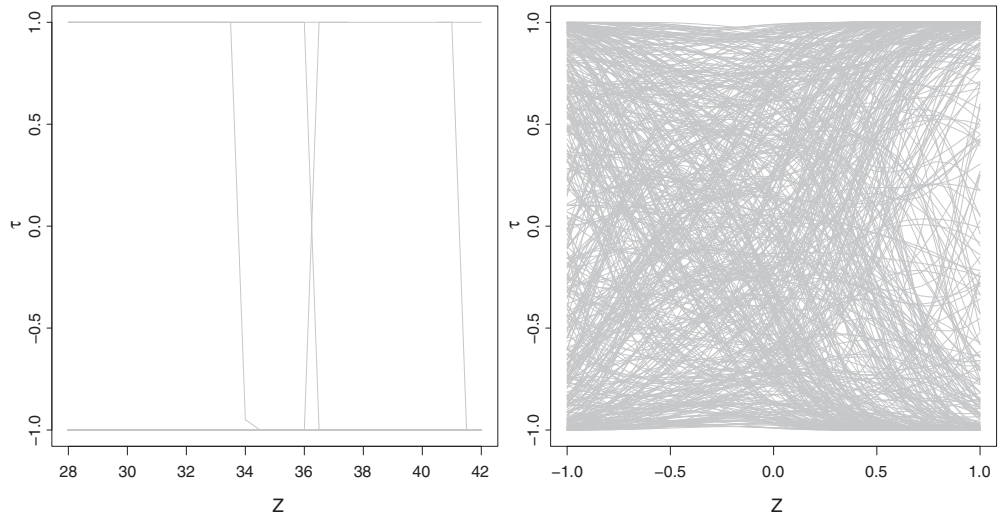


Figure 1. Five hundred realizations of $\eta_h(z)$ drawn from the prior distribution. The simulation set-up is inspired by the real data example from Section 5: in the left panel, Z is uniform on (28, 42), and in the right panel, Z has been standardized using the transformation $h(Z) = (Z - 35)/7$.

$$\eta_h(X_h) = \sum_{j=1}^{3} \alpha_j^{(h)} X_h^j + \sum_{k=1}^{K_{\max}^{(h)}} \zeta_k^{(h)} \psi_k^{(h)} (X_h - \gamma_k^{(h)})_+^3, \tag{8}$$

and one can see that the number of non-zero terms in the sum depends on the values of $\zeta_1^{(h)}, \ldots, \zeta_{K_{\max}^{(h)}}^{(h)}$. For each η_h, we construct a hierarchical prior for $\{\zeta_1^{(h)}, \ldots, \zeta_{K_{\max}^{(h)}}\}$. Specifically, if we let $|\zeta^{(h)}| = \sum_{k=1}^{K_{\max}^{(h)}} \zeta_k^{(h)}$ be the number of knots that are used in the model for η_i, then

$$p(|\zeta^{(h)}| \mid \lambda^{(h)}) \propto \frac{\lambda^{(h)|\zeta^{(h)}|}}{|\zeta^{(h)}|!} \mathbf{1}_{\{|\zeta^{(h)}| \le K_{\max}^{(h)}\}}, \tag{9}$$

that is, $|\zeta^{(h)}|$ follows the right truncated Poisson distribution with parameter $\lambda^{(h)}$ and maximum value $K_{\max}^{(h)}$. In addition,

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

The form of $p(\zeta^{(h)} \mid |\zeta^{(h)}|)$ 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_h are chosen regardless of the type of outcome as

$$\begin{aligned} \lambda^{(h)} &\sim \text{Bin}(K_{\max}^{(h)}, p = 0.5), \\ \alpha_0 &\sim \mathcal{N}(0, 10), \\ \alpha_j^{(h)} &\sim \mathcal{N}(0, 10), \ \forall 1 \le j \le 3, \end{aligned} \tag{10}$$

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

The ISI's Journal for the Rapid Dissemination of Statistics Research

$$\begin{aligned} \psi_k^{(h)} &\sim \mathcal{N}(0, 10), \ \forall 1 \leq k \leq K_{\max}^{(h)}, \\ \gamma_k^{(h)} &\sim \text{Unif}[I_k^{(h)}], \ \forall 1 \leq k \leq K_{\max}^{(h)}. \end{aligned} \tag{11}$$

