Copula Link-Based Additive Models for Bivariate Time-to-Event Outcomes with General Censoring Scheme

Danilo Petti[a,*], Alessia Eletti[b], Giampiero Marra[b], Rosalba Radice[c]

[a]*Department of Economics and Statistical Science, University of Salerno, Via Giovanni Paolo II, Fisciano*
[b]*Department of Statistical Science, University College London, Gower Street, WC1E 6BT London, UK*
[c]*Faculty of Actuarial Science and Insurance, Bayes Business School, City, University of London, 106 Bunhill Row, EC1Y 8TZ London, UK*

Abstract

Bivariate survival outcomes arise frequently in applied studies where the occurrence of two events of interest are associated. Often the exact event times are unknown due to censoring which can manifest in various forms. A general and flexible copula regression model that can handle bivariate survival data subject to various censoring mechanisms, which include a mixture of uncensored, left-, right-, and interval-censored data, is proposed. The proposal permits to specify all model parameters as flexible functions of covariate effects, flexibly model the baseline survival functions by means of monotonic P-splines, characterise the marginals via transformations of the survival functions which yield, e.g., the proportional hazards and odds models as special cases, and model the dependence between events using a wide variety of copulae. The algorithm is based on a computationally efficient and stable penalised maximum likelihood estimation approach with integrated automatic multiple smoothing parameter selection. The proposed model is evaluated in a simulation study and illustrated using data from the Age-Related Eye Disease Study. The modelling framework has been incorporated in the newly-revised `R` package `GJRM`, hence allowing any user to fit the desired model(s) and produce easy-to-interpret numerical and visual summaries.

Keywords: Additive predictor, Bivariate survival data, Copula, Link function, Mixed censoring scheme, Simultaneous penalised parameter estimation.

1. Introduction

Bivariate survival outcomes arise frequently in many research areas such as health and epidemiology. For example, bivariate survival data are often used in clinical trials studying diseases concerning paired organs, where the outcomes of interest are measured on the same individual ~~and as a consequence are associated~~. The main feature of survival data is censoring. For instance, bivariate interval censoring occurs when the events are not precisely observed due to intermittent assessment times and are indeed only known to belong to intervals. When individuals do not experience the two events at their last assessment times, the event statuses are undefined (bivariate right censoring). If some

*Corresponding author
Email addresses: `dpetti@unisa.it` (Danilo Petti), `alessia.eletti.19@ucl.ac.uk` (Alessia Eletti), `giampiero.marra@ucl.ac.uk` (Giampiero Marra), `rosalba.radice@city.ac.uk` (Rosalba Radice)

[1]Additional experiments and implementation details can be found in the Supplement Material. The code used for the case study, instead, can be found on link of relevant CSDA url.

individuals have already experienced both events at the times they enter the study then the data are bivariate left-censored. Sometimes various types of censoring arise simultaneously. This would be the case when, e.g., a disease occurs in one of the paired organs between two consecutive visits and the condition does not occur in the other organ by the end of the study. The aim of this paper is to introduce a flexible regression modelling framework that can handle bivariate survival data under any censoring mechanisms.

Several approaches for modelling bivariate censored data have been proposed. The literature is vast and here we mention a handful of works. Some of them are based on the frailty technique (e.g., Chen et al., 2009, 2014; Martins et al., 2019; Wen & Chen, 2013; Wang et al., 2015; Zhou et al., 2017; Zeng et al., 2017). Others, based on copule and hence more relevant to this paper, are Barthel et al. (2018), Cook & Tolusso (2009), Hu et al. (2017), Kwon et al. (2021), Lo et al. (2020), Marra & Radice (2020), Romeo et al. (2018), Sujica & Van Keilegom (2018), Sun & Ding (2021a) and Wang et al. (2008). These works are not as general and versatile as our proposal. In fact, our modelling framework allows for: a) any bivariate combination of censoring types, whether left-, right-, interval-, or non-censored; b) the exploration of a wide array of dependence structures via copulae; c) all model parameters to be specified as functions of flexible covariate effects via the penalised regression spline methodology (e.g., Wood, 2017); d) the margins of the copula to be modeled via transformations of the survival functions, which give rise to link-based models with the proportional hazards and odds models being particular cases (e.g., Liu et al., 2018); e) the baseline survival functions to be modeled by means of monotonic P-splines which are theoretically advantageous and computationally tractable (e.g., Pya & Wood, 2015). There are currently no such models (and related fitting procedures) available in the literature nor software implementations.

