Citation: Marra, G. and Radice, R. ORCID: 0000-0002-6316-3961 (2017). Bivariate copula additive models for location, scale and shape. Computational Statistics & Data Analysis, 112, pp. 99-113. doi: 10.1016/j.csda.2017.03.004

Bivariate Copula Additive Models for Location, Scale and Shape

Giampiero Marra* Rosalba Radice†

Wednesday 15$^{\text{th}}$ February, 2017

Abstract

In generalized additive models for location, scale and shape (GAMLSS), the response distribution is not restricted to belong to the exponential family and all the model's parameters can be made dependent on additive predictors that allow for several types of covariate effects (such as linear, non-linear, random and spatial effects). In many empirical situations, however, modeling simultaneously two or more responses conditional on some covariates can be of considerable relevance. The scope of GAMLSS is extended by introducing bivariate copula models with continuous margins for the GAMLSS class. The proposed computational tool permits the copula dependence and marginal distribution parameters to be estimated simultaneously, and each parameter to be modeled using an additive predictor. Simultaneous parameter estimation is achieved within a penalized likelihood framework using a trust region algorithm with integrated automatic multiple smoothing parameter selection. The introduced approach allows for straightforward inclusion of potentially any parametric marginal distribution and copula function. The models can be easily used via the `copulaReg()` function in the `R` package `SemiParBIVProbit`. The proposal is illustrated through two case studies and simulated data.

Key Words: additive predictor, marginal distribution, copula, simultaneous parameter estimation.

*Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK, `giampiero.marra@ucl.ac.uk`.

†Department of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK, `r.radice@bbk.ac.uk`.

1 Introduction

Regression models typically involve a response variable and a set of covariates. However, modeling simultaneously two or more responses conditional on some covariates can be of considerable empirical relevance. Some examples can be drawn from health economics (e.g., modeling self-selection and dependence between health insurance and health care demand among married couples), engineering and econometrics (e.g., building time-series models for electricity price and demand), biostatistics (e.g., modeling adverse birth outcomes), actuarial science (e.g., studying the interdependence between mortality and losses) and finance (e.g., modeling jointly the prices of different assets); see Trivedi & Zimmer (2006) for more examples. The copula approach offers a convenient and computationally tractable framework to model multivariate responses in a regression context and it has been the subject of many methodological developments over the last few years (e.g., Cherubini et al., 2004; Kolev & Paiva, 2009; Nelsen, 2006; Radice et al., 2016, and references therein).

Rigby & Stasinopoulos (2005) extended the class of generalized additive models (GAM; Hastie & Tibshirani, 1990; Wood, 2006) by introducing generalized additive models for location, scale and shape (GAMLSS). Here, the response distribution is not restricted to belong to the exponential family and its parameters can be made dependent on flexible functions of explanatory variables. A similar idea was followed by Yee & Wild (1996) and has been recently exploited by Klein et al. (2015b). This article extends the scope of GAMLSS by introducing a computational tool for fitting bivariate copula models with continuous margins for the GAMLSS class. The method permits the copula dependence and marginal distribution parameters to be estimated simultaneously, and each parameter to depend on an additive predictor incorporating several types of covariate effects (such as linear, non-linear, random and spatial effects). The framework allows for the use of potentially any parametric marginal response distribution, several dependence structures between the margins as implied by parametric copulae, and as many additive predictors as the number of distributional parameters. The proposed approach is a direct competitor of the technique by Vatter & Chavez-Demoulin (2015). The main difference is that these authors' method is based on a two-stage technique where the parameters of the marginal distributions and of the copula function are estimated separately, whereas our method is based on the simultaneous estimation of all the model's parameters. The proposal can also be regarded as an extension of the works by Radice et al. (2016) and Yee (2016), and as a frequentist counterpart of the Bayesian implementation by Klein & Kneib (2016). Other existing bivariate copula regression approaches and software implementations (e.g., Acar et al., 2013; Gijbels et al., 2011; Kramer et al., 2012; Kraemer & Silvestrini, 2015; Sabeti et al., 2014; Yan, 2007) cover only parts of the flexibility of the proposed modeling tool.

The methodology developed in this article is most useful when the main interest is in relating the parameters of a bivariate copula distribution to covariate effects. Otherwise, semi/non-parametric extensions where, for instance, the margins and/or copula function are estimated using kernels, wavelets or orthogonal polynomials may be considered instead (e.g., Kauermann et al., 2013; Lambert, 2007; Segers et al., 2014; Shen et al., 2008). Such techniques are in principle more flexible in determining the shape of an underlying bivariate distribution. In practice, however, they are very limited with regard to the inclusion of flexible covariate effects and may require the imposition of functional identifying restrictions.

The remainder of the paper is organized as follows. Section 2 introduces the proposed class of models, discusses parameter estimation and inference, and provides some guidelines for model building. Sections 3 and 4 illustrate the approach on simulated and real data, whereas Section 5 discusses some potential directions for future research. The models discussed in this paper can be easily used via the `copulaReg()` function in the `R` package `SemiParBIVProbit` (Marra & Radice, 2017), and the reader can reproduce the analyses in this paper using the `R` scripts in the on-line Supplementary Material.

2 Methodology

This section introduces bivariate copula additive models and describes its main building blocks. Parameter estimation, inference and model building are also discussed.

2.1 Copula models for the GAMLSS class

Let us express the joint cumulative distribution function (cdf) of two continuous random variables, Y_1 and Y_2, conditional on a generic vector of covariates $\mathbf{z}$ as

$$F(y_1, y_2|\vartheta) = C\left(F_1(y_1|\mu_1, \sigma_1, \nu_1), F_2(y_2|\mu_2, \sigma_2, \nu_2); \zeta, \theta\right), \quad (1)$$

where $\vartheta = (\mu_1, \sigma_1, \nu_1, \mu_2, \sigma_2, \nu_2, \zeta, \theta)^{\mathsf{T}}$, $F_1(y_1|\mu_1, \sigma_1, \nu_1)$ and $F_2(y_2|\mu_2, \sigma_2, \nu_2)$ are marginal cdfs of Y_1 and Y_2 taking values in $(0, 1)$, μ_m, σ_m and ν_m, for $m = 1, 2$ are marginal distribution parameters, C is a uniquely defined two-place copula function with dependence coefficient θ (e.g., Sklar, 1959, 1973), ζ represents in this case the number of degrees of freedom of the Student-t copula (which only appears in C and ϑ when such copula is employed), and the parameters in ϑ are linked to **z** via additive predictors (see next section). Note that equation (1) also allows for copulae with two association parameters in which case ζ would represent an additional dependence coefficient. A substantial advantage of the copula approach is that a joint cdf can be conveniently expressed in terms of (arbitrary) univariate marginal cdfs and a function C that binds them together. The copulae implemented in `SemiParBIVProbit` are reported in Table 1. Counter-clockwise rotated versions of the Clayton, Gumbel and Joe copulae are obtained using the formulae in Brechmann & Schepsmeier (2013). Furthermore, each of the Clayton, Gumbel and Joe copulae can be mixed with a rotated version of the same family. For instance, mixing the Clayton copula with its 90 degree (counter-clockwise) rotation allows one to model positive and negative tail dependence. More details on copulae can be found in Nelsen (2006), for instance.

The marginal distributions of Y_1 and Y_2 are specified using parametric cdfs and densities denoted respectively as $F_m(y_m|\mu_m, \sigma_m, \nu_m)$ and $f_m(y_m|\mu_m, \sigma_m, \nu_m)$, for $m = 1, 2$. In this work, we have considered two and three parameter distributions, hence the notation adopted. However, the computational framework can be conceptually easily extended to parametric distributions with more than three parameters. The distributions implemented in `SemiParBIVProbit` are described in Table 2 and have been parametrized according to Rigby & Stasinopoulos (2005); sometimes μ_m, σ_m and ν_m represent location, scale and shape.

As it should be clear from our model specification, this paper is concerned with conditional copula models. We would like to stress that the theory for such models is relatively new. Patton (2002) first discussed conditional copulae by assuming that the conditioning covariate vector is the same for the margins and copula, whereas Fermanian & Wegkamp (2012) relaxed this assumption and developed the concept of pseudo-copula. Vatter & Chavez-Demoulin (2015) pointed out that a regression-like theory for the dependence parameter is possible when using conditional copulae and assuming exogeneity of the covariates.

In the above notation, observation index i was suppressed to avoid clutter. In fact, our focus

Copula	$C(u,v;\zeta,\theta)$	Ranges of θ and ζ	Transformation	Kendall's τ
AMH (`"AMH"`)	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\theta \in [-1,1]$	$\tanh^{-1}(\theta)$	$-\frac{2}{3\theta^2}\left\{\theta+(1-\theta)^2 \log(1-\theta)\right\}+1$
Clayton (`"C0"`)	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$\theta \in (0,\infty)$	$\log(\theta-\epsilon)$	$\frac{\theta}{\theta+2}$
FGM (`"FGM"`)	$uv\left\{1+\theta(1-u)(1-v)\right\}$	$\theta \in [-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{9}\theta$
Frank (`"F"`)	$-\theta^{-1}\log\left\{1+(e^{-\theta u}-1)(e^{-\theta v}-1)/(e^{-\theta}-1)\right\}$	$\theta \in \mathbb{R}\backslash\{0\}$	$-$	$1-\frac{4}{\theta}\left[1-D_1(\theta)\right]$
Gaussian (`"N"`)	$\Phi_2\left(\Phi^{-1}(u),\Phi^{-1}(v);\theta\right)$	$\theta \in [-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$
Gumbel (`"G0"`)	$\exp\left[-\left\{(-\log u)^\theta+(-\log v)^\theta\right\}^{1/\theta}\right]$	$\theta \in [1,\infty)$	$\log(\theta-1)$	$1-\frac{1}{\theta}$
Joe (`"J0"`)	$1-\left\{(1-u)^\theta+(1-v)^\theta-(1-u)^\theta(1-v)^\theta\right\}^{1/\theta}$	$\theta \in (1,\infty)$	$\log(\theta-1-\epsilon)$	$1+\frac{4}{\theta^2}D_2(\theta)$
Student-t (`"T"`)	$t_{2,\zeta}\left(t_\zeta^{-1}(u),t_\zeta^{-1}(v);\zeta,\theta\right)$	$\theta \in [-1,1]$ $\zeta \in (2,\infty)$	$\tanh^{-1}(\theta)$ $\log(\zeta-2-\epsilon)$	$\frac{2}{\pi}\arcsin(\theta)$

Table 1: Definition of copulae implemented in `SemiParBIVProbit`, with corresponding parameter ranges of association parameter θ and number of degrees of freedom ζ (when present), transformation/link function of θ and ζ, and relation between Kendall's τ and θ. $\Phi_2(\cdot,\cdot;\theta)$ denotes the cdf of a standard bivariate normal distribution with correlation coefficient θ, and $\Phi(\cdot)$ the cdf of a univariate standard normal distribution. $t_{2,\zeta}(\cdot,\cdot;\zeta,\theta)$ indicates the cdf of a standard bivariate Student-t distribution with correlation θ and ζ degrees of freedom, and $t_\zeta(\cdot)$ denotes the cdf of a univariate Student-t distribution with ζ degrees of freedom. $D_1(\theta)=\frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1}dt$ is the Debye function and $D_2(\theta)=\int_0^1 t\log(t)(1-t)^{\frac{2(1-\theta)}{\theta}}dt$. Quantity ϵ is set to 1e-07 and is used to ensure that the restrictions on the space of θ are maintained. Argument `BivD` of `copulaReg()` in `SemiParBIVProbit` allows the user to employ the desired copula function and can be set to any of the values within brackets next to the copula names in the first column; for example, `BivD = "J0"`. For Clayton, Gumbel and Joe, the number after the capital letter indicates the degree of rotation required: the possible values are `0`, `90`, `180` and `270`. Each of the Clayton, Gumbel and Joe copulae is allowed to be mixed with a rotated version of the same family. This allows the user to simultaneously model positive and negative tail dependence. The options available are `"C0C90"`, `"C0C270"`, `"C180C90"`, `"C180C270"`, `"G0G90"`, `"G0G270"`, `"G180G90"`, `"G180G270"`, `"J0J90"`, `"J0J270"`, `"J180J90"` and `"J180J270"`. Note that for copulae with two association parameters, ζ would represent an additional dependence coefficient.