Without additional information on the shape of η_h, we would like to be as vague as possible *a priori*. Note that the prior distributions given in equations (9) and (11) induce a prior distribution on the set of all possible maps $\eta_h : \mathbb{R} \to \mathbb{R}$. This prior is too complex to characterize analytically, but easy to sample from. Specifically, given a response index h, each sample of spline parameters $\{\zeta_k^{(h)}, \gamma_k^{(h)}, \psi_k^{(h)} : 1 \leq k \leq K_{\max}^{(h)}\}$, $\{\alpha_j^{(h)} : 1 \leq j \leq 3\}$ and α_0 from (9) and (11) will produce, when plugged into equation (8), a curve η_h. If the priors used are indeed not too informative about the shape of η_h, then we do not expect to see emerging any particular patterns. Our simulations show that the prior is not too sensitive to changes in the values used in (9) and (11) but is sensitive to the covariate's range. In Figure 1, we show 500 maps $\eta_h(z)$ on Kendall's tau scale where it is easier to see emerging patterns because the range is bounded. The left panel illustrates the case where the covariate is uniform on the interval $(28, 42)$ (the range was chosen to match the data example in Section 4), and the curves in the right panel are obtained after standardizing the covariate so that the values are in the interval $(-1, 1)$. One can see that when the covariate has a wider range, we tend to favour heavily, *a priori*, extreme dependence patterns with Kendall's tau close to 1 or -1. Such trends are undesirable as they can potentially bias the inference. However, after standardizing the covariate, the prior bias seems to vanish. For this reason, we recommend standardizing all covariates used in the conditional copula model.

2.1. The computational algorithm

If ω is the vector of all the parameters involved in the model and $\mathcal{D}$ denotes all the observed data, the posterior distribution, $\pi(\omega \mid \mathcal{D})$, cannot be studied analytically because of its complicated form. Instead, we construct a Markov chain Monte Carlo (MCMC) algorithm to sample from $\pi(\omega|\mathcal{D})$. The form of the sampling algorithm follows the generic design of the Gibbs sampler (Gelfand, 2000) in which every component ω_j is updated by sampling from its conditional distribution $\pi(\omega_j \mid \omega \backslash \omega_j, \mathcal{D})$. Some of the components of the chain cannot be sampled directly from the conditional distribution, so a Metropolis–Hastings update is needed (for details on using Metropolis–Hastings updates within the Gibbs sampler, see, for instance, Craiu & Rosenthal, 2014). The strategies used to update each parameter at step $t+1$ (the values at step t are marked using supra index $^{(t)}$) are described in the following.

β's: Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be the matrix whose rows are X_j^T, $1 \leq j \leq n$, and $\mathbf{Y}_1, \mathbf{Y}_2$ the vector of responses, $\mathbf{Y}_i = \{Y_{ij} : 1 \leq j \leq n\}$. If we had not considered the dependence between the outcomes, then the posterior conditional distribution of β_1, β_2 would have been

$$\tilde{\pi}_i\left(\beta_i|\mathcal{D}, \sigma_i^{(t)}\right) = \tilde{\pi}\left(\beta_i|\mathbf{X}, \mathbf{Y}_i, \sigma_i^{(t)}\right) = n(\beta_i; \mu_i, \Sigma_i), \ i = 1, 2, \tag{12}$$

where $n(x; a, b)$ is the density of a normal with mean vector a and variance matrix b, and

$$\begin{aligned} \mu_i &= \left(\mathbf{I} + \mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{X}^T\mathbf{Y}_i, \\ \Sigma_i &= \left(\sigma_i^{(t)}\right)^2 \left(\mathbf{I} + \mathbf{X}^T\mathbf{X}_i\right)^{-1}, \ i = 1, 2. \end{aligned} \tag{13}$$

The update of each β_i involves a mixture of transition kernels. With probability $\lambda = 0.8$, we update using an independent Metropolis (IM) transition kernel in which the proposal distribution is $\tilde{\pi}_i(\beta_i \mid \mathcal{D}, \sigma_i^{(t)})$, and with probability $1-\lambda = 0.2$, we update using a random-walk Metropolis (RWM) with a Gaussian proposal with mean at the current value of β_i and variance chosen so that the acceptance rate is between 20% and 30%.

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

σ's: Once again, without the copula component of the likelihood, the posterior conditional distribution of σ_i given the data and β_1, β_2 available in closed form is

$$\begin{aligned}\tilde{\pi}\left(\sigma_i|\mathcal{D}, \beta_i^{(t+1)}\right) &= \tilde{\pi}\left(\sigma|\mathbf{X}, \mathbf{Y}_i, \beta_i^{(t+1)}\right) = \\ &= \text{IG}\left(0.1 + \tfrac{p+n}{2}, 0.1 + \frac{\left(\beta_i^{(t+1)}\right)^T\beta_i^{(t+1)} + \left(\mathbf{Y}_i - \mathbf{X}\beta_i^{(t+1)}\right)^T\left(\mathbf{Y}_i - \mathbf{X}\beta_i^{(t+1)}\right)}{2}\right),\ i = 1, 2.\end{aligned} \tag{14}$$