Despite the proposed model is complex in that it allows for many layers of structure, there is no price to pay in terms of usability and interpretability. In fact, the model has been incorporated in the newly-revised software package `GJRM` (Marra & Radice, 2022), written for the programming language `R` (R Development Core Team, 2022), which significantly eases the use of the framework. An additional benefit is that post estimation functions have been extended and integrated within `GJRM` to allow any user to produce interpretable results. Parameter estimation relies on an extension of the stable and fast algorithm presented in Marra & Radice (2020) which is based on a simultaneous penalised maximum likelihood approach with integrated automatic multiple smoothing parameter selection. The proposed model together with fast and reliable software implementation represents a significant advance in modelling bivariate survival data. An interesting feature of the proposal is that it is very flexible and at the same time parametric. Sir David R. Cox, among others, has encouraged the broader use of parametric models for empirical modelling (e.g., Reid, 1994). In that spirit, our modelling framework enables a large amount of exploration via many and diverse functional structures which may help to uncover new patterns and trends in the data.

The potential of the approach is illustrated via a simulation study as well as using data from the Age-Related Eye Disease Study (AREDS), a multi-center randomised clinical trial exploring the development and progression of age-related macular degeneration (AMD), sponsored by the National Eye Institute (Group, 1999). The analysis aims

to quantify the effect of clinical risk factors on the joint risks of AMD progression as well as to predict the progression profiles of AMD patients with different characteristics.

The article is organised as follows. Section 2 discusses various details of the proposed model. Section 3 introduces the model log-likelihood and explains how parameter estimation, whereas Section 4 shows some inferential results. In Section 5, data from the AREDS are analysed and the main findings presented. Section 6 concludes the paper with a discussion. The On-line Supplementary Material provides more details on the log-likelihood construction, reports the analytical expressions for the score and Hessian matrix, discusses the findings of a simulation study, and illustrates the use of `GJRM` on the AREDS data.

2. The Model

Let us consider the pair of survival times (T_{1i}, T_{2i}), a vector of covariates $\mathbf{x}_i$, for $i = 1, 2, \ldots, n$ where n represents the sample size, and a generic parameter vector $\boldsymbol{\delta} \in \mathbb{R}^W$ of dimension W. We assume that T_{1i} and T_{2i} have marginal survival functions written as $S_v(t_{vi}|\mathbf{x}_{vi}; \boldsymbol{\beta}_v) = P(T_{vi} > t_{vi}|\mathbf{x}_{vi}; \boldsymbol{\beta}_v) \in (0, 1)$, for $v = 1, 2$, and a joint survival function expressed as $S(t_{1i}, t_{2i}|\mathbf{x}_i; \boldsymbol{\delta}) = P(T_{1i} > t_{1i}, T_{2i} > t_{2i}|\mathbf{x}_i; \boldsymbol{\delta})$. The survival times are linked via a copula as follows

$$S(t_{1i}, t_{2i}|\mathbf{x}_i; \boldsymbol{\delta}) = C\left(S_1(t_{1i}|\mathbf{x}_{1i}; \boldsymbol{\beta}_1), S_2(t_{2i}|\mathbf{x}_{2i}; \boldsymbol{\beta}_2); m\left\{\eta_{3i}(\mathbf{x}_{3i}; \boldsymbol{\beta}_3)\right\}\right),$$

where $\boldsymbol{\delta}^\mathsf{T} = (\boldsymbol{\beta}_1^\mathsf{T}, \boldsymbol{\beta}_2^\mathsf{T}, \boldsymbol{\beta}_3^\mathsf{T})$, $\mathbf{x}_{1i}$, $\mathbf{x}_{2i}$ and $\mathbf{x}_{3i}$ are vectors of covariates, which can be sub-vectors of or equal to $\mathbf{x}_i$, with associated coefficient vectors $\boldsymbol{\beta}_1 \in \mathbb{R}^{W_1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{W_2}$ and $\boldsymbol{\beta}_3 \in \mathbb{R}^{W_3}$, $W = W_1 + W_2 + W_3$, $C : (0, 1)^2 \rightarrow (0, 1)$ is a uniquely defined 2-dimensional copula function with coefficient $\theta_i = m\left\{\eta_{3i}(\mathbf{x}_{3i}; \boldsymbol{\beta}_3)\right\}$ modelling the potentially varying dependence of (T_{1i}, T_{2i}) across observations, $\eta_{3i}(\mathbf{x}_{3i}; \boldsymbol{\beta}_3) \in \mathbb{R}$ is a predictor which includes generic additive covariate effects, and m is a monotonic and differentiable one-to-one transformation function ensuring that the restriction on the space of the parameter being considered is not violated. A similar specification has been previously adopted; see, e.g., Emura et al. (2021), Geerdens et al. (2018) and Marra & Radice (2020). The copulae implemented in `GJRM` are reported in Table 1, which also shows the relation between θ and the Kendall's $\tau \in [-1, 1]$. If a copula can only account for positive dependence (e.g., Gumbel) then its counter-clockwise rotated versions can also be obtained (Brechmann & Schepsmeier, 2013).