	$F_m(y_m \mid \mu_m, \sigma_m, \nu_m)$	$f_m(y_m \mid \mu_m, \sigma_m, \nu_m)$	$\mathbb{E}(Y_m)$	$\mathbb{V}(Y_m)$	Support of y_m Parameter ranges
beta ("BE")	$I(y;\alpha_1,\alpha_2)$ $\alpha_1 = \frac{\mu(1-\sigma^2)}{\sigma^2}$ $\alpha_2 = \frac{(1-\mu)(1-\sigma^2)}{\sigma^2}$	$\frac{y^{\alpha_1-1}(1-y)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)}$	μ	$\sigma^2\mu(1-\mu)$	$0<y<1$ $0<\mu<1, 0<\sigma<1$
Dagum ("DAGUM")	$\left\{1+\left(\frac{y}{\mu}\right)^{-\sigma}\right\}^{-\nu}$	$\frac{\sigma\nu}{y}\left[\frac{\left(\frac{y}{\mu}\right)^{\sigma\nu}}{\left\{\left(\frac{y}{\mu}\right)^{\sigma}+1\right\}^{\nu+1}}\right]$	$-\frac{\mu}{\sigma}\frac{\Gamma(-\frac{1}{\sigma})\Gamma(\frac{1}{\sigma}+\nu)}{\Gamma(\nu)}$ if $\sigma>1$	$-\left(\frac{\mu}{\sigma}\right)^2\left[2\sigma\frac{\Gamma(-\frac{2}{\sigma})\Gamma(\frac{2}{\sigma}+\nu)}{\Gamma(\nu)} + \left\{\frac{\Gamma(-\frac{1}{\sigma})\Gamma(\frac{1}{\sigma}+\nu)}{\Gamma(\nu)}\right\}^2\right]$ if $\sigma>2$	$y>0$ $\mu>0, \sigma>0, \nu>0$
Fisk ("FISK")	$\left\{1+\left(\frac{y}{\mu}\right)^{-\sigma}\right\}^{-1}$	$\frac{\sigma y^{\sigma-1}}{\mu^{\sigma}\left\{1+\left(\frac{y}{\mu}\right)^{\sigma}\right\}^2}$	$\frac{\mu\pi/\sigma}{\sin(\pi/\sigma)}$ if $\sigma>1$	$\mu^2\left\{\frac{2\pi/\sigma}{\sin(2\pi/\sigma)}-\frac{(\pi/\sigma)^2}{\sin(\pi/\sigma)^2}\right\}$ if $\sigma>2$	$y>0$ $\mu>0, \sigma>0$
gamma ("GA")	$\frac{1}{\Gamma(\frac{1}{\sigma^2})}\gamma\left(\frac{1}{\sigma^2},\frac{y}{\mu\sigma^2}\right)$	$\frac{1}{(\mu\sigma^2)^{\frac{1}{\sigma^2}}}\frac{y^{\frac{1}{\sigma^2}-1}\exp\left(-\frac{y}{\mu\sigma^2}\right)}{\Gamma(\frac{1}{\sigma^2})}$	μ	$\mu^2\sigma^2$	$y>0$ $\mu>0, \sigma>0$
Gumbel ("GU")	$1-\exp\left\{-\exp\left(\frac{y-\mu}{\sigma}\right)\right\}$	$\frac{1}{\sigma}\exp\left\{\left(\frac{y-\mu}{\sigma}\right)-\exp\left(\frac{y-\mu}{\sigma}\right)\right\}$	$\mu-0.57722\sigma$	$\frac{\pi^2\sigma^2}{6}$	$-\infty<y<\infty$ $-\infty<\mu<\infty, \sigma>0$
inverse Gaussian ("iG")	$\Phi\left\{\frac{1}{\sqrt{y\sigma^2}}\left(\frac{y}{\mu}-1\right)\right\}$ + $\exp\left(\frac{2}{\mu\sigma^2}\right)$ $\Phi\left\{-\frac{1}{\sqrt{y\sigma^2}}\left(\frac{y}{\mu}+1\right)\right\}$	$\frac{1}{\sqrt{2\pi\sigma^2y^3}}\exp\left\{-\frac{1}{2\mu^2\sigma^2y}(y-\mu)^2\right\}$	μ	$\mu^3\sigma^2$	$y>0$ $\mu>0, \sigma>0$
log-normal ("LN")	$\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left\{\frac{\log(y)-\mu}{\sigma\sqrt{2}}\right\}$	$\frac{1}{y\sigma\sqrt{2\pi}}\exp\left[-\frac{\{\log(y)-\mu\}^2}{2\sigma^2}\right]$	$\sqrt{\exp(\sigma^2)}\exp(\mu)$	$\exp(\sigma^2)\{\exp(\sigma^2) -1\}\exp(2\mu)$	$y>0$ $\mu>0, \sigma>0$
logistic ("LO")	$\frac{1}{1+\exp\left(-\frac{y-\mu}{\sigma}\right)}$	$\frac{1}{\sigma}\left\{\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}\left\{1+\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}^{-2}$	μ	$\frac{\pi^2\sigma^2}{3}$	$-\infty<y<\infty$ $-\infty<\mu<\infty, \sigma>0$
normal ("N")	$\frac{1}{2}\left\{1+\mathrm{erf}\left(\frac{y-\mu}{\sigma\sqrt{2}}\right)\right\}$	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$	μ	σ^2	$-\infty<y<\infty$ $-\infty<\mu<\infty, \sigma>0$
reverse Gumbel ("rGU")	$\exp\left\{-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}$	$\frac{1}{\sigma}\exp\left\{\left(-\frac{y-\mu}{\sigma}\right)-\exp\left(-\frac{y-\mu}{\sigma}\right)\right\}$	$\mu+0.57722\sigma$	$\frac{\pi^2\sigma^2}{6}$	$-\infty<y<\infty$ $-\infty<\mu<\infty, \sigma>0$
Singh-Maddala ("SM")	$1-\left\{1+\left(\frac{y}{\mu}\right)^{\sigma}\right\}^{-\nu}$	$\frac{\sigma\nu y^{\sigma-1}}{\mu^{\sigma}\left\{1+\left(\frac{y}{\mu}\right)^{\sigma}\right\}^{\nu+1}}$	$\mu\frac{\Gamma(1+\frac{1}{\sigma})\Gamma(-\frac{1}{\sigma}+\nu)}{\Gamma(\nu)}$ if $\sigma\nu>1$	$\mu^2\left\{\Gamma\left(1+\frac{2}{\sigma}\right)\Gamma(\nu)\Gamma\left(-\frac{2}{\sigma}+\nu\right) -\Gamma\left(1+\frac{1}{\sigma}\right)^2\Gamma\left(-\frac{1}{\sigma}+\nu\right)^2\right\}$ if $\sigma\nu>2$	$y>0$ $\mu>0, \sigma>0, \nu>0$
Weibull ("WEI")	$1-\exp\left\{-\left(\frac{y}{\mu}\right)^{\sigma}\right\}$	$\frac{\sigma}{\mu}\left(\frac{y}{\mu}\right)^{\sigma-1}\exp\left\{-\left(\frac{y}{\mu}\right)^{\sigma}\right\}$	$\mu\Gamma\left(\frac{1}{\sigma}+1\right)$	$\mu^2\left[\Gamma\left(\frac{2}{\sigma}+1\right) -\left\{\Gamma\left(\frac{1}{\sigma}+1\right)\right\}^2\right]$	$y>0$ $\mu>0, \sigma>0$

Table 2: Definition and some properties of the distributions implemented in `SemiParBIVProbit`. These have been parametrised according to Rigby & Stasinopoulos (2005) and are defined in terms of μ, σ and ν (which sometimes represent location, scale and shape). Subscript m can take values 1 and 2; to avoid clutter in the notation we have suppressed m in the main body of the table. The means and variances of `DAGUM`, `FISK` (also known as log-logistic) and `SM` are indeterminate for certain values of σ and ν. If a parameter can only take positive values then the transformation/link function $\log(\cdot - \epsilon)$ is employed, where ϵ is defined and its use explained in the caption of Table 1. If a parameter can take values in $(0, 1)$ then the inverse of the cumulative distribution function of a standardized logistic is used. $I(\cdot;\cdot,\cdot)$ is the regularized beta function, $B(\cdot,\cdot)$ is the beta function, $\Gamma(\cdot)$ is the gamma function, $\gamma(\cdot,\cdot)$ is the lower incomplete gamma function, $\Phi(\cdot)$ is the cdf of a univariate standard normal distribution, and $\mathrm{erf}(\cdot)$ is the error function. Argument `margins` of `copulaReg()` in `SemiParBIVProbit` allows the user to employ the desired marginal distributions and can be set to any of the values within brackets next to the names in the first column; for example, `margins = c("WEI", "DAGUM")`. In many cases the parameters of the distributions determine $\mathbb{E}(y_m)$ and $\mathbb{V}(y_m)$ through functions of them.

is on modeling independent bivariate realizations $(y_{i1}, y_{i2})^{\mathsf{T}}$ as functions of additive predictors, where $i = 1, \ldots, n$ and n is the sample size.

2.2 Predictor specification

All the model's parameters are related to covariates and regression coefficients via additive predictors η's and known monotonic link functions which ensure that the restrictions on the parameter spaces are maintained (see Table 1 and the caption of Table 2 for the links employed). As an example, if σ_1 and σ_2 can only take positive values and we wish to model them as functions of covariates and regression coefficients then we can specify $g(\sigma_{1i}) = \eta_{\sigma_1 i}$ and $g(\sigma_{2i}) = \eta_{\sigma_2 i}$, where the link function $g(\cdot)$ is equal to $\log(\cdot)$. As for the copula parameter, similarly to Klein & Kneib (2016) and Sabeti et al. (2014), we can use $\log(\theta_i - 1) = \eta_{\theta i}$ in the Gumbel case; this would allow for the strength of the (upper tail) dependence between the marginals to vary across observations. Copula models in which each parameter in ϑ is related to an additive predictor can also be regarded as instances of the distributional regression framework described by Klein et al. (2015a).

Let us define a generic predictor η_i as a function of an intercept and smooth functions of sub-vectors of a generic covariate vector $\mathbf{z}_i$. That is,

$$\eta_i = \beta_0 + \sum_{k=1}^{K} s_k(\mathbf{z}_{ki}),\ i = 1, \ldots, n, \tag{2}$$

where $\beta_0 \in \mathbb{R}$ is an overall intercept, $\mathbf{z}_{ki}$ denotes the k^{th} sub-vector of the complete covariate vector $\mathbf{z}_i$ (containing, for example, binary, categorical, continuous, and spatial variables) and the K functions $s_k(\mathbf{z}_{ki})$ represent generic effects which are chosen according to the type of covariate(s) considered. Each $s_k(\mathbf{z}_{ki})$ can be approximated as a linear combination of J_k basis functions $b_{kj_k}(\mathbf{z}_{ki})$ and regression coefficients $\beta_{kj_k} \in \mathbb{R}$, i.e.

$$\sum_{j_k=1}^{J_k} \beta_{kj_k} b_{kj_k}(\mathbf{z}_{ki}). \tag{3}$$

Equation (3) implies that the vector of evaluations $\{s_k(\mathbf{z}_{k1}), \ldots, s_k(\mathbf{z}_{kn})\}^{\mathsf{T}}$ can be written as $\mathbf{Z}_k\boldsymbol{\beta}_k$ with $\boldsymbol{\beta}_k = (\beta_{k1}, \ldots, \beta_{kJ_k})^{\mathsf{T}}$ and design matrix $Z_k[i, j_k] = b_{kj_k}(\mathbf{z}_{ki})$. This allows the predictor in equation (2) to be written as

$$\boldsymbol{\eta} = \beta_0 \mathbf{1}_n + \mathbf{Z}_1 \boldsymbol{\beta}_1 + \ldots + \mathbf{Z}_K \boldsymbol{\beta}_K, \quad (4)$$

where $\mathbf{1}_n$ is an n-dimensional vector made up of ones. Equation (4) can also be written in a more compact way as $\boldsymbol{\eta} = \mathbf{Z}\boldsymbol{\beta}$, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{Z}_1, \ldots, \mathbf{Z}_K)$ and $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^\mathsf{T}, \ldots, \boldsymbol{\beta}_K^\mathsf{T})^\mathsf{T}$.