The updates are made according to an IM kernel in which the proposal density is $\tilde{\pi}(\sigma \mid \mathbf{X}, \mathbf{Y}_i, \beta_i^{(t+1)})$ for each $i = 1, 2$. The updating steps for β and σ lead to faster mixing compared with the sampling algorithm defined in Craiu & Sabeti (2012) where only RWM updates were used, because the IM transition kernel has a good acceptance rate ($\approx$30%) and allows the chain to jump around the target space, which leads to a decrease in the autocorrelation.
In the absence of additional information regarding which covariates are more likely to induce changes in η, we use the same $K_{\text{max}} = K_{\text{max}}^{(h)}$ value for each $h = 1, 2, \ldots, p$.

α's: Because there is no range restriction for each $\alpha_k^{(h)}$ and no direct sampling strategy is possible, we use the RWM within Gibbs with proposal variance tuned so that the acceptance rates are between 20% and 50%.

ζ's: The updates are performed using the Metropolis-within-Gibbs strategy for the entire latent variable vector $\vec{\zeta}^{(h)} = (\zeta_1^{(h)}, \ldots, \zeta_{K_{\text{max}}}^{(h)})$. For updating $\vec{\zeta}^{(h)}$, 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 half to add/delete a component chosen at random and otherwise to permute two components of $\vec{\zeta}$ that are selected at random. Each proposed move is accepted or rejected based on a Metropolis–Hastings rule.

ψ's: If $\zeta_k^{(h)} = 1$, we use the RWM-within-Gibbs strategy to update $\psi_k^{(h)}$ using proposals tuned so that the acceptance rates are between 20% and 50%. If $\zeta_k^{(h)} = 0$, $\psi_k^{(h)}$ is updated using a random draw from its prior distribution that is automatically accepted.

γ's: If $\zeta_k^{(h)} = 1$, we use an IM update for $\gamma_k^{(h)}$ using as proposal the prior distribution of $\gamma_k^{(h)}$. If $\zeta_k^{(h)} = 0$, then the next state $\gamma_k^{(h)}$ is sampled from its prior and automatically accepted.

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

2.2. Pseudo-marginal likelihood model selection

The cross-validated pseudo-marginal likelihood (CVML) criterion of Geisser & Eddy (1979) is used to compare the predictive power of various models considered. Denote by $\mathcal{M}$ such a generic model, characterized by regression parameters $\{\beta_i, \sigma_i : i = 1, 2\}$ corresponding to a subset of covariates, $\mathbf{X}$, and all the spline parameters involved in modelling the calibration function $\eta(\mathbf{X})$. Denote the parameters in the model as ω, the data are $\mathcal{D}$, and for each $1 \le j \le n$, $\mathcal{D}_{-j}$ denotes the remaining data after we have removed the covariates and responses pertaining to the jth item, (Y_{1j}, Y_{2j}, X_j). A selection criterion based on the CVML will choose the model $\mathcal{M}$ that maximizes the sum

$$\text{CVML}(\mathcal{M}) = \sum_{j=1}^{n} \log p(Y_{1j}, Y_{2j}|\mathcal{D}_{-j}, \mathcal{M}). \tag{15}$$

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

The ISI's Journal for the Rapid Dissemination of Statistics Research

One can see from (15) that the CVML criterion favours models that exhibit the best average predictive power. The average is taken with respect to the parameters in the model, so (15) is a function of the model and the observed data only. From a Bayesian standpoint, the computation of the criterion would be impractical if we were to proceed by performing separately n data analyses, one for each sample of size $n-1$. However, the following derivation can be used to compute CVML($\mathcal{M}$) from a single Bayesian analysis of the *whole* data (see also Hanson et al., 2011). We have

$$\begin{aligned} E[\,p(Y_{1j}, Y_{2j}|\omega)^{-1}] &= \frac{1}{p(\mathcal{D}|\mathcal{M})} \int \frac{p(\mathcal{D}|\omega,\mathcal{M})p(\omega|\mathcal{M})}{p(Y_{1j},Y_{2j}|\omega,\mathcal{M})}\,\mathrm{d}\omega = \frac{1}{p(\mathcal{D}|\mathcal{M})} \int p(\mathcal{D}_{-j}|\omega,\mathcal{M})p(\omega|\mathcal{M})\,\mathrm{d}\omega \\ &= \frac{p(\mathcal{D}_{-j}|\mathcal{M})}{p(\mathcal{D}|\mathcal{M})} = \frac{1}{p(Y_{1j},Y_{2j}|\mathcal{D}_{-j},\mathcal{M})}, \end{aligned} \tag{16}$$