The marginal survival functions can be written as

$$g_v\left[S(t_{vi}|\mathbf{x}_{vi}; \boldsymbol{\beta})\right] = \eta_{vi}(t_{vi}, \mathbf{x}_{vi}; \mathbf{f}_v(\boldsymbol{\beta}_v)), \tag{1}$$

where $g_v : (0, 1) \rightarrow \mathbb{R}$ is a monotone and twice continuously differentiable link function with bounded derivatives, $\eta_{vi}(t_{vi}, \mathbf{x}_{vi}; \mathbf{f}_v(\boldsymbol{\beta}_v)) \in \mathbb{R}$ is an additive predictor which models the baseline hazard and several types of covariate effects, and $\mathbf{f}_v(\boldsymbol{\beta}_v)$ has the role of imposing a monotonicity constraint when evaluating the baseline function of time contained in the additive predictor (see the next section). Equation (1) can also be written as $S(t_{vi}|\mathbf{x}_{vi}; \boldsymbol{\beta}_v) =$

Table 1: Definition of the copulae implemented in the R package GJRM, with corresponding parameter range of association parameter θ, one-to-one transformation function of θ, relation between Kendall's τ and θ, and range of τ. $\Phi_2(\cdot,\cdot;\theta)$ denotes the cumulative distribution function (cdf) of the standard bivariate normal distribution with correlation coefficient θ, and $\Phi(\cdot)$ the cdf of the univariate standard normal distribution. $t_{2,\zeta}(\cdot,\cdot;\zeta,\theta)$ indicates the cdf of the standard bivariate Student-t distribution with correlation θ and fixed $\zeta \in (2,\infty)$ degrees of freedom, and $t_\zeta(\cdot)$ denotes the cdf of the univariate Student-t distribution with ζ degrees of freedom. $A(t) = 1 - \left[t^{-\theta} + (1-t)^{-\theta}\right]^{-\frac{1}{\theta}}$ is the Pickands dependence function of the Galambos copula. $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1}dt$ is the Debye function and $D_2(\theta) = \int_0^1 t\log(t)(1-t)^{\frac{2(1-\theta)}{\theta}}dt$. Quantities Q and R are given by $1+(\theta-1)(u_1+u_2)$ and $Q^2-4\theta(\theta-1)u_1u_2$, respectively. The Kendall's τ for "PL" is computed numerically since no analytical expression is available. Argument BivD of gjrm() in GJRM allows the user to employ the desired copula and can be set to any of the values within brackets next to the copula names in the first column; for example, BivD = "C0". For Clayton, Galambos, Gumbel and Joe, the number after the capital letter indicates the degree of rotation required: the possible values are 0, 90, 180 and 270. The rotations are defined as $C_{90}(u_1,u_2;\theta) = u_2 - C(1-u_1,u_2)$, $C_{180}(u_1,u_2;\theta) = u_1+u_2-1+C(1-u_1,1-u_2)$ and $C_{270}(u_1,u_2;\theta) = u_1 - C(u_1,1-u_2)$.

Copula	$C(u_1,u_2;\theta)$	Range of θ	Transf.of θ	Kendall's τ	Range of τ
AMH ("AMH")	$\frac{u_1u_2}{1-\theta(1-u_1)(1-u_2)}$	$[-1,1]$	$\tanh^{-1}(\theta)$	$-\frac{2}{3\theta^2}\left\{\theta+(1-\theta)^2\log(1-\theta)\right\}+1$	$[-0.1817, 1/3]$
Clayton ("C0")	$\left(u_1^{-\theta}+u_2^{-\theta}-1\right)^{-1/\theta}$	$(0,\infty)$	$\log(\theta)$	$\frac{\theta}{\theta+2}$	$(0,1]$
FGM ("FGM")	$u_1u_2\left\{1+\theta(1-u_1)(1-u_2)\right\}$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{9}\theta$	$[-2/9, 2/9]$
Frank ("F")	$-\theta^{-1}\log\left\{1+(\exp\{-\theta u_1\}-1)(\exp\{-\theta u_2\}-1)/(\exp\{-\theta\}-1)\right\}$	$\mathbb{R}\setminus\{0\}$	$-$	$1-\frac{4}{\theta}\left[1-D_1(\theta)\right]$	$(-1,1)\setminus\{0\}$
Galambos ("GAL")	$u_1u_2\exp\left[\left\{(-\log u_1)^{-\theta}+(-\log u_2)^{-\theta}\right\}^{-1/\theta}\right]$	$(0,\infty)$	$\log(\theta)$	$\int_0^1 \frac{t(1-t)}{A(t)}A''(t)dt$	$(0,1]$
Gaussian ("N")	$\Phi_2\left(\Phi^{-1}(u_1),\Phi^{-1}(u_2);\theta\right)$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$	$[-1,1]$
Gumbel ("G0")	$\exp\left[-\left\{(-\log u_1)^{\theta}+(-\log u_2)^{\theta}\right\}^{1/\theta}\right]$	$[1,\infty)$	$\log(\theta-1)$	$1-\frac{1}{\theta}$	$[0,1]$
Joe ("J0")	$1-\left\{(1-u_1)^{\theta}+(1-u_2)^{\theta}-(1-u_1)^{\theta}(1-u_2)^{\theta}\right\}^{1/\theta}$	$(1,\infty)$	$\log(\theta-1)$	$1+\frac{4}{\theta^2}D_2(\theta)$	$(0,1]$
Plackett ("PL")	$\left(Q-\sqrt{R}\right)/\left\{2(\theta-1)\right\}$	$(0,\infty)$	$\log(\theta)$	$-$	$(-1,1]$
Student's t ("T")	$t_{2,\zeta}\left(t_\zeta^{-1}(u_1),t_\zeta^{-1}(u_2);\zeta,\theta\right)$	$[-1,1]$	$\tanh^{-1}(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$	$[-1,1]$