Each $\boldsymbol{\beta}_k$ has an associated quadratic penalty $\lambda_k \boldsymbol{\beta}_k^\mathsf{T} \mathbf{D}_k \boldsymbol{\beta}_k$ whose role is to enforce specific properties on the k^{th} function, such as smoothness. It is important to note that $\mathbf{D}_k$ only depends on the choice of basis functions. The smoothing parameter $\lambda_k \in [0, \infty)$ controls the trade-off between fit and smoothness, and plays a crucial role in determining the shape of $\hat{s}_k(\mathbf{z}_{ki})$. The overall penalty can be defined as $\boldsymbol{\beta}^\mathsf{T} \mathbf{D} \boldsymbol{\beta}$, where $\mathbf{D} = \text{diag}(0, \lambda_1 \mathbf{D}_1, \ldots, \lambda_K \mathbf{D}_K)$. Finally, the smooth functions are subject to centering (identifiability) constraints (Wood, 2006). In practice, the model's additive predictors and corresponding penalties are set up using the `R mgcv` package (Wood, 2016); Supplementary Material 1 (SM-1) gives some examples of smooth function specification for the reader's convenience.

2.3 Some estimation details

It is well known that the log-likelihood function for a copula model with continuous margins can be written as (e.g., Kolev & Paiva, 2009; Vatter & Chavez-Demoulin, 2015)

$$\ell(\boldsymbol{\delta}) = \sum_{i=1}^{n} \log \left\{ c \left(F_1(y_{1i}|\mu_{1i}, \sigma_{1i}, \nu_{1i}), F_2(y_{2i}|\mu_{2i}, \sigma_{2i}, \nu_{2i}); \zeta_i, \theta_i \right) \right\} + \sum_{i=1}^{n} \sum_{m=1}^{2} \log \left\{ f_m(y_{mi}|\mu_{mi}, \sigma_{mi}, \nu_{mi}) \right\},$$

where c is the copula density and is given by $\frac{\partial^2 C(F_1(y_{1i}), F_2(y_{2i}))}{\partial F_1(y_{1i}) \partial F_2(y_{2i})}$ (here the conditioning on parameters has been suppressed for notational simplicity). The distributional parameters are defined as $\mu_{mi} = g_{\mu_m}^{-1}(\eta_{\mu_m i})$, $\sigma_{mi} = g_{\sigma_m}^{-1}(\eta_{\sigma_m i})$, $\nu_{mi} = g_{\nu_m}^{-1}(\eta_{\nu_m i})$, for $m = 1, 2$, $\zeta_i = g_{\zeta}^{-1}(\eta_{\zeta i})$ and $\theta_i = g_{\theta}^{-1}(\eta_{\theta i})$, where the g's are link functions. The parameter vector $\boldsymbol{\delta}$ is defined as $(\boldsymbol{\beta}_{\mu_1}^\mathsf{T}, \boldsymbol{\beta}_{\mu_2}^\mathsf{T}, \boldsymbol{\beta}_{\sigma_1}^\mathsf{T}, \boldsymbol{\beta}_{\sigma_2}^\mathsf{T}, \boldsymbol{\beta}_{\nu_1}^\mathsf{T}, \boldsymbol{\beta}_{\nu_2}^\mathsf{T}, \boldsymbol{\beta}_{\zeta}^\mathsf{T}, \boldsymbol{\beta}_{\theta}^\mathsf{T})^\mathsf{T}$, which is indeed made up of the coefficient vectors associated with $\eta_{\mu_1 i}$, $\eta_{\mu_2 i}$, $\eta_{\sigma_1 i}$, $\eta_{\sigma_2 i}$, $\eta_{\nu_1 i}$, $\eta_{\nu_2 i}$, $\eta_{\zeta i}$ and $\eta_{\theta i}$. Because of the flexible predictors' structures employed here, the use of a classic (unpenalized) optimization algorithm is likely to result in unduly wiggly estimates (e.g., Ruppert et al., 2003; Wood, 2006). Therefore, we maximize

$$\ell_p(\boldsymbol{\delta}) = \ell(\boldsymbol{\delta}) - \frac{1}{2}\boldsymbol{\delta}^{\mathsf{T}}\mathbf{S}\boldsymbol{\delta}, \tag{5}$$

where ℓ_p is the penalized log-likelihood, $\mathbf{S} = \mathrm{diag}(\mathbf{D}_{\mu_1}, \mathbf{D}_{\mu_2}, \mathbf{D}_{\sigma_1}, \mathbf{D}_{\sigma_2}, \mathbf{D}_{\nu_1}, \mathbf{D}_{\nu_2}, \mathbf{D}_{\zeta}, \mathbf{D}_{\theta})$. The smoothing parameters contained in the $\mathbf{D}$ components make up the overall vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{\mu_1}^{\mathsf{T}}, \boldsymbol{\lambda}_{\mu_2}^{\mathsf{T}}, \boldsymbol{\lambda}_{\sigma_1}^{\mathsf{T}}, \boldsymbol{\lambda}_{\sigma_2}^{\mathsf{T}}, \boldsymbol{\lambda}_{\nu_1}^{\mathsf{T}}, \boldsymbol{\lambda}_{\nu_2}^{\mathsf{T}}, \boldsymbol{\lambda}_{\zeta}^{\mathsf{T}}, \boldsymbol{\lambda}_{\theta}^{\mathsf{T}})^{\mathsf{T}}$.

To maximize (5), we have extended the efficient and stable trust region algorithm with integrated automatic multiple smoothing parameter selection by Marra et al. (2017) to incorporate two and three parametric continuous marginal distributions and two-parameter copula functions, and to link all the model's parameters to additive predictors. Estimation of $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ is carried out as follows. At iteration a, holding $\boldsymbol{\lambda}$ fixed at a vector of values and for a given $\boldsymbol{\delta}^{[a]}$, we maximize equation (5) using a trust region algorithm. That is,

$$\boldsymbol{\delta}^{[a+1]} = \boldsymbol{\delta}^{[a]} + \underset{\mathbf{e}:\|\mathbf{e}\|\leq\Delta^{[a]}}{\arg\min}\ \breve{\ell}_p(\boldsymbol{\delta}^{[a]}), \tag{6}$$

where $\breve{\ell}_p(\boldsymbol{\delta}^{[a]}) = -\left\{\ell_p(\boldsymbol{\delta}^{[a]}) + \mathbf{e}^{\mathsf{T}}\mathbf{g}_p(\boldsymbol{\delta}^{[a]}) + \frac{1}{2}\mathbf{e}^{\mathsf{T}}\boldsymbol{H}_p(\boldsymbol{\delta}^{[a]})\mathbf{e}\right\}$, $\mathbf{g}_p(\boldsymbol{\delta}^{[a]}) = \mathbf{g}(\boldsymbol{\delta}^{[a]}) - \mathbf{S}\boldsymbol{\delta}^{[a]}$ and $\boldsymbol{H}_p(\boldsymbol{\delta}^{[a]}) = \boldsymbol{H}(\boldsymbol{\delta}^{[a]}) - \mathbf{S}$. Vector $\mathbf{g}(\boldsymbol{\delta}^{[a]})$ consists of $\mathbf{g}_{\mu_1}(\boldsymbol{\delta}^{[a]}) = \partial\ell(\boldsymbol{\delta})/\partial\boldsymbol{\beta}_{\mu_1}|_{\boldsymbol{\beta}_{\mu_1}=\boldsymbol{\beta}_{\mu_1}^{[a]}}, \ldots,$ $\mathbf{g}_{\theta}(\boldsymbol{\delta}^{[a]}) = \partial\ell(\boldsymbol{\delta})/\partial\boldsymbol{\beta}_{\theta}|_{\boldsymbol{\beta}_{\theta}=\boldsymbol{\beta}_{\theta}^{[a]}}$, the Hessian matrix has elements $\boldsymbol{H}(\boldsymbol{\delta}^{[a]})_{o,h} = \partial^2\ell(\boldsymbol{\delta})/\partial\boldsymbol{\beta}_o\partial\boldsymbol{\beta}_h^{\mathsf{T}}|_{\boldsymbol{\beta}_o=\boldsymbol{\beta}_o^{[a]},\boldsymbol{\beta}_h=\boldsymbol{\beta}_h^{[a]}}$ where $o, h = \mu_1, \mu_2, \sigma_1, \sigma_2, \nu_1, \nu_2, \zeta, \theta$, $\|\cdot\|$ denotes the Euclidean norm and $\Delta^{[a]}$ is the radius of the trust region which is adjusted through the iterations. The first line of (6) uses a quadratic approximation of $-\ell_p$ about $\boldsymbol{\delta}^{[a]}$ (the so-called model function) in order to choose the best $\mathbf{e}^{[a+1]}$ within the ball centered in $\boldsymbol{\delta}^{[a]}$ of radius $\Delta^{[a]}$, the trust-region. Estimation is made precise and quick by using the analytical score and Hessian. Also, close to the converged solution, the trust-region usually behaves like a classic unconstrained optimization algorithm.

Then, holding the model's parameter vector value fixed at $\boldsymbol{\delta}^{[a+1]}$, we solve the problem

$$\boldsymbol{\lambda}^{[a+1]} = \underset{\boldsymbol{\lambda}}{\arg\min}\ \|\mathbf{M}^{[a+1]} - \mathbf{A}^{[a+1]}\mathbf{M}^{[a+1]}\|^2 - \check{n} + 2\mathrm{tr}(\mathbf{A}^{[a+1]}), \tag{7}$$

where $\mathbf{M}^{[a+1]} = \sqrt{-\boldsymbol{H}(\boldsymbol{\delta}^{[a+1]})}\boldsymbol{\delta}^{[a+1]} + \sqrt{-\boldsymbol{H}(\boldsymbol{\delta}^{[a+1]})}^{-1}\mathbf{g}(\boldsymbol{\delta}^{[a+1]})$, $\mathbf{A}^{[a+1]} = \sqrt{-\boldsymbol{H}(\boldsymbol{\delta}^{[a+1]})}\left(-\boldsymbol{H}(\boldsymbol{\delta}^{[a+1]}) + \mathbf{S}\right)^{-1}\sqrt{-\boldsymbol{H}(\boldsymbol{\delta}^{[a+1]})}$, $\mathrm{tr}(\mathbf{A}^{[a+1]})$ is the number of effective degrees of freedom (edf) of the penalized model and $\breve{n} = 8n$ (if three parameter marginal distributions and the Student-t copula are employed). The formulation of (7) is derived in Marra et al. (2017), whereas the problem is solved using the computational approach by Wood (2004) which is based on the performance iteration idea of Gu (1992). Since $\boldsymbol{H}$ and $\mathbf{g}$ are obtained as a byproduct of the estimation of $\boldsymbol{\delta}$, little computational effort is required to set up the quantities required for (7).