where the first expectation is taken with respect to the posterior distribution of all parameters in the model, $\pi(\omega \mid \mathcal{D},\mathcal{M}) = p(\mathcal{D}|\omega,\mathcal{M})/p(\mathcal{D} \mid \mathcal{M})$. Based on (16), we deduce that a Monte Carlo estimator of (15) is

$$\widehat{\mathrm{CVML}}(\mathcal{M}) = \sum_{j=1}^{n} -\log\left[\frac{1}{M}\sum_{m=1}^{M} p(Y_{1j}, Y_{2j}|\omega^{(m)},\mathcal{M})^{-1}\right], \tag{17}$$

where $\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(M)}$ are draws from the posterior distribution $\pi(\omega \mid \mathcal{D},\mathcal{M})$ obtained via the MCMC algorithm described in the previous section.

3 Simulations

The simulation study provides information about the average errors incurred when implementing the proposed estimation approach and illustrates the performance of the CVML criterion when it is used to select the copula family and the influential covariates in model (6).

3.1. Simulation details

We have generated data using the Clayton copula using either a univariate or bivariate calibration function. Marginally, the outcomes follow the distributions defined by the linear models specified in (3). All covariate values are independently sampled from the Uniform$[0,1]$ distribution. For the dependence structure, we have considered two non-linear calibration functions η_{S1} and η_{S2} defined as

$$\eta_{S1}(x) = \log[\,4.5 - 1.5\sin(\pi x)]$$

and

$$\eta_{S2}(x_1, x_2) = \log[\,4.5 - \sin(x_1) - \sin(x_2)]\,.$$

Under scenario **S1**, we simulate data using only one covariate so the true calibration function is η_{S1}, and under scenario **S2**, we generate data using the calibration η_{S2}. Marginally, under **S1** and **S2**, each response variable is linked to, respectively, one or two covariates via a linear model with Gaussian errors, as specified in (3).

Each analysis has been replicated 50 times for samples of size $n = 450$. We kept $K_{\max} = 4$ fixed throughout the simulation study. The MCMC sampler was run for 10,000 iterations, and the first 3,000 samples were discarded as burn-in. The simulation parameters used in the MCMC samplers were selected to produce acceptance probabilities in

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

the range 20–40%. The copula model data were generated using the `copula` library within R. The main steps of the MCMC sampler were implemented in C++ with the results processed in R.

3.2. Estimation of the calibration function

In this section, we present plots and measures of the goodness of fit for the proposed estimating procedure. We focus on scenario **S2**, which is more challenging to fit

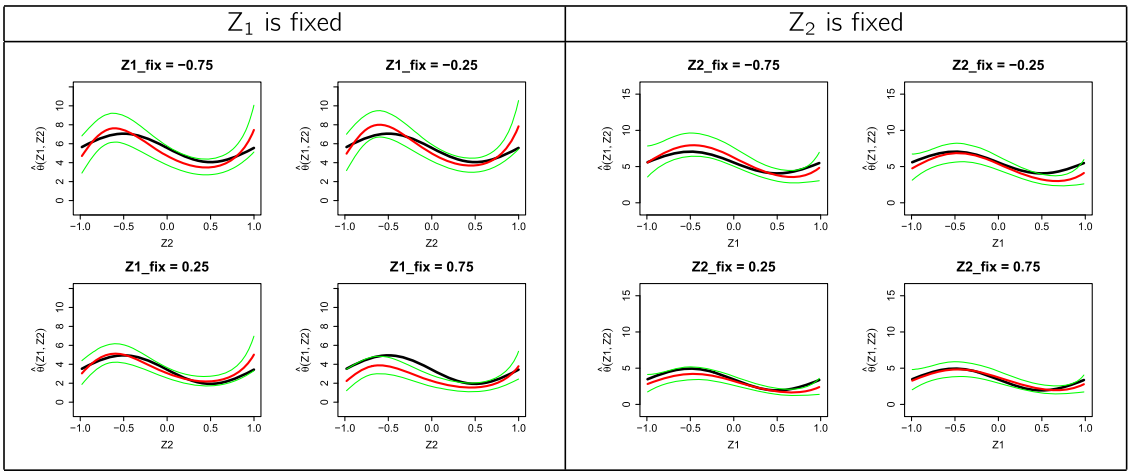
To provide a graphical illustration of the fit, in Table I, we show one-dimensional slices in the true surface (black line), the estimated surface (red line) and the two surfaces delimitating the 95% credible region (green lines). The slices are obtained when one of the two covariates is fixed at values in the set $\{-0.75, -0.25, 0.25, 0.75\}$. We observe that the credible bands grow wider near the boundaries of the covariate range and the bias gets also bigger when the covariate is closer to 1 or -1.