$G_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}$, where G_v is an inverse link function. The cumulative hazard and hazard functions are defined as $H_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v) = -\log\left[G_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}\right]$, and

$$h_v(t_{vi}|\mathbf{x}_{vi};\boldsymbol{\beta}_v) = -\frac{G_v'\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}}{G_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}(\boldsymbol{\beta}_v))\right\}}\frac{\partial\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))}{\partial t_{vi}}, \tag{2}$$

respectively, where $G_v'\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\} = \partial G_v\left\{\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))\right\}/\partial\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))$. Table 2 displays the functions g, G and G' implemented in GJRM.

Table 2: Link functions implemented in GJRM. Φ and ϕ are the cumulative distribution and density functions of a univariate standard normal distribution.

Model	Link $g(S)$	Inverse link $g^{-1}(\eta) = G(\eta)$	$G'(\eta)$
Prop. hazards ("PH")	$\log\{-\log(S)\}$	$\exp\{-\exp(\eta)\}$	$-G(\eta)\exp(\eta)$
Prop. odds ("PO")	$-\log\left(\frac{S}{1-S}\right)$	$\frac{\exp(-\eta)}{1+\exp(-\eta)}$	$-G^2(\eta)\exp(-\eta)$
probit ("probit")	$-\Phi^{-1}(S)$	$\Phi(-\eta)$	$-\phi(-\eta)$

2.1. Predictor specification

The key difference between $\eta_{vi}(t_{vi},\mathbf{x}_{vi};\mathbf{f}_v(\boldsymbol{\beta}_v))$, for $v = 1,2$, and $\eta_{3i}(\mathbf{x}_{3i};\boldsymbol{\beta}_3)$, where in the latter $\mathbf{f}_3$ is the identity vector function, is that the two former predictors must include smooth functions of times t_{vi} which can be treated as regressors. In fact, the construction of the design matrices for the three additive predictors follows the same philosophy. We, therefore, consider a generic $\eta_{\nu i}$ ($\nu = 1,2,3$), where the dependence on the covariates

and parameters is momentarily dropped, an overall covariate vector $\mathbf{z}_{\nu i}$ containing $\mathbf{x}_{\nu i}$ and $t_{\nu i}$ when $\nu = 1, 2$, and $\mathbf{z}_{3i} = \mathbf{x}_{3i}$. For simplicity, the dimensions of $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are assumed to be W_1 and W_2.

An additive predictor can be defined as

$$\eta_{\nu i} = \beta_{\nu 0} + \sum_{k_\nu=1}^{K_\nu} s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i}),\ i = 1, \ldots, n, \tag{3}$$

where $\beta_{\nu 0} \in \mathbb{R}$ is an overall intercept, $\mathbf{z}_{\nu k_\nu i}$ denotes the k_ν^{th} sub-vector of the complete vector $\mathbf{z}_{\nu i}$ and the K_ν functions $s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$ represent generic effects which are chosen according to the type of covariate(s) considered. Each $s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu i})$ can be represented as a linear combination of $J_{\nu k_\nu}$ basis functions $b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i})$ and regression coefficients $f_{\nu k_\nu j_{\nu k_\nu}}(\beta_{\nu k_\nu j_{\nu k_\nu}}) \in \mathbb{R}$, that is (e.g., Wood, 2017)

$$\sum_{j_{\nu k_\nu}=1}^{J_{\nu k_\nu}} f_{\nu k_\nu j_{\nu k_\nu}}(\beta_{\nu k_\nu j_{\nu k_\nu}}) b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i}). \tag{4}$$