Trust-region algorithms have several advantages over classical alternatives. For instance, in line-search methods, when an iteration falls in a long plateau region, the search for step $\boldsymbol{\delta}^{[a+1]}$ can occur so far away from $\boldsymbol{\delta}^{[a]}$ that the evaluation of the model's log-likelihood may be indefinite or not finite, in which case user's intervention is required. Trust-region methods, on the other hand, always solve sub-problem (6) before evaluating the objective function. So, if this is not finite at the proposed $\boldsymbol{\delta}^{[a+1]}$ then step $\mathbf{e}^{[a+1]}$ is rejected, the trust-region shrunken, and the optimization computed again. The radius is also reduced if there is not agreement between the model and objective functions (i.e., the proposed point in the region is not better than the current one). Reversibly, if such agreement occurs, the trust region is expanded for the next iteration. In summary, $\boldsymbol{\delta}^{[a+1]}$ is accepted if it improves on $\boldsymbol{\delta}^{[a]}$ and allows for the evaluation of $\breve{\ell}_p$, whereas the reduction/expansion of $\Delta^{[a+1]}$ is based on the similarity between model and objective functions. Theoretical and practical details of the method can be found in Nocedal & Wright (2006, Chapter 4), Conn et al. (2000) and Geyer (2015). The latter also discusses the necessary modifications to the sub-problem (6) and the radius for ill-scaled variables.

The methods for estimating $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ are iterated until the algorithm satisfies the criterion $\frac{\left|\ell(\boldsymbol{\delta}^{[a+1]})-\ell(\boldsymbol{\delta}^{[a]})\right|}{0.1+\left|\ell(\boldsymbol{\delta}^{[a+1]})\right|} < 1e-07$. Proving algorithmic convergence when smoothing parameters are estimated in a performance iteration fashion is difficult and to the best of our knowledge this is still an open issue (see, for instance, Gu (2002) for a full discussion). However, the ease of implementation and success of this approach for practical modeling justifies its use in many contexts including the one considered in this paper (e.g., Marra et al., 2017; Wood, 2004, 2006; Yee, 2016).

Starting values for the parameters of the marginals are obtained using the function `gamlss()` within `SemiParBIVProbit`, which has been designed to fit GAMLSS with two or three parameter response distributions and additive predictors, using the estimation approach of this paper. An initial value for the copula parameter is obtained by using a transformation of the empirical

Kendall's association between the responses. The analytical score and Hessian of $\ell(\boldsymbol{\delta})$ required for estimation have been derived in a modular way. It will therefore be easy to extend our algorithm to other copulae and marginal distributions not included in Tables 1 and 2 as long as their cdfs and probability density functions (pdfs) are known and their derivatives with respect to their parameters exist. If a derivative is difficult and/or computationally expensive to compute then appropriate numerical approximations can be employed. The score and Hessian for all combinations of copulae and marginal distributions considered here have been verified using the functions available in the `numDeriv R` package (Gilbert & Varadhan, 2015).

2.4 Further details

At convergence, reliable point-wise confidence intervals for linear and non-linear functions of the model coefficients (e.g., smooth components, copula parameter, joint and conditional predicted probabilities) are obtained using the Bayesian large sample approximation

$$\boldsymbol{\delta} \dot{\sim} \mathcal{N}(\hat{\boldsymbol{\delta}}, -\boldsymbol{H}_p(\hat{\boldsymbol{\delta}})^{-1}). \tag{8}$$

The rationale for using this result is provided in Marra & Wood (2012) for GAMs, whereas some examples of interval construction are given in Radice et al. (2016) for copula binary models. This result can be justified using the distribution of $\mathbf{M}$ discussed in Marra et al. (2017), making the large sample assumption that $\boldsymbol{H}(\boldsymbol{\delta})$ can be treated as fixed, and making the usual Bayesian assumption on the prior of $\boldsymbol{\delta}$ for smooth models (e.g., Silverman, 1985; Wood, 2006). Note that (8) neglects smoothing parameter uncertainty. However, as argued by Marra & Wood (2012) this is not problematic provided that heavy oversmoothing is avoided (so that the bias is not too large a proportion of the sampling variability) and in our experience we found that (8) works well in practice (see Section 3.2 for some simulation-based evidence). To test smooth components for equality to zero, the results discussed in Wood (2013a) and Wood (2013b) are employed.

Proving consistency of the proposed estimator is beyond the scope of this paper although, under certain assumptions on the size of the spline bases and on the asymptotic behavior of the smoothing parameters, this could be straightforwardly demonstrated (e.g., Radice et al., 2016; Vatter & Chavez-Demoulin, 2015).

The proposed approach has generally proved to be fast and reliable. In our experience, convergence failure typically occurs when the model is misspecified and/or the sample size is low compared to the complexity of the model. Examples of misspecification include using marginal distributions that do not fit the responses satisfactorily, and employing a copula which does not accommodate the type and/or strength of dependence between the margins (e.g., using the AMH copula when the association between the margins is strong). `copulaReg()` produces a warning message if there is a convergence issue, and `conv.check()` provides some detailed diagnostics about the fitted model.

2.5 Model building

The flexibility of the proposed framework means that the researcher has to be able to choose a suitable copula function and response distributions as well as select relevant covariates in the model's additive predictors. To this end, we recommend using the Akaike information criterion (AIC) and/or Bayesian information criterion (BIC), normalized quantile residuals (Dunn & Smyth, 1996) and hypothesis testing. Since many choices need to be made, model building can become a time consuming and daunting process when working with large data sets and many candidate regressors. To facilitate the process, we suggest following roughly the guidelines of Klein et al. (2015a) who argue that each of the above criteria is most useful for specific aspects of model building. In short, quantile residuals can be used to assess the goodness of fit of the marginal distributions and AIC/BIC to find a best fitting model given some pre-selected marginal distributions. The criteria are discussed below in more detail.

Quantile residuals for each margin are defined as $\hat{r}_{mi} = \Phi^{-1}\left\{F_m(y_{mi}|\hat{\mu}_{mi}, \hat{\sigma}_{mi}, \hat{\nu}_{mi})\right\}$, for $i = 1, \ldots, n$ and $m = 1, 2$, where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution. If $F_m(y_{mi}|\hat{\mu}_{mi}, \hat{\sigma}_{mi}, \hat{\nu}_{mi})$ is close to the true distribution then the $\hat{r}_{mi}$ follow approximately a standard normal distribution, hence a normal Q-Q plot of such residuals is a useful graphical tool for detecting lack of fit of the marginal distributions. We observed that, in practice, quantile residuals are fairly robust to the exact specification of the additive predictors of the distribution's parameters; this has also been found by Klein et al. (2015a). Therefore, the choice of marginal distributions can be based, for example, on more or less complex model specifications. Also, note that adequate marginal fits are necessary but not sufficient conditions for a satisfactory fit of the multivariate model. Function `post.check()` in `SemiParBIVProbit` produces histograms and normal Q-Q plots of normalized quantile residuals.

AIC and BIC are defined as $-2\ell(\hat{\boldsymbol{\delta}})+2edf$ and $-2\ell(\hat{\boldsymbol{\delta}})+\log(n)edf$, respectively, where the log-likelihood is evaluated at the penalized parameter estimates and $edf = \text{tr}(\hat{\mathbf{A}})$. Given some marginal distributions, AIC/BIC can be used to select a copula function and the most relevant covariates in the model's predictors (using stepwise backward and/or forward selection, for instance). To favor more parsimonious models, small differences in the AIC/BIC values of competing models can be assisted by looking at the significance of the estimated effects; for instance, a covariate with linear effect could be excluded if the respective parameter's p-value is larger than 5%. Here, the relevant `R` functions are `AIC()`, `BIC()`, `summary()` and `plot()`.

3 Simulation study

3.1 Setup

We consider two continuous outcomes, one binary covariate and two continuous regressors. The responses are assumed to follow inverse Gaussian and Singh-Maddala distributions, respectively. The responses are joined using several copulae. Linear and non-linear effects of the regressors on the parameters of the resulting bivariate distributions are also introduced. In practice, this is achieved as follows.

```
library(copula); library(gamlss)
library(SemiParBIVProbit)

cor.cov <- matrix(0.5, 3, 3); diag(cor.cov) <- 1

s1 <- function(x) x*sin(3*x)
s2 <- function(x) sin(2*pi*x) # sin(6*pi*x)

data.gen <- function(cor.cov, s1, s2, FAM){

cov <- rMVN(1, rep(0,3), cor.cov)
cov <- pnorm(cov)
z1 <- cov[, 1]
z2 <- cov[, 2]
z3 <- round(cov[, 3])

eta_mu1    <- 0.5 - 1.25*z2 - 0.8*z3
eta_mu2    <- 0.1 - 0.9*z1 + s1(z2)
```

```
eta_sigma1 <- 1.8
eta_sigma2 <- 0.1
eta_nu     <- 0.2 + z3
eta_theta  <- 0.2 + z1 + s2(z2)

if( FAM == "clayton") theta.para <- exp(eta_theta) + 1e-07
if( FAM == "gumbel")  theta.para <- exp(eta_theta) + 1
if( FAM %in% c("normal", "t")) theta.para <- tanh(eta_theta)

if( FAM %in% c("clayton", "gumbel") ) Cop <- archmCopula(family = FAM,
                                          dim = 2, param = theta.para)
else Cop <- ellipCopula(family = FAM, dim = 2, param = theta.para, df = 4)

speclist1 <- list( mu = exp(eta_mu1), sigma = exp(eta_sigma1) )
speclist2 <- list( mu = exp(eta_mu2), sigma = exp(eta_sigma2), nu = 1,
                                                   tau = exp(eta_nu) )

spec <- mvdc(copula = Cop, c("IG", "GB2"), list(speclist1, speclist2) )
c(rMvdc(1, spec), z1, z2, z3)
}
```

Package `copula` (Yan, 2007) contains functions `archmCopula()`, `ellipCopula`, `mvdc()` and `rMvdc()` which allow us to simulate from the desired copula. Package `gamlss` (Stasinopoulos et al., 2016) contains all functions required to simulate inverse Gaussian (`IG`) and Singh-Maddala (`GB2` with `nu = 1`) deviates, and `rMVN()` (from `SemiParBIVProbit`) allows us to simulate Gaussian correlated variables. The correlation matrix used to associate three simulated Gaussian covariates is `cor.cov`, whereas `cov <- pnorm(cov)` allows us to obtain Uniform(0,1) correlated covariates (e.g., Gentle, 2003). A balanced binary variable is created using `round(cov[, 3])`. Following Vatter & Nagler (2016), `FAM` is set to `"clayton"`, `"gumbel"`, `"normal"` and `"t"`. Functions `s1` and `s2` produce curves with different complexity; the latter is from Vatter & Nagler (2016). The various `eta` refer to the predictors of the marginal and copula parameters. These are transformed in `speclist1` and `speclist2` to ensure that the restrictions on the parameters' spaces of the marginal distributions are maintained (see Table 2). For the same reason, `theta.para` transforms the respective predictor based on the chosen copula (see Table 1). Since `archmCopula()` and `ellipCopula()` do not allow for the use of vectors for `param`, function `data.gen()` is executed as many times as the number of observations that the user wishes to simulate.