Table II contains the trace plots, the autocorrelation plots (up to lag 200) and the histograms of the posterior samples for $\theta(-0.25, 0.75)$, $\theta(0.75, -0.25)$ and $\theta(0.75, 0.75)$. In general, the autocorrelation function plots and the trace plots look similar across different parameter values. In the histograms, the red line shows the true value of the calibration function. We observe that the samples for $\theta(0.75, 0.75)$ are further from the true value when compared with the samples for $\theta(0.75, -0.25)$. This is consistent with our previous observation concerning the fit when covariate values are close to the boundary.

We also look at the model estimates for the normal regression parameters. Table III shows the trace plots, the autocorrelation plots and the histograms obtained from posterior samples corresponding to the linear regression model for the first outcomes β_{11} and β_{12}, and the residual standard deviation σ_1. The parameters used in the second regression model yield similar plots.

The red line in the histograms represents the true value of the parameter. Although the autocorrelation function seems to be high for these estimates, the histograms suggest that the samples provide good estimates of the marginal model parameters.

Table I. One-dimensional projections of the true calibration surface (black), the estimated surface (red) and credible bands (green) produced under scenario **S2**.



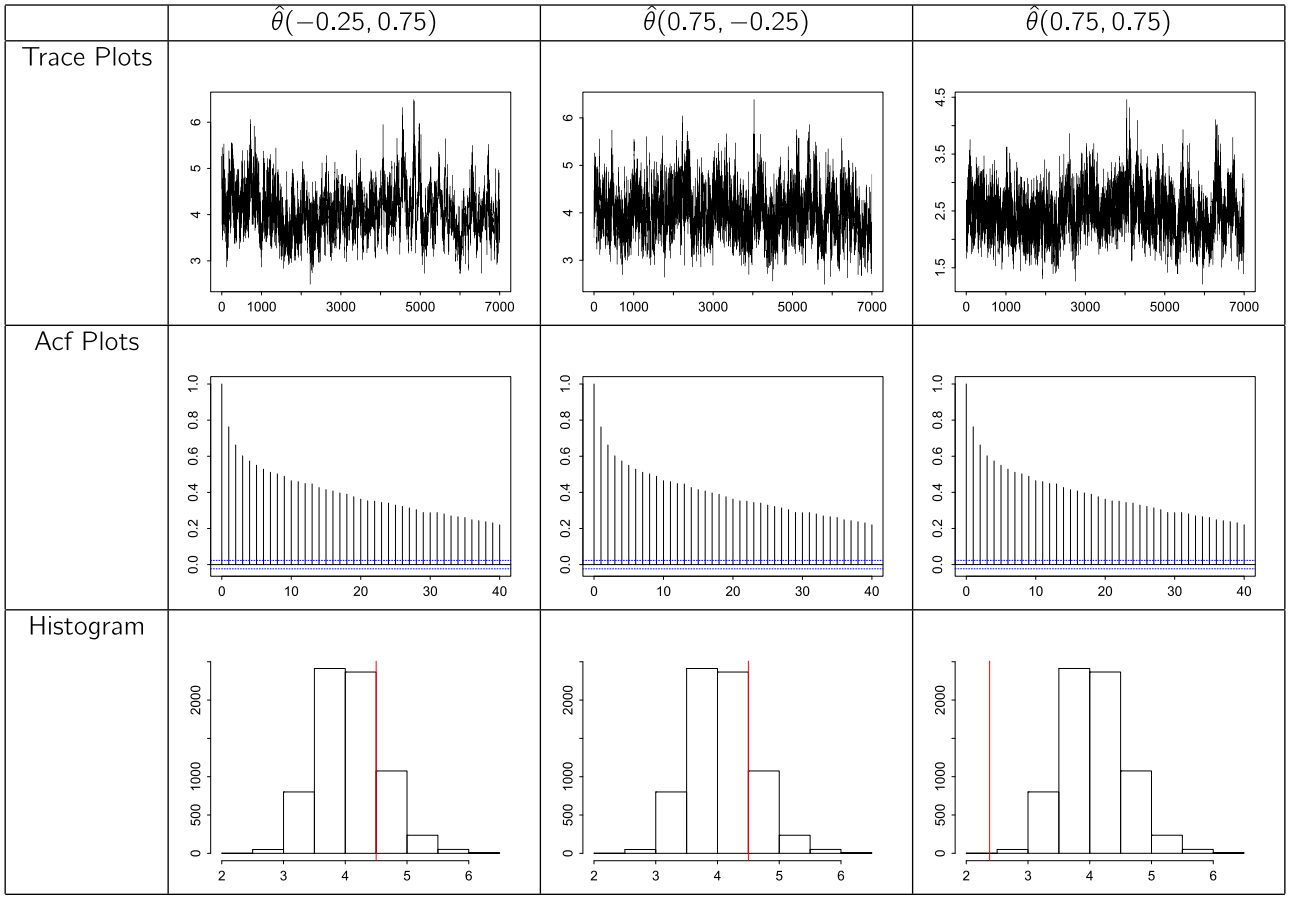
Copyright © 2014 John Wiley & Sons, Ltd

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table II. Trace (top row), autocorrelation function (middle row) and histogram (bottom row) plots for the calibration parameters $\theta(-0.25, 0.75)$, $\theta(0.75, -0.25)$ and $\theta(0.75, 0.75)$.



For a more global summary, we approximate numerically the integrated variance, squared bias and mean squared error using a grid of 400 equidistant points in the covariate space. The values reported in Table IV are computed when the copula family used is the true one. When comparing these measures across the two simulation scenarios, we notice an increase in bias when the number of covariates is increased. This is not surprising because we fit a significantly more complex model under scenario **S2** than under **S1**, but the sample size is the same.

3.3. Copula selection

The performance of the CVML criterion for choosing the correct copula family is illustrated in Table V. Specifically, we fit the generated data using Clayton, Frank and Gumbel copula families. In Table V, we report the percentage of correct decisions computed from 50 independent replicates. It can be noticed that there is a small decrease in accuracy for scenario **S2** compared with **S1**, which is not surprising given that the former model is more complex than the latter.

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

Table III. Trace (top row), autocorrelation function (middle row) and histogram (bottom row) plots for the regression coefficients, β_{11} and β_{12}, of the first outcome, Y_1, and the corresponding residual standard deviation, σ_1.