The above formulation implies that the vector of evaluations $\{s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu 1}), \ldots, s_{\nu k_\nu}(\mathbf{z}_{\nu k_\nu n})\}^\mathsf{T}$ can be written as $\mathbf{Z}_{\nu k_\nu}\mathbf{f}_{\nu k_\nu}(\boldsymbol{\beta}_{\nu k_\nu})$ with $\mathbf{f}_{\nu k_\nu}(\boldsymbol{\beta}_{\nu k_\nu}) = (f_{\nu k_\nu 1}(\beta_{\nu k_\nu 1}), \ldots, f_{\nu k_\nu J_{\nu k_\nu}}(\beta_{\nu k_\nu J_{\nu k_\nu}}))^\mathsf{T}$ and design matrix $\mathbf{Z}_{\nu k_\nu}[i, j_{\nu k_\nu}] = b_{\nu k_\nu j_{\nu k_\nu}}(\mathbf{z}_{\nu k_\nu i})$. Therefore, equation (3) can be written as

$$\boldsymbol{\eta}_\nu = \beta_{\nu 0}\mathbf{1}_n + \mathbf{Z}_{\nu 1}\mathbf{f}_{\nu 1}(\boldsymbol{\beta}_{\nu 1}) + \ldots + \mathbf{Z}_{\nu K_\nu}\mathbf{f}_{\nu K_\nu}(\boldsymbol{\beta}_{\nu K_\nu}), \tag{5}$$

where $\mathbf{1}_n$ is an n-dimensional vector made up of ones, or in a more compact way as $\boldsymbol{\eta}_\nu = \mathbf{Z}_\nu\mathbf{f}_\nu(\boldsymbol{\beta}_\nu)$, where $\mathbf{Z}_\nu = (\mathbf{1}_n, \mathbf{Z}_{\nu 1}, \ldots, \mathbf{Z}_{\nu K_\nu})$ and $\mathbf{f}_\nu(\boldsymbol{\beta}_\nu) = (\beta_{\nu 0}, \mathbf{f}_{\nu 1}(\boldsymbol{\beta}_{\nu 1})^\mathsf{T}, \ldots, \mathbf{f}_{\nu K_\nu}(\boldsymbol{\beta}_{\nu K_\nu}^\mathsf{T}))^\mathsf{T}$. Note that smooth functions are subject to centering identifiability constraints (Wood, 2017). Each $\boldsymbol{\beta}_{\nu k}$ has an associated quadratic penalty $\lambda_{\nu k_\nu}\boldsymbol{\beta}_{\nu k_\nu}^\mathsf{T}\mathbf{D}_{\nu k_\nu}\boldsymbol{\beta}_{\nu k_\nu}$ which has to be used during model fitting to enforce specific properties on the k_ν^{th} function, such as smoothness. Smoothing parameter $\lambda_{\nu k_\nu} \in [0, \infty)$ controls the trade-off between fit and smoothness, whereas $\mathbf{D}_{\nu k_\nu}$ only depends on the choice of the basis functions. The overall penalty can be defined as $\boldsymbol{\beta}_\nu^\mathsf{T}\mathbf{D}_\nu\boldsymbol{\beta}_\nu$, where $\mathbf{D}_\nu = \text{diag}(0, \lambda_{\nu 1}\mathbf{D}_{\nu 1}, \ldots, \lambda_{\nu K_\nu}\mathbf{D}_{\nu K\nu})$. The above formulation allows for many types of flexible covariate effects (e.g., non-linear, random, spatial, interactions). In fact, several definitions of basis functions and penalty terms are supported in `GJRM` which are based on Wood (2017). The time effects are instead modelled using the monotonic P-spline approach which will guarantee that the estimated survival functions are monotonically decreasing or equivalently that the hazard functions are positive. Specifically, using a slightly simplified notation, let $s_v(t_{vi}) = \sum_{j_v=1}^{J_v} f_{vj_v}(\beta_{vj_v})b_{vj_v}(t_{vi})$, where the b_{vj_v} are B-spline basis functions of at least second order built over the interval $[a, b]$, based on equally spaced knots, and the $f_{vj_v}(\beta_{vj_v})$ are spline coefficients. A sufficient condition for $s_v'(t_{vi}) \geq 0$ over $[a, b]$ is that $f_{vj_v}(\beta_{vj_v}) \geq f_{vj_v}(\beta_{vj_v-1}), \forall j$ (e.g., Leitenstorfer & Tutz, 2006). Such condition can be imposed by re-parametrising the spline coefficient vector so that $\mathbf{f}_v(\boldsymbol{\beta}_v) = \boldsymbol{\Sigma}_v\{\beta_{v1}, \exp(\beta_{v2}), \ldots, \exp(\beta_{vJ_v})\}^\mathsf{T}$ and $\boldsymbol{\Sigma}_v[\iota_{v1}, \iota_{v2}] = 0$ if $\iota_{v1} < \iota_{v2}$

and $\mathbf{\Sigma}_v[\iota_{v1}, \iota_{v2}] = 1$ if $\iota_{v1} \geq \iota_{v2}$, with ι_{v1} and ι_{v2} denoting the row and column entries of the respective matrix. When setting up the penalty term we penalise the squared differences between adjacent β_{vj_v}, starting from β_{v2}, using $\mathbf{D}_v = \mathbf{D}_v^{*\mathsf{T}}\mathbf{D}_v^*$ where $\mathbf{D}_v^*$ is a $(J_v - 2) \times J_v$ matrix made up of zeros except that $\mathbf{D}_v^*[\iota_v, \iota_v + 1] = -\mathbf{D}_v^*[\iota_v, \iota_v + 2] = 1$ for $\iota_v = 1, ..., J_v - 2$ (Pya & Wood, 2015). Matrix $\mathbf{\Sigma}_v$ can be absorbed into $\mathbf{Z}_v$. An alternative approach to modelling baseline hazards in the copula context is provided by Kwon et al. (2021) who adopted M-splines.