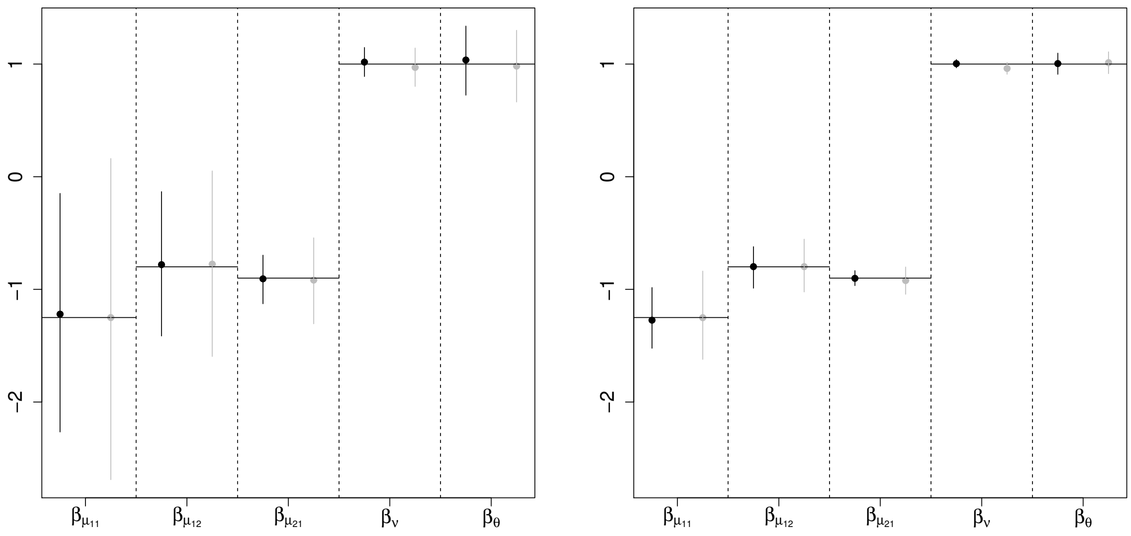
Sample sizes are set to 500 and 5000, the number of replicates to 1000, and the copulae employed are `C0C90`, `G0G90`, `N`, and `T` with 4 degrees of freedom. Models are fitted using `copulaReg()` in `SemiParBIVProbit` and the two-stage approach by Vatter & Chavez-Demoulin (2015) (in this case, using `gamlss()` from the `gamlss R` package for the margins and `gamCopula` (Vatter & Nagler, 2017) for the copula parameter). In both approaches, each smooth function is represented using a penalized low rank thin plate spline with second order penalty and 10 basis functions. For each replicate, smooth function estimates are constructed using 200 equally spaced fixed values in the $(0, 1)$ range (e.g., Radice et al., 2016).

3.2 Results

This section compares the performance of `copulaReg()` with that of `gamlss` + `gamCopula` by assessing the accuracy and precision of the linear and non-linear effect estimates. Some model selection and computation time considerations are also provided. We only discuss the results obtained for the case of data simulated using `FAM = "t"` as they are virtually identical to those obtained using the other copulae and hence do not lead to any further insight. We also repeated the experiments using a Weibull distribution instead of Singh-Maddala for the second margin; the substantive conclusions did not change.

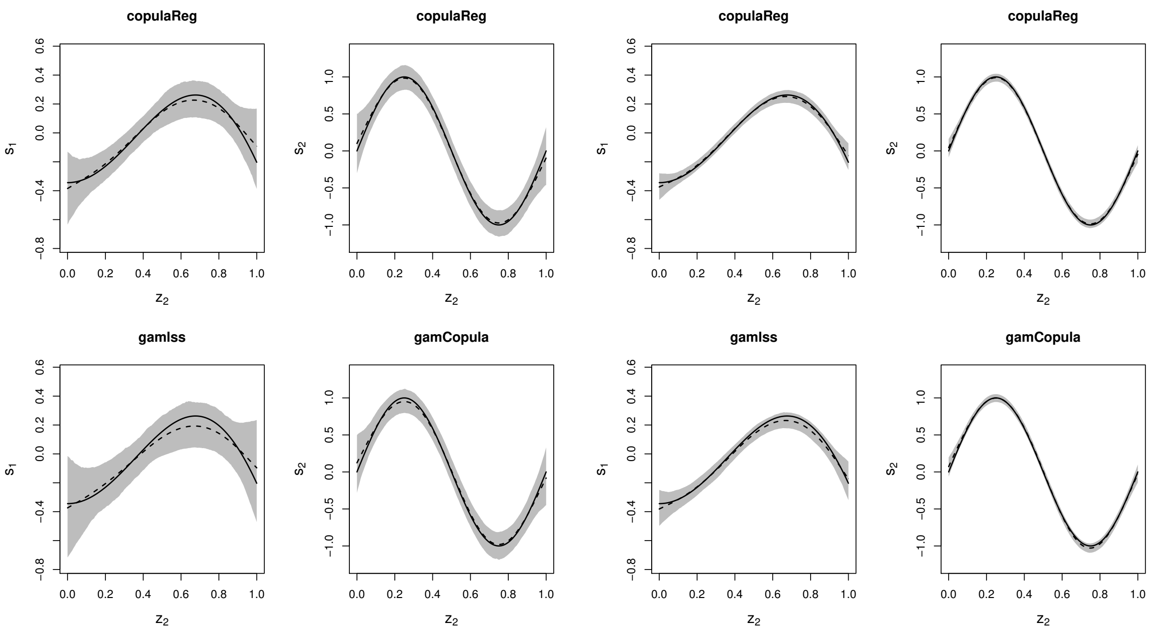
Figures 1 and 2 depict linear and non-linear estimates under the two methods whereas Table 3 reports the bias and root mean squared error of the parameter estimators. All mean estimates are very close to the true values and, as expected, their variability decreases as the sample size increases. Our proposal generally delivers less biased and more efficient estimates. The main exceptions are the estimates related to the copula parameter which are fairly close, with `copulaReg()` yielding slightly better results. Using a more complex function for `s2` (i.e., `sin(6*pi*x)` from Vatter & Nagler (2016)) did not lead to different conclusions when comparing the two methods. We also calculated 95% average coverage probabilities for the smooth functions using point-wise intervals based on the result of Section 2.4. The coverages delivered by `copulaReg()` for s_1 and s_2 were 0.961 and 0.943 for $n = 500$, and 0.952 and 0.956 for $n = 5000$, hence confirming the good performance of the employed approximation.

We also explored whether the correct model is selected by AIC/BIC in the presence of several misspecified models. The latter were based on the 'non-correct' copulae with correct margins, and



(a) $n = 500$ (b) $n = 5000$

Figure 1: Linear coefficient estimates obtained by applying `copulaReg()` (black circles and black vertical bars) and the two-stage method by Vatter & Chavez-Demoulin (2015) (grey circles and grey vertical bars) to data simulated from Student-t copula models with inverse Gaussian (IG) and Singh-Maddala (SM) margins. The two stage approach is implemented using `gamlss()` from the `gamlss R` package for the margins and `gamCopula` for the copula parameter. `copulaReg()` and `gamlss` + `gamCopula` were specified using the Student-t copula with IG and SM margins. Circles indicate mean estimates while bars represent the estimates' ranges resulting from 5% and 95% quantiles. True values are indicated by black solid horizontal lines.



(a) $n = 500$ (b) $n = 5000$

Figure 2: Smooth function estimates obtained by applying `copulaReg()` and `gamlss` + `gamCopula` to data simulated from Student-t copula models with inverse Gaussian and Singh-Maddala margins. True functions are represented by black solid lines, mean estimates by dashed lines and pointwise ranges resulting from 5% and 95% quantiles by shaded areas.

		Bias		RMSE	
		copulaReg	gamlss + gamCopula	copulaReg	gamlss + gamCopula
$n = 500$	$\beta_{\mu_{11}}$	0.070	0.000	0.650	0.911
	$\beta_{\mu_{12}}$	0.019	0.024	0.400	0.504
	$\beta_{\mu_{21}}$	-0.006	-0.018	0.137	0.235
	β_{ν}	0.018	-0.030	0.080	0.108
	β_{θ}	0.029	-0.018	0.188	0.190
	s_1	0.021	0.035	0.084	0.117
	s_2	0.023	0.029	0.118	0.120
$n = 5000$	$\beta_{\mu_{11}}$	0.017	-0.001	0.167	0.245
	$\beta_{\mu_{12}}$	0.002	0.002	0.110	0.146
	$\beta_{\mu_{21}}$	-0.002	-0.022	0.041	0.077
	β_{ν}	0.004	-0.039	0.023	0.051
	β_{θ}	0.005	0.013	0.057	0.061
	s_1	0.008	0.017	0.029	0.043
	s_2	0.010	0.014	0.039	0.042

Table 3: Bias and root mean squared error (RMSE) obtained by applying the `copulaReg()` and `gamlss + gamCopula` parameter estimators to data simulated from Student-t copula models with inverse Gaussian and Singh-Maddala margins. Bias and RMSE for the smooth terms are calculated, respectively, as $n_s^{-1} \sum_{i=1}^{n_s} |\bar{\hat{s}}_i - s_i|$ and $n_s^{-1} \sum_{i=1}^{n_s} \sqrt{n_{rep}^{-1} \sum_{rep=1}^{n_{rep}} (\hat{s}_{rep,i} - s_i)^2}$, where $\bar{\hat{s}}_i = n_{rep}^{-1} \sum_{rep=1}^{n_{rep}} \hat{s}_{rep,i}$, n_s is the number of equally spaced fixed values in the $(0, 1)$ range, and n_{rep} is the number of simulation replicates. In this case, $n_s = 200$ and $n_{rep} = 1000$. The bias for the smooth terms is based on absolute differences in order to avoid compensating effects when taking the sum.

non-correct copulae with one incorrect margin (i.e., Singh-Maddala was replaced with Weibull). For each scenario and replicate, the correct model was always chosen by both criteria.

Finally, comparing the computation times of the two methods for the scenarios considered here when using a 2.20-GHz Intel(R) Core(TM) computer running Windows 7, we generally found that `copulaReg` is 1.3 times faster than `gamlss + gamCopula`. As for the latter approach, the margins' estimation step was the most expensive as it typically amounted to 93% of the total computation time.

4 Empirical illustrations

The next sections illustrate the proposed bivariate copula additive modeling framework using two empirical case studies based on electricity and birth data.

4.1 Analysis of Spanish electricity price and demand data

The aim of this section is to build an explanatory bivariate time-series model for electricity price and demand. In the engineering and econometric literature electricity demands are related with electricity prices throughout the time and one way of capturing this is via transfer function models (e.g., Nogales & Conejo, 2006). Here, we take a different approach by relating price and demand of energy using copulae. We also quantify the effect of prices of raw materials (oil, gas and coal) on electricity price and demand. In the last decade, the issue of modeling electricity price and demand has been the key question to determine the causes of price behavior as well as the macroeconomic significance of the prices of raw materials, since Spain is an importer country. We use working-daily data from January 1, 2002 to October 31, 2008 which are available from the `R` package `MSwM` (Sanchez-Espigares & Lopez-Moreno, 2014).