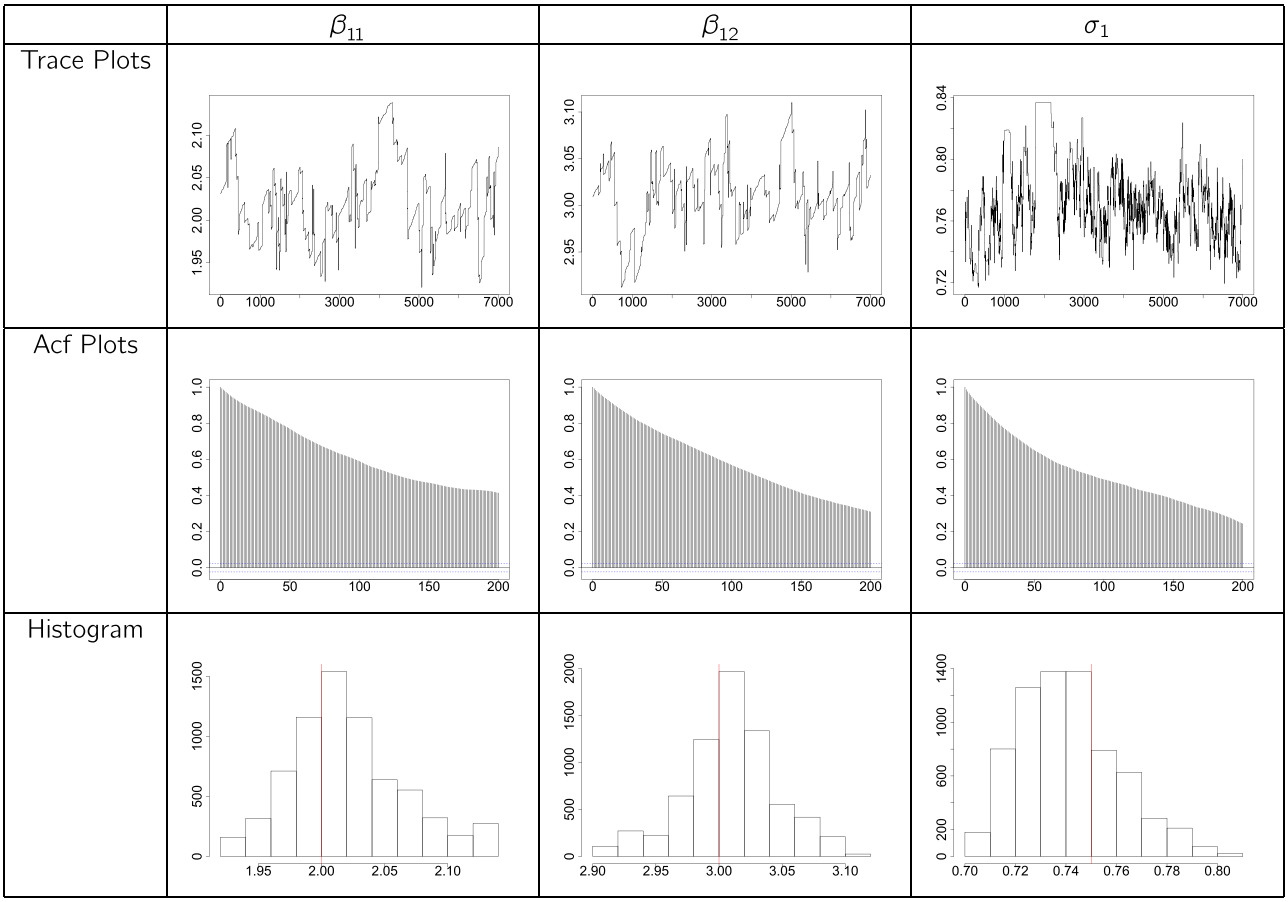


Table IV. Performance of the estimation procedure under scenarios **S1** and **S2**.

Scenario	Integrated squared bias	Integrated variance	Integrated mean squared error
S1	0.061	0.433	0.494
S2	0.132	0.515	0.647

3.4. Variable selection

We have also examined the performance of CVML in selecting the covariates to be included in the model. We focused on data generated under scenario **S2**, and we fitted them using models with one, two or three covariates. All results reported in this section are obtained under the correct Clayton copula family.

Copyright © 2014 John Wiley & Sons, Ltd

Stat

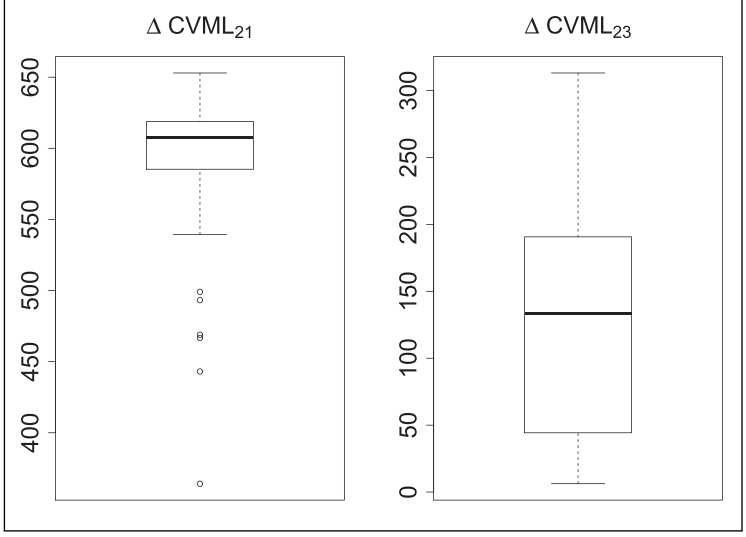
(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

The ISI's Journal for the Rapid Dissemination of Statistics Research

Table V. Performance of cross-validated marginal likelihood in selecting the correct Clayton family over Frank or Gumbel family under scenarios **S1** and **S2**. The numbers in the table represent the percentage of correct decisions.

Scenario/Copula	Frank (%)	Gumbel (%)
S1	100	98
S2	96	94

Table VI. Comparison of the cross-validated marginal likelihood (CVML) criterion values for models with one, two or three covariates. Left panel: box plot of 50 independently replicated values of the difference CVML($\mathcal{M}_2$) − CVML($\mathcal{M}_1$). Right panel: box plot of 50 independently replicated values of the difference CVML($\mathcal{M}_2$) − CVML($\mathcal{M}_3$).



If we denote by $\mathcal{M}_i$ the model with the first i covariates included, then we see from the box plots shown in Table VI that CVML always selects $\mathcal{M}_2$ over $\mathcal{M}_1$ or $\mathcal{M}_3$.The difference in CVML values is larger between $\mathcal{M}_2$ and $\mathcal{M}_1$ than between $\mathcal{M}_2$ and $\mathcal{M}_3$, which is natural given the criterion's connection to the model's predictive power.

4 Application to the twin birth data

The additive model approach is applied to a subset of the matched multiple birth data set. The data containing all twin births in the USA from 1995 to 2000 enable detailed investigation of twin gestations. We consider the twin live births in which both babies survived their first year of life with mothers aged between 18 and 40 years. Of interest

Copyright © 2014 John Wiley & Sons, Ltd

(wileyonlinelibrary.com) DOI: 10.1002/sta4.64

Table VII. Twin birth data: cross-validated marginal likelihood (CVML) values for three copula families under model $\mathcal{M}_1$ (bottom row) and $\mathcal{M}_2$ (top row). The criterion suggests that the model $\mathcal{M}_1$ with the Frank copula is most suitable for fitting the data.