2.2. Remarks

When working with interval-censored observations, the model set up needs to account for the information contained in the lower and upper bounds of the censoring intervals. Therefore, for each margin, two distinct design matrices (based on the two bounds) and hence additive predictors are required. The covariates and parameter vector $\boldsymbol{\beta}_v$ used in their construction will be the same.

In equation (2), $\partial\eta_{vi}(t_{vi}, \mathbf{x}_{vi}; \mathbf{f}_v(\boldsymbol{\beta}_v))/\partial t_{vi}$ is required. Based on the results of the previous paragraph, $\eta_{vi}(t_{vi}, \mathbf{x}_{vi}; \mathbf{f}_v(\boldsymbol{\beta}_v))$ can be written as $\mathbf{Z}_{vi}(t_{vi}, \mathbf{x}_{vi})^{\mathsf{T}}\mathbf{f}_v(\boldsymbol{\beta}_v)$ which means that the quantity of interest can be calculated as $\lim_{\varepsilon \to 0} \left\{ \frac{\mathbf{Z}_{vi}(t_{vi}+\varepsilon, \mathbf{x}_{vi}) - \mathbf{Z}_{vi}(t_{vi}-\varepsilon, \mathbf{x}_{vi})}{2\varepsilon} \right\}^{\mathsf{T}} \mathbf{f}_v(\boldsymbol{\beta}_v) = \mathbf{Z}_{vi}'^{\mathsf{T}}\mathbf{f}_v(\boldsymbol{\beta}_v)$, where $\mathbf{Z}'_{vi}$ can be conveniently obtained by finite differencing.

Formulation (4) requires a value for $J_{\nu k_\nu}$. This is especially relevant when modelling the effects of continuous covariates. As explained by Vatter & Chavez-Demoulin (2015), among others, all that is required is to set $J_{\nu k_\nu}$ to an arbitrary value that allows for enough flexibility in estimating the related smooth term; penalisation during model fitting will then ensure that a good balance between fit and parsimony is achieved.

The general model formulation introduced in the previous two sections yields the proportional hazards and odds models as special cases; for details on this, we refer the reader to, e.g., Liu et al. (2018) whose developments are based on the same conceptual survival modelling framework adopted here. Other important benefits are that quantities such as $h_v(t_{vi}|\mathbf{x}_{vi}; \boldsymbol{\beta}_v)$ can be directly obtained without the need for numerical integration, and that time-dependent effects can be easily incorporated in the model via terms like $s_{vk_v}(t_{vi})\mathbf{x}_{vk_vi}$.

3. Parameter Estimation

Let T_{vi} denote the true event time, for $v = 1, 2$. In the case of censoring, T_{vi} is only known to lie within the interval (L_{vi}, R_{vi}), where L_{vi} and R_{vi} represent left and right censoring times. If $L_{vi} = 0$ then the i^{th} observation for the v margin is defined as left-censored. When $R_{vi} = \infty$, the observation is classified as right-censored. If L_{vi} and R_{vi} take on finite distinct non-zero values then the observation is interval-censored. Exact observations relate to the case $L_{vi} = R_{vi}$. Since we are dealing with a bivariate response, there will be sixteen possible censoring combinations to account for; these can be characterised through the indicator functions $\gamma_{I_{vi}}$ and $\gamma_{U_{vi}}$, where $\gamma_{I_{vi}}$ takes value 1 if the i^{th} observation is interval-, right- or left-censored and 0 otherwise. Similarly, $\gamma_{U_{vi}}$ is 1 if the i^{th} observation is uncensored and 0 otherwise.