The first step is to choose the margins. Following the guidelines of Section 2.5, we choose the normal and Gumbel distributions for price and demand, respectively. As for the choice of copula we start off with the normal. We also allow the dependence between the margins, location and scale parameters to vary with raw material prices. In addition to these covariates, we employ a time variable as the underlying electricity prices and demands tend to vary with time, for reasons which may have little or nothing to do with material prices. When we attempt to fit a copula model in which all variables (time, oil, gas and coal prices) enter the five equations (two equations for the location parameters, two for the scale parameters and one equation for the association parameter) the algorithm fails to converge. This suggests that the sample size is perhaps low compared to the complexity of the model. We, therefore, try out more parsimonious specifications. Specifically, we always keep the time variable in all the model's equations and, in a forward selection fashion, choose the best (converged) model as judged by AIC and BIC. This leads to

```
eq.mu.1      <- Price  ~ s(t, k = 60) + s(Oil)          + s(Coal)
eq.mu.2      <- Demand ~ s(t, k = 60) + s(Oil) + s(Gas) + s(Coal)
eq.sigma2.1 <-        ~ s(t, k = 60)
eq.sigma2.2 <-        ~ s(t, k = 60) + s(Oil) + s(Gas)
eq.theta    <-        ~ s(t, k = 60)

fl <- list(eq.mu.1, eq.mu.2, eq.sigma2.1, eq.sigma2.2, eq.theta)

outN <- copulaReg(fl, margins = c("N", "GU"), data = energy, ...)
```

where the `s` are smooth functions of time, oil, gas and coal represented using penalized low rank thin plate splines with default second order penalties and `k` (the number of basis functions) equal to 10, unless otherwise stated; see SM-1 for more details. The value of `k = 60` for the smooth of `t` has been chosen to be a fraction (about 3.4%) of the sample size ($n = 1784$). This value implies that there are approximately 10 basis functions per year. As explained, for instance, in

Peng & Dominici (2008), if `k` per year is small (say 2) then only the long-term trend and seasonality are accounted for and other sub-seasonal and shorter-term variations remain in the data. When building an explanatory time-series model, using 10 or 12 bases per year is more appropriate as variation in the data longer than a timescale of about one week is modeled. As a sensitivity analysis, we increased the `k` values for the `s` terms by several multiples of their original values; the smooth functions of `t` became increasingly wigglier and the effects of raw material prices progressively smoother. This suggested that allowing the time variable to capture very short timescale variation in the data has a detrimental impact on the explanatory power of the model (e.g., Peng & Dominici, 2008; Wood, 2006). The final estimated edfs for the smooth components are 57.7, 1, 6.6 for `eq.mu.1`, 55.2, 8.4, 7.4, 8.8 for `eq.mu.2`, 55.2 for `eq.sigma2.1`, 53.6, 8.3, 7.4 for `eq.sigma2.2`, and 50.7 for `eq.theta`. Recall that when the edf is equal to 1, the respective estimated effect is linear, hence the covariate can enter the model parametrically. On the other hand, the higher the edf the more complex the estimated curve. The total number of estimated parameter is 363 and the computation time was about 12 minutes. The `R` code used for this analysis is given in SM-2.

The overall Kendall's $\hat{\tau}$ and $\hat{\theta}$ are positive and significant (see `summary(outN)`), however some of the individual $\hat{\tau}$ and $\hat{\theta}$ assume negative values. We, therefore, tried all mixed combinations of the Clayton, Gumbel and Joe families, `T`, `F`, `AMH` and `FGM` where the last two can only account for weak dependencies ($-0.18 \leq \tau \leq 0.33$ and $-0.22 \leq \tau \leq 0.22$, respectively). The Gaussian copula is the most supported model by AIC and BIC. Using the Student-t copula virtually yielded the same results as those obtained under the Gaussian copula; this did not come as a surprise as the estimated value for the ζ was 249.15.

Marginal residual plots for the final model are shown in Figure 3 and suggest that the choice of distribution for the first margin is sound, whereas that for the second margin is questionable as the lower-tail residuals are off the reference line. Unfortunately, in this case, it was not possible to find a better fitting distribution. Autocorrelation plots of the response variable and quantile residuals (obtained after fitting the model) show that, while most of the structure has been modeled, short term auto-correlation is still present in the data; see Figure 1 in SM-2. This could be addressed by incorporating in the model autoregressive and/or moving average components but it is beyond the scope of this article.

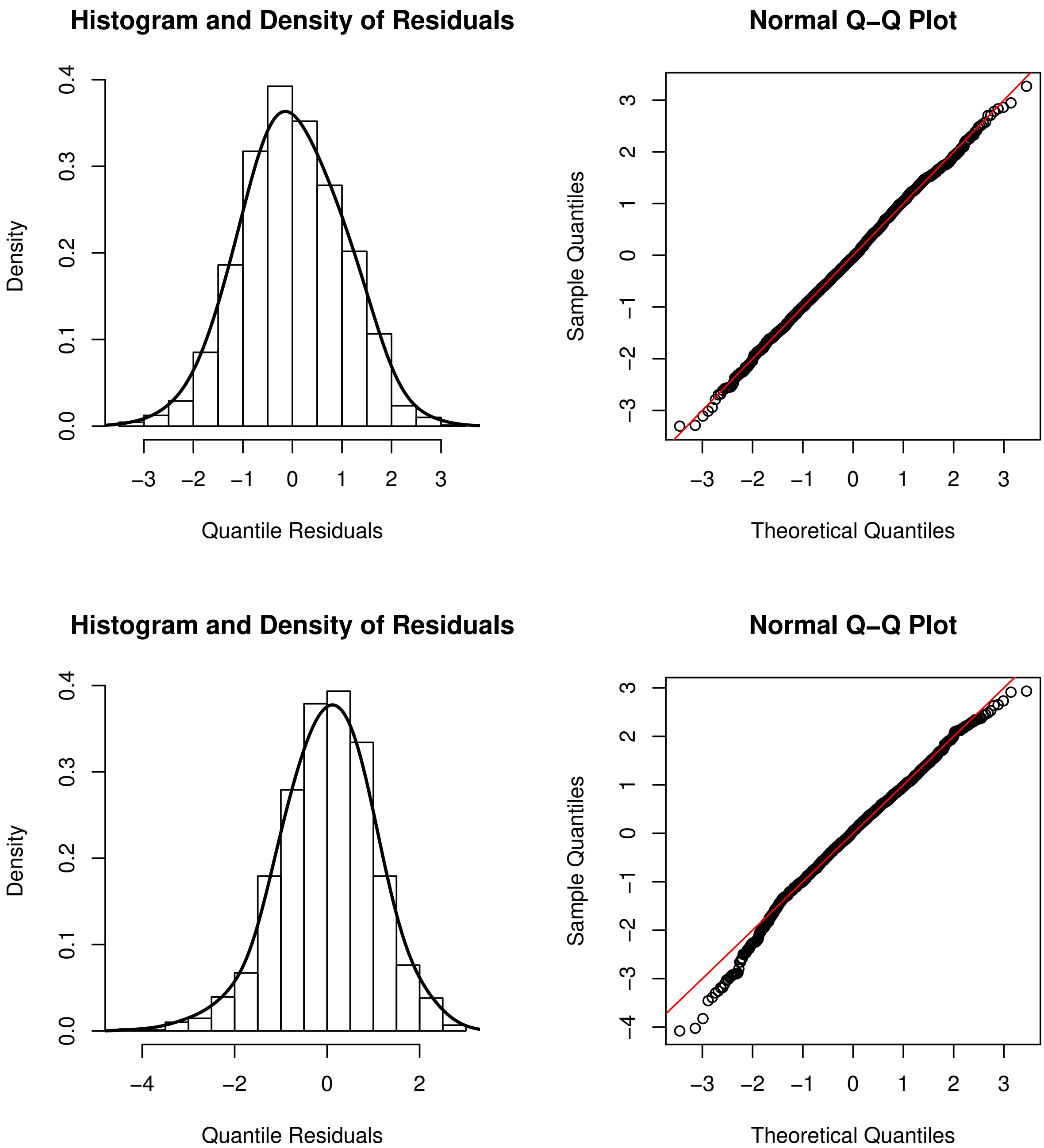


Figure 3: Histograms and normal Q-Q plots of normalized quantile residuals for electricity price (top) and demand (bottom) produced after fitting a Gaussian copula model with normal and Gumbel margins to electricity data.

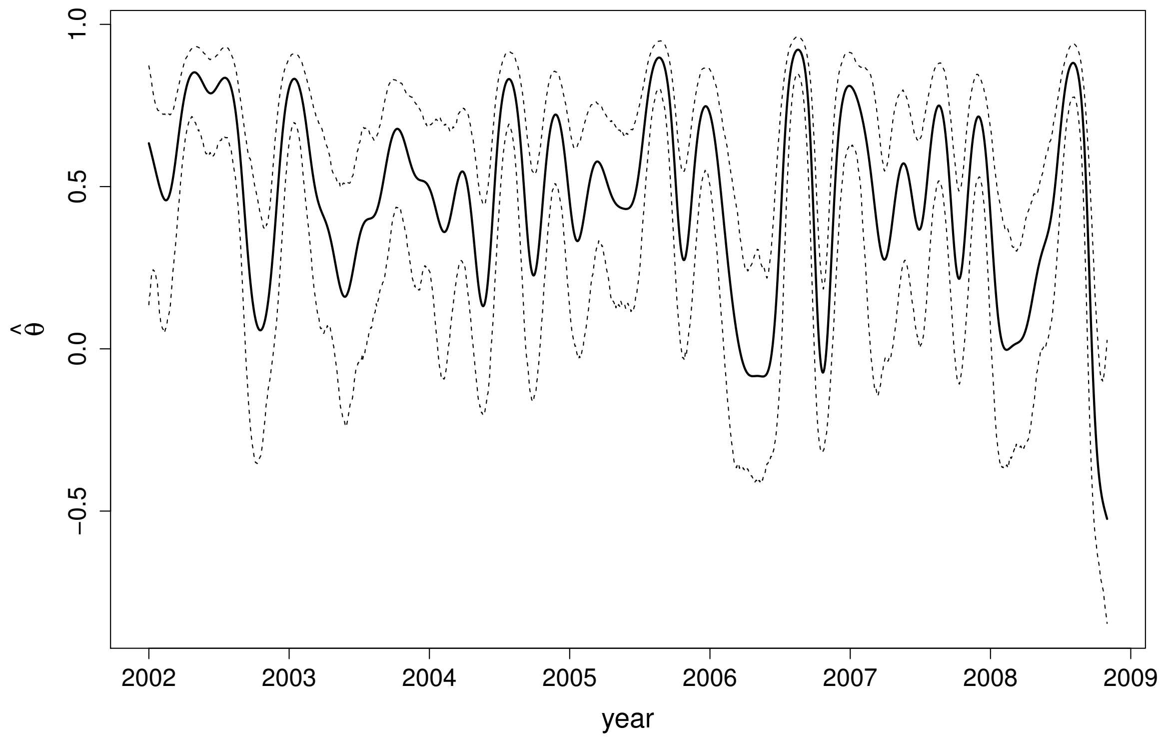


Figure 4: Estimates and 95% intervals for θ over time from a Gaussian copula model with normal and Gumbel margins fitted to electricity price and demand data.

Using the fitted model, we build the plot in Figure 4 which shows that the correlation between `Price` and `Demand` fluctuates around 0.5 (a similar plot could be produced for Kendall's τ). Many of the intervals do not contain zero: after accounting for raw material prices, a significant association between the two responses which varies over time still persists. The reason of these fluctuations are most likely due to variables, such as weather conditions and human habits, that we could not control for because they were not available. Moving on to the covariate effects and focusing, for instance, on the first equation, Figure 5 displays the impacts of `t`, `Oil` and `Coal` on `Price`. The plots show a cyclic trend with maximum and minimum peaks and suggest that on average electricity price tends to linearly increase with `Oil`, and decrease and then stabilize with `Coal` although there is substantial uncertainty related to the last half of the curve. The estimated effect of `Coal` is counter-intuitive and further research is needed to shed light on this. Figure 2 in SM-2 reports the estimates and intervals for σ_1^2. We could also predict joint and conditional probabilities of interest from the model. This point is illustrated in the next section.