Criterion	Clayton	Frank	Gumbel
CVML($\mathcal{M}_2$)	−10,213.4	−7683.7	−54,763.2
CVML($\mathcal{M}_1$)	−7405.3	−5569.4	−49,947.9

is the dependence between the birth weights of twins (in grammes), denoted by BW1 and BW2, respectively. We consider a random sample of 450 twin live births and investigate the effect of two covariates, gestational age (GA) and maternal age (MA), on the dependence between BW1 and BW2.

We compare the model $\mathcal{M}_1$ in which the GA is the only covariate considered and model $\mathcal{M}_2$ in which GA and MA are the included covariates. We also compare three analyses based on three parametric copula families: Clayton, Frank and Gumbel. For each copula family, we compute the CVML criterion for the models with both covariates (GA and MA) included. The results shown in the first row of Table VII suggest that the Frank copula is more suitable for analysing the data.

Under the Frank copula, model $\mathcal{M}_1$ is preferred with a CVML value of −5569.4 compared with −7683.7 obtained for $\mathcal{M}_2$. After deciding that M_1 is preferred, we compare again the fit for M_1 under each of the three copulas, and the results are shown on the second row of Table VII. This finding is concordant with the single covariate analysis of Acar et al. (2011).

5 Conclusions and future work

We propose Bayesian inference for the conditional copula model in a regression context with multiple covariates. We implement spline approximation within the additive model framework and propose a model selection criterion that selects the model with the best predictive power. The simulations show that the efficiency of the method decreases as the dimension of the covariate vector increases, and we would like to explore theoretically the rate of the decay. It is conceivable that when the number of covariates grows large, the approach proposed here may become too computationally expensive, and simpler formulations of the calibration function and improvements of the MCMC algorithm needed to sample the posterior distribution are worth investigating.

Acknowledgements

This work was supported by an individual Natural Sciences and Engineering Research Council of Canada research grant.

References

Abegaz, F, Gijbels, I & Veraverbeke, N (2012), 'Semiparametric estimation of conditional copulas', *Journal of Multivariate Analysis*, **110**, 43–73.

 Copyright © 2014 John Wiley & Sons, Ltd

Acar, E, Craiu, RV & Yao, F (2011), 'Dependence calibration in conditional copulas: a nonparametric approach ', *Biometrics*, **67**, 445–453.

Acar, E, Genest, C & Nešlehová, J (2012), 'Beyond simplified pair-copula constructions', *Journal of Multivariate Analysis*, **110**, 74–90.

Craiu, RV & Rosenthal, JS (2014), 'Bayesian computation via Markov chain Monte Carlo', *Annual Reviews of Statistics and Its Application*, **1**, 179–201.

Craiu, RV & Sabeti, A (2012), 'In mixed company: Bayesian inference for bivariate conditional copula models with discrete and continuous outcomes', *Journal of Multivariate Analysis*, **110**, 106–120.

Geisser, S & Eddy, WF (1979), 'A predictive approach to model selection', *Journal of the American Statistical Association*, **74**, 153–160.

Gelfand, AE (2000), 'Gibbs sampling', *Journal of the American Statistical Association*, **95**(452), 1300–1304.

Genest, C & Remillard, B (2004), 'Tests of independence and randomness based on the empirical copula process', *Test*, **13**, 335–369.

Hanson, T, Branscum, A & Johnson, W (2011), 'Predictive comparison of joint longitudinal–survival modelling: a case study illustrating competing approaches', *Lifetime Data Analysis*, **17**, 2–28.

Hastie, TJ & Tibshirani, RJ (1990), *Generalized Additive Models*, *Chapman & Hall*, London.

Kojadinovic, I & Holmes, M (2009), 'Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process', *Journal of Multivariate Analysis*, **100**, 1137–1154.

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

Omelka, M, Gijbels, I & Veraverbeke, N (2009), 'Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing', *Annals of Statistics*, **37**(5B), 3023–3058.

Patton, AJ (2006), 'Modelling asymmetric exchange rate dependence', *International Economic Review*, **47**, 527–556.

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

Smith, M & Kohn, R (1996), 'Nonparametric regression using Bayesian variable selection', *Journal of Econometrics*, **75**, 317–343.

Székely, GJ, Rizzo, ML & Bakirov, NK (2007), 'Measuring and testing dependence by correlation of distances', *Annals of Statistics*, **35**, 2769–2794.

Veraverbeke, N, Omelka, M & Gijbels, I (2011), 'Estimation of a conditional copula and association measures', *Scandinavian Journal of Statistics*, **38**, 766–780.

Copyright © 2014 John Wiley & Sons, Ltd