$$
\begin{aligned}
\ell(\boldsymbol{\delta}) &= \gamma_{U_{1i}}\gamma_{U_{2i}} \sum_{i=1}^{n} \log f(t_{1i}, t_{2i}) + \gamma_{I_{1i}}\gamma_{I_{2i}} \sum_{i=1}^{n} \log P(T_{1i} \in (l_{1i}, r_{1i}], T_{2i} \in (l_{2i}, r_{2i}]) \\
&+ \gamma_{U_{1i}}\gamma_{I_{1i}} \sum_{i=1}^{n} \log \left[\int_{l_{21}}^{r_{2i}} f(t_{1i}, y) dy \right] + \gamma_{I_{1i}}\gamma_{U_{1i}} \sum_{i=1}^{n} \log \left[\int_{l_{1i}}^{r_{1i}} f(y, t_{2i}) dy \right] \\
&= \gamma_{U_{1i}}\gamma_{U_{2i}} \sum_{i=1}^{n} \log \left[\frac{\partial^2}{\partial t_{1i} \partial t_{2i}} C \left\{ G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i})); \theta_i \right\} \right] \\
&+ \gamma_{I_{1i}}\gamma_{I_{2i}} \sum_{i=1}^{n} \log \Big[C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} \\
&- C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} \Big] \\
&+ \gamma_{U_{1i}}\gamma_{I_{1i}} \sum_{i=1}^{n} \log \left[\frac{\partial}{\partial t_{1i}} \Big(C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} \Big) \right] \\
&+ \gamma_{I_{1i}}\gamma_{U_{1i}} \sum_{i=1}^{n} \log \left[\frac{\partial}{\partial t_{2i}} \Big(C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(t_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(t_{2i})); \theta_i\} \Big) \right].
\end{aligned}
$$

The case of interval censoring incorporates both right and left censoring. So, if the i^{th} observation for the v margin is right-censored then $r_{vi} = \infty$. If it is left-censored then $l_{vi} = 0$. The terms of the above log-likelihood have been derived as follows:

- T_{1i} uncensored and T_{2i} uncensored (in this case, $t_{1i} = r_{1i} = l_{1i}$ and $t_{2i} = r_{2i} = l_{2i}$):

$$
\begin{aligned}
f(t_{1i}, t_{2i}) &= \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} F(t_{1i}, t_{2i}) = \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} [1 - S(t_{1i}) - S(t_{2i}) + S(t_{1i}, t_{2i})] \\
&= \frac{\partial^2}{\partial t_{1i} \partial t_{2i}} C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(t_{2i})); \theta_i\}.
\end{aligned}
$$

- T_{1i} interval-censored and T_{2i} interval-censored:

$$
\begin{aligned}
&P(l_{1i} < T_{1i} < r_{1i}, l_{2i} < T_{2i} < r_{2i}) = P(T_{1i} < r_{1i}, T_{2i} < r_{2i}) - P(T_{1i} < l_{1i}, T_{2i} < r_{2i}) \\
&- P(T_{1i} < r_{1i}, T_{2i} < l_{2i}) + P(T_{1i} < l_{1i}, T_{2i} < l_{2i}) \\
&= F(r_{1i}, r_{2i}) - F(l_{1i}, r_{2i}) - F(r_{1i}, l_{2i}) + F(l_{1i}, l_{2i}) \\
&= S(l_{1i}, l_{2i}) - S(l_{1i}, r_{2i}) - S(r_{1i}, l_{2i}) + S(r_{1i}, r_{2i}) \\
&= C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(l_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} \\
&\quad - C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} + C\{G_1(\eta_{1i}(r_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\}.
\end{aligned}
$$

Recall that, using the above formulation, all scenarios deriving from any combination of right-, left- and interval-censored bivariate outcomes can be produced.

- T_{1i} uncensored and T_{2i} interval-censored (the "swapped" case can be trivially derived by switching the subscripts where required):

$$\int_{l_{2i}}^{r_{2i}} f(t_{1i}, y)dy = \int_{0}^{r_{2i}} f(t_{1i}, y)dy - \int_{0}^{l_{2i}} f(t_{1i}, y)dy = \frac{\partial}{\partial t_{1i}} F(t_{1i}, r_{2i}) - \frac{\partial}{\partial t_{1i}} F(t_{1i}, l_{2i})$$
$$= \frac{\partial}{\partial t_{1i}} [1 - S_1(t_{1i}) - S_2(r_{2i}) + S(t_{1i}, r_{2i})] - \frac{\partial}{\partial t_{1i}} [1 - S_1(t_{1i}) - S_2(l_{2i}) + S(t_{1i}, l_{2i})]$$
$$= \frac{\partial}{\partial t_{1i}} \left[C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(r_{2i})); \theta_i\} - C\{G_1(\eta_{1i}(t_{1i})), G_2(\eta_{2i}(l_{2i})); \theta_i\} \right].$$

  As above, the right- and left-censored cases can be easily worked out.

The reader is referred to Supplementary Material-Section A for the more explicit version of the log-likelihood. As explained in Section 2.1, quadratic penalties have to be employed during model fitting to calibrate the trade-off between fit and smoothness. Therefore, we maximise

$$\ell_p(\boldsymbol{\delta}) = \ell(\boldsymbol{\delta}) - \frac{1}{2}\boldsymbol{\delta}^\mathsf{T}\mathbf{S}\boldsymbol{\delta}, \tag{6}$$

where ℓ_p is the penalised log-likelihood, $\mathbf{S} = \text{diag}(\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3)$, $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ are overall penalties that take the form specified in Section 2.1 and ~~which contain~~ include $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$ and $\boldsymbol{\lambda}_3$ ~~, and $\boldsymbol{\lambda}_v = (\lambda_{v1}, \ldots, \lambda_{vK_v})^\mathsf{T}$~~. The smoothing parameters can be collected in the vector $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^\mathsf{T}, \boldsymbol{\lambda}_2^\mathsf{T}, \boldsymbol{\lambda}_3^\mathsf{T})^\mathsf{T}$.