When we employed the two-stage estimation approach by Vatter & Chavez-Demoulin (2015), the algorithm converged in about 16 minutes (after setting `control = gamlss.control(n.cyc = 1e2)` for the marginal univariate models); the results were very similar to those reported

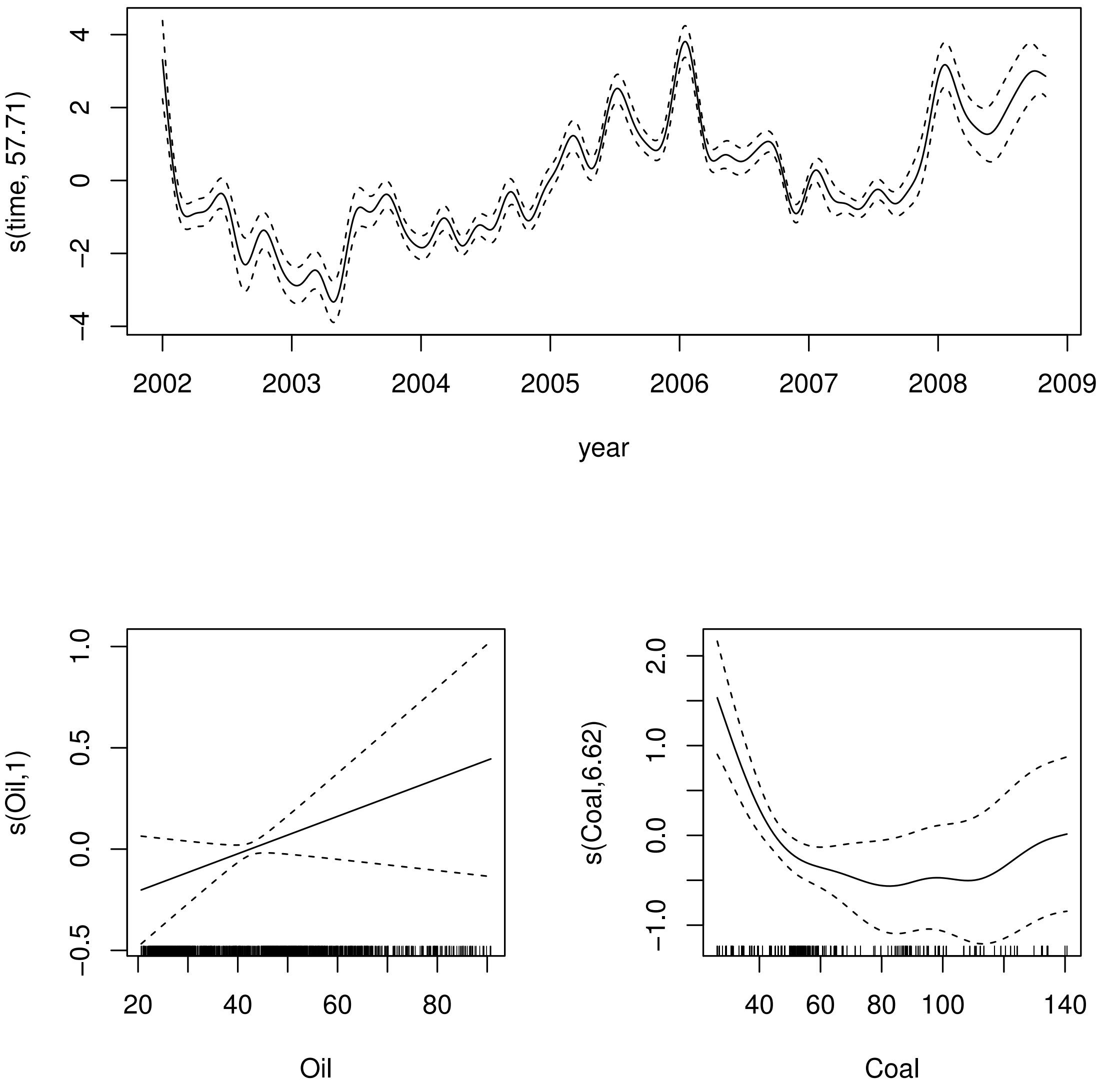


Figure 5: Estimated smooth effects of time, oil and coal prices on electricity price and associated 95% point-wise intervals obtained when fitting a Gaussian copula model with normal and Gumbel margins to electricity price and demand data. The jittered rug plot, at the bottom of each graph, shows the covariate values. The number in brackets in the y-axis caption represents the edf of the respective smooth curve.

above.

It would be interesting to compare the performance of bivariate and univariate GAMLSS in a context of forecasting. However, this would make more sense if the models were specified with this goal in mind. A possibility would be to follow the approach by Marx et al. (2010) and Lee & Durban (2012) in which case the scope of the models would have to be extended accordingly.

4.2 North Carolina birth data analysis

The analysis in this section uses 2010 birth data from the North Carolina Center for Health Statistics (`http://www.schs.state.nc.us/`) which provides details on all the live births occurred within the State of North Carolina, including information on infant and maternal health and parental characteristics. The data cover maternal demographic information, pregnancy related events and outcomes, maternal medical complications, newborn conditions and maternal health behaviors. The choice of variables largely follows the work by Neelon et al. (2012) and the analysis reported below is for female infants (similar results were obtained for male infants). The responses are birth weight in grams (`bwgram`) and gestational age in weeks (`wksgest`). The covariates are maternal ethnicity (`nonhisp`, categorized as non-Hispanic and Hispanic), singleton birth (`multbirth`, born as a multiple or single birth), maternal age (`mage` in years), mother's marital status (`married`) and county (`county`, indicating the North Carolina county of residence of the mother).

Birth weight and gestational age are important determinants of infant and child health; recent evidence has also shown that these factors affect long-term health throughout adulthood (Oreopoulos et al., 2008; Hack et al., 2002). Although both birth weight and gestational age are predictors of future health, modelling these outcomes jointly is essential for a number of reasons. First, birth weight and gestational age are highly correlated and confounded by factors such as intrauterine growth restriction (Slattery & Morrison, 2002). In addition, risk factors for low birth weight, such as maternal age, are also the same risk factors for preterm birth. Finally, evidence suggests that the impact of low birth weight on health may be elevated by low gestational age, and vice-versa (Hediger et al., 2002). Thus, modelling these outcomes independently would present a confounded picture of who is most vulnerable to poor infant health and how best to intervene; a more accurate picture is revealed by modelling these outcome jointly. The goal is, therefore, to build a bivariate copula regression model for the simultaneous analysis of `bwgram` and `wksgest`. The resulting model can, for instance, be used to estimate the association (adjusted for covariates) between `bwgram` and `wksgest` by county, to quantify the effects of covariates on `bwgram` and `wksgest`, and to calculate joint and conditional probabilities of interest.

We first choose the marginal distributions for `bwgram` and `wksgest` based on the guidelines outlined in Section 2.5. The normal Q-Q plots of the normalized quantile residuals and AIC/BIC suggest that the best fits for `bwgram` and `wksgest` are achieved using the logistic and Gumbel distributions (see Figure 3 in SM-4). Using backward selection, we fit bivariate models for `bwgram` and `wksgest` and choose the best model as judged by AIC and BIC. Several copulae are also tried out in a similar way as described in the previous section. The final model is

```
eq.mu.1      <- bwgram  ~ nonhisp + multbirth + married + s(mage) +
                          s(county, bs = "mrf", xt = xt)
eq.mu.2      <- wksgest ~ nonhisp + multbirth + married + s(mage) +
                          s(county, bs = "mrf", xt = xt)
eq.sigma2.1 <-          ~ nonhisp + multbirth + married + s(mage) +
                          s(county, bs = "mrf", xt = xt)
eq.sigma2.2 <-          ~           multbirth + married + s(mage) +
                          s(county, bs = "mrf", xt = xt)
eq.theta    <-          ~ nonhisp + multbirth           + s(mage) +
                          s(county, bs = "mrf", xt = xt)

fl <- list(eq.mu.1, eq.mu.2, eq.sigma2.1, eq.sigma2.2, eq.theta)

outC0 <- copulaReg(fl, margins = c("LO", "GU"), BivD = "C0",
                   data = datNC, ...)
```

where the default number of basis functions for the Gaussian Markov random field smooth term (`mrf`) is equal to the number of covariate's levels or regions (in this case, number of North Carolina counties which is 100), and a Clayton copula is used to join the logistic and Gumbel distributions for the two responses. The first two equations refer to the μ parameters of `bwgram` and `wksgest`, the third and fourth to the σ^2 parameters and the last to θ. These parameters are modeled using additive predictors involving factor, continuous and regional variables. The use of `mrf` smoothers in all equations ensures that the distribution parameters vary smoothly across counties. The total number of observations and estimated parameters are 56940 and 558, respectively, and the computation time was about 25 minutes. Increasing the `k` values for the `s` terms did not affect the final result and only increased the computation time. The estimated edfs for the smooth components are 5.5, 52.0 for `eq.mu.1`, 4.3, 68.6 for `eq.mu.2`, 5.3, 3.9 for `eq.sigma2.1`, 2.6, 45.6 for `eq.sigma2.2`, and 5.3, 36.6 for `eq.theta`. Note that a low value for the edf of the `mrf` smooth term indicates that the estimated county effects are similar with each other and vice-versa. The `R` code used for this analysis is given in SM-3.

Figure 6 shows the joint probabilities of low weight birth babies and premature deliveries in North Carolina when using a copula model and an independence model (which assumes that `bwgram` an `wksgest` are not associated after accounting for covariates). This joint probability was calculated for all the observations in the dataset and then averaged by county. The AIC/BIC values for the copula and independence models suggest that the former provides a better fit to the data. As it can be seen from Figure 6, in this case, assuming independence leads to smaller probabilities. Looking at the copula model's results, the probabilities vary across counties, ranging from around 2% to 7%. The least favorable places to be born are clustered in the northeast of the state, specifically Hertford, Northampton, Halifax, Edgecombe, Gates, Chowan, Perquimans, Pasquotank, Camden, Currituck, Tyrell and Hyde counties.

For comparison purposes, we also employed the two stage approach with the same model specification as that used for `copulaReg()`. Computatio time was around 34 minutes where the majority of time was spent at the margins' stage. Results are similar to those obtained using `copulaReg()` (see Figures 4, 5 and 6 in SM-4).

An analysis similar to that produced in Section 4.1, showing for instance some estimated smooth function and Kendall's $\hat{\tau}$ by county is given in SM-4 to save space.

5 Discussion

We have introduced a modeling framework for bivariate copula additive models for location, scale and shape. The modularity of the estimation approach allows for easy inclusion of potentially any parametric continuous marginal distribution and copula function as long as the cdfs and pdfs are known and their derivatives with respect to their parameters exist. Parameter estimation is carried out within a penalized maximum likelihood estimation framework with integrated automatic multiple smoothing parameter selection, and known and reliable inferential results from the smoothing literature are employed for interval construction and hypothesis testing. The proposed models can

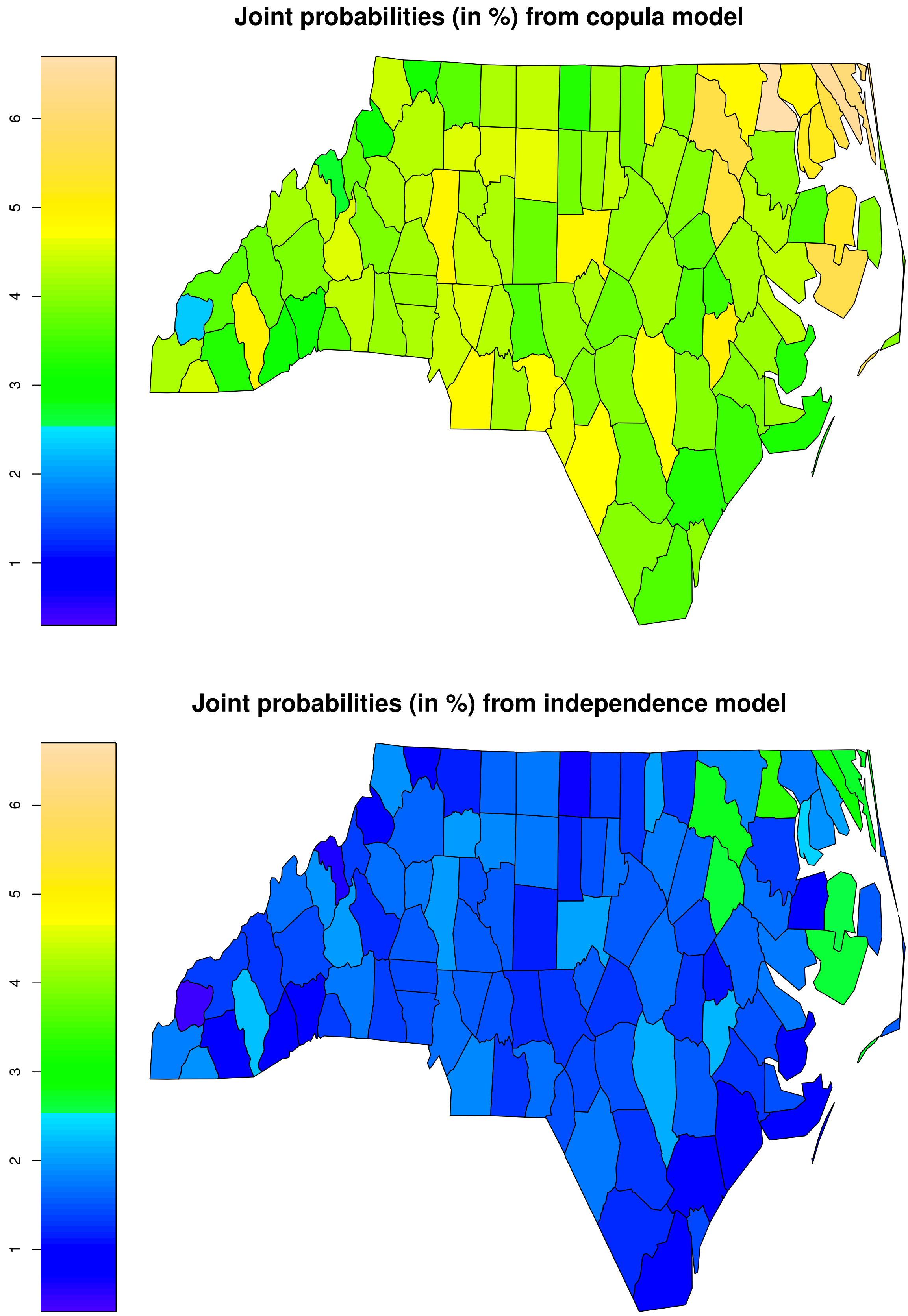


Figure 6: Joint probabilities that `bwgram` is less than or equal to 2500 grams and that `wksgest` is less than or equal to 36 weeks by county in North Carolina. These have been calculated using a Clayton copula model and an independence model (assuming that `bwgram` and `wksgest` are not associated after accounting for covariates).

be easily used via `copulaReg()` in `SemiParBIVProbit` and the potential of the approach has been demonstrated using simulated and real data.

Future releases of `SemiParBIVProbit` will incorporate more copulae and marginal distributions as well as facilities for comparing the predictive ability of competing models based, for instance, on proper scoring rules (Gneiting & Raftery, 2007). Copula models with binary-discrete, binary-continuous, discrete-discrete and discrete-continuous margins will also be made available in the near future. These developments will obviously involve writing and implementing the respective log-likelihood functions, score vectors and Hessian matrices, but the estimation and inferential framework will essentially be unaffected by such changes.

Future research will look into the feasibility of strengthening the framework described in this article by incorporating two-parameter and non-exchangeable copulae (e.g., Durante, 2009; Frees & Valdez, 1998; Brechmann & Schepsmeier, 2013). Another interesting extension would be to consider systems involving more than two responses using C- and D-Vine copulae (e.g., Brechmann & Schepsmeier, 2013).

Acknowledgement

We would like to thank two reviewers for their well thought out suggestions which improved considerably the presentation and message of the article. One reviewer also provided very useful comments on `SemiParBIVProbit` and asked us to implement the Student-t copula as well as mixed versions of the Clayton, Gumbel and Joe families. Finally, we are indebted to Thibault Vatter for various clarifications on the use of `gamCopula`.

References

Acar, E. F., Craiu, V. R., & Yao, F. (2013). Statistical testing of covariate effects in conditional copula models. *Electronic Journal of Statistics*, (pp. 2822–2850).

Brechmann, E. C. & Schepsmeier, U. (2013). Modeling dependence with c- and d-vine copulas: The R package CDVine. *Journal of Statistical Software*, 52(3), 1–27.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula Methods in Finance*. Wiley, London.

Conn, A. R., Gould, N. I. M., & Toint, P. L. (2000). *Trust Region Methods*. Society for Industrial and Applied Mathematics.

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236–245.

Durante, F. (2009). Construction of non-exchangeable bivariate distribution functions. *Statistical Papers*, 50, 383–391.

Fermanian, J. D. & Wegkamp, M. H. (2012). Time-dependent copulas. *Journal of Multivariate Analysis*, 110, 19–29.

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

Gentle, J. E. (2003). *Random number generation and Monte Carlo methods*. Springer-Verlag, London.

Geyer, C. J. (2015). *trust: Trust Region Optimization*. R package version 0.1-6.

Gijbels, I., Veraverbeke, N., & Omelka, M. (2011). Conditional copulas, association measures and their applications. *Computational Statistics and Data Analysis*, 55, 1919–1932.

Gilbert, P. & Varadhan, R. (2015). *numDeriv: Accurate Numerical Derivatives*. R package version 2014.2-1.

Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102, 359–378.

Gu, C. (1992). Cross validating non-gaussian data. *Journal of Computational and Graphical Statistics*, 1, 169–179.

Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer-Verlag, London.

Hack, M., Flannery, D. J., Schluchter, M., Cartar, L., Borawski, E., & Klein, N. (2002). Outcomes in young adulthood for very-low-birth-weight infants. *New England Journal of Medicine*, 346(3), 149–157. 01097.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC, London.

Hediger, M. L., Overpeck, M. D., Ruan, W. J., & Troendle, J. F. (2002). Birthweight and gestational age effects on motor and social development. *Paediatric and Perinatal Epidemiology*, 16(1), 33–46. 00140.

Kauermann, G., Schellhase, C., & Ruppert, D. (2013). Flexible copula density estimation with penalized hierarchical b-splines. *Scandinavian Journal of Statistics*, 40, 685–705.

Klein, N. & Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: a unifying bayesian approach. *Statistics and Computing*, 26(4), 841–860.

Klein, N., Kneib, T., Klasen, S., & Lang, S. (2015a). Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society, Series C*, 64, 569–591.

Klein, N., Kneib, T., Lang, S., & Sohn, A. (2015b). Bayesian structured additive distributional regression with an application to regional income inequality in germany. *Annals of Applied Statistics*, 9(2), 1024–1052.

Kolev, N. & Paiva, D. (2009). Copula-based regression models: A survey. *Journal of Statistical Planning and Inference*, 139, 3847–3856.

Kraemer, N. & Silvestrini, D. (2015). *CopulaRegression: Bivariate Copula Based Regression Models*. R package version 0.1-5.

Kramer, N., Brechmann, E. C., Silvestrini, D., & Czado, C. (2012). Total loss estimation using copula-based regression models. *Insurance: Mathematics and Economics*, 53, 829–839.

Lambert, P. (2007). Archimedean copula estimation using bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, 51, 6307–6320.

Lee, D. J. & Durban, M. (2012). Seasonal modulation smoothing mixed models for time series forecasting. *Proceedings in the 27th International Workshop on Statistical Modelling, Prague, Czech Republic*.

Marra, G. & Radice, R. (2017). *SemiParBIVProbit: Semiparametric Copula Bivariate Probit Modelling*. R package version 3.8-1.

Marra, G., Radice, R., Bärnighausen, T., Wood, S. N., & McGovern, M. E. (2017). A simultaneous equation approach to estimating hiv prevalence with non-ignorable missing responses. *Journal of the American Statistical Association*.

Marra, G. & Wood, S. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics*, 39, 53–74.

Marx, B. D., Eilers, P. H. C., Gampe, J., & Rau, R. (2010). Bilinear modulation models for seasonal tables of counts. *Statistics and Computing*, 20(2), 191–202.

Neelon, B., Anthopolos, R., , & Miranda, M. L. (2012). A spatial bivariate probit model for correlated binary data with application to adverse birth outcomes. *Statistical Methods in Medical Research*, 23(2), 119–133.

Nelsen, R. (2006). *An Introduction to Copulas*. New York: Springer, second edition.

Nocedal, J. & Wright, S. J. (2006). *Numerical Optimization*. New York: Springer-Verlag.

Nogales, F. J. & Conejo, A. J. (2006). Electricity price forecasting through transfer function models. *Journal of the Operational Research Society*, 57, 350–356.

Oreopoulos, P., Stabile, M., Walld, R., & Roos, L. L. (2008). Short-, Medium-, and Long-Term Consequences of Poor Infant Health An Analysis Using Siblings and Twins. *Journal of Human Resources*, 43(1), 88–138. 00278.

Patton, A. J. (2002). Applications of copula theory in financial econometrics. *Ph.D. thesis, University of California, San Diego*.

Peng, R. D. & Dominici, F. (2008). *Statistical Methods for Environmental Epidemiology with R: A Case Study in Air Pollution and Health*. Springer.

Radice, R., Marra, G., & Wojtys, M. (2016). Copula regression spline models for binary outcomes. *Statistics and Computing*, 26(5), 981–995.

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society, Series C*, 54, 507–554.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric Regression.* Cambridge University Press, New York.

Sabeti, A., Wei, M., & Craiu, R. V. (2014). Additive models for conditional copulas. *STAT*, 3, 300–312.

Sanchez-Espigares, J. A. & Lopez-Moreno, A. (2014). *MSwM: Fitting Markov Switching Models*. R package version 1.2.

Segers, J., van den Akker, R., & Werker, B. J. M. (2014). Linear b-spline copulas with applications to nonparametric estimation of copulas. *Annals of Statistics*, 42, 1911–1940.

Shen, X., Zhu, Y., & Song, L. (2008). Linear b-spline copulas with applications to nonparametric estimation of copulas. *Computational Statistics and Data Analysis*, 52, 3806–3819.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of The Royal Statistical Society Series B*, 47, 1–52.

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

Sklar, A. (1973). Random variables, joint distributions, and copulas. *Kybernetica*, 9, 449–460.

Slattery, M. M. & Morrison, J. J. (2002). Preterm delivery. *The Lancet*, 360(9344), 1489–1497. 00704.

Stasinopoulos, M., Rigby, B., Voudouris, V., Akantziliotou, C., Enea, M., & Kiose, D. (2016). *gamlss: Generalised Additive Models for Location Scale and Shape*. R package version 5.0-0.

Trivedi, P. K. & Zimmer, D. M. (2006). *Copula Modeling: An Introduction for Practitioners*.

Vatter, T. & Chavez-Demoulin, V. (2015). Generalized additive models for conditional dependence structures. *Journal of Multivariate Analysis*, 141, 147–167.

Vatter, T. & Nagler, T. (2016). Generalized Additive Models for Pair-Copula Constructions. *ArXiv e-prints*.

Vatter, T. & Nagler, T. (2017). *gamCopula: Generalized Additive Models for Bivariate Conditional Dependence Structures and Vine Copulas*. R package version 0.0-2.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99, 673–686.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction With R*. Chapman & Hall/CRC, London.

Wood, S. N. (2013a). On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100, 221–228.

Wood, S. N. (2013b). A simple test for random effects in regression models. *Biometrika*, 100, 1005–1010.

Wood, S. N. (2016). *mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation*. R package version 1.8-11.

Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *Journal of Statistical Software*, 21, 1–21.

Yee, T. W. (2016). *VGAM: Vector Generalized Linear and Additive Models*. R package version 1.0-3.

Yee, T. W. & Wild, C. J. (1996). Vector generalized additive models. *Journal of Royal Statistical Society, Series B*, 58, 481–493.