Model fitting is challenging in this context because of the non-linear dependence of $\mathbf{f}_v(\boldsymbol{\beta}_v)$ on $\boldsymbol{\beta}_v$, the requirement of estimating $\boldsymbol{\lambda}$ in a data driven manner, and the need for providing a stable and fast implementation that is computationally solid and practically usable. To this end, we employ the stable and fast trust region algorithm presented in Marra & Radice (2020) which is based on a simultaneous penalised maximum likelihood approach with integrated automatic multiple smoothing parameter selection. A major challenge with the implementation of such algorithm is that the analytical score vector and Hessian matrix of $\ell(\boldsymbol{\delta})$ are required. Given the generality and complexity of the model, deriving such quantities has been a rather tedious and time-consuming task; these are given in Sections B and C of the Supplementary Material, and have been thoroughly checked and verified numerically. Starting values for the marginal survival models are obtained by combining the use of the shape constrained smoothing approach of Pya & Wood (2015) with the procedure detailed in Liu et al. (2018). An initial value for the copula parameter is worked out by using a transformation of the empirical τ between the responses. The simulation study in Supplementary Material-Section D supports the empirical effectiveness of the estimation framework. Briefly, several sample sizes ($n = 300, 1000, 1500$ and 2000) are considered as well as both mild (62.86% and 44.98%) and high (84.82% and 77.13%) censoring levels. Overall, the modelling framework performs consistently well, even for the lowest sample size. The parametric effects and smooth effects were properly recovered in all of the scenarios considered, exhibiting both low bias and RMSE. The estimation of the quantities related to the copula dependence parameter is

more challenging and shows some bias when compared to the other model parameters, although performance is still deemed satisfactory. As expected, parameter estimation is more difficult in the presence of high censoring, due to the loss of information implied by censoring itself. Note, however, that both of these challenging settings improve markedly as the sample size increases.

The number of parameters in the model can be quantified using the notion of number of effective degrees of freedom (edf). The edf for a model containing only unpenalised terms would clearly be equal to W, whereas that for a penalised model can be written as $W-\zeta$, where $\zeta = \text{tr}\left\{(-\boldsymbol{H}+\mathbf{S})^{-1}\mathbf{S}\right\}$ and $\boldsymbol{H}$ is the Hessian matrix. This shows the role that $\boldsymbol{\lambda}$ (contained in $\mathbf{S}$) plays in determining the model edf, which indeed is a value in the range $[W-\zeta, W]$. The definition of the edf of a single smooth or penalised term follows the same logic and has a value smaller than or equal to $J_{\nu k_\nu}$.

4. Inference

Inferential results can be borrowed from known theory for general penalised likelihood-based models. Specifically, at convergence, reliable confidence intervals for any linear or non-linear function of $\boldsymbol{\delta}$ are obtained by exploiting the Bayesian large sample approximation (e.g., Wahba, 1983; Wood et al., 2016)

$$\boldsymbol{\delta} \dot{\sim} \mathcal{N}(\hat{\boldsymbol{\delta}}, \mathbf{V}_{\boldsymbol{\delta}}), \tag{7}$$

where $\hat{\boldsymbol{\delta}} = \arg\max_{\boldsymbol{\delta}} \ell_p(\boldsymbol{\delta})$ and $\mathbf{V}_{\boldsymbol{\delta}} = (-\boldsymbol{H}(\hat{\boldsymbol{\delta}}) + \mathbf{S})^{-1}$.

Employing the Bayesian framework for penalised models implicitly assumes that overly complex models are less likely than simpler or smoother ones; this translates into the prior specification $f_{\boldsymbol{\delta}} \propto \exp\left(-1/2\boldsymbol{\delta}^\mathsf{T}\mathbf{S}\boldsymbol{\delta}\right)$. As elaborated by Wood (2017, Section 6.10, see also references therein), the Bayesian covariance matrix gives close to across-the-function *frequentist* coverage probabilities since it includes both the bias and variance components in a frequentist sense. Intervals for nonlinear functions of $\boldsymbol{\delta}$ can be conveniently obtained via posterior simulation (see, e.g., Marra & Radice (2020) for an example. P-values for the terms in the model can be reliably obtained by using the results summarised in Wood (2017, Section 6.12) which are based on $\mathbf{V}_{\boldsymbol{\delta}}$. Note that for the parametric (unpenalised) terms in the model, the corresponding entries in $\mathbf{S}$ (contained in $\mathbf{V}_{\boldsymbol{\delta}}$) are equal to zero. This would be equivalent to using the classical frequentist result, based on $-\boldsymbol{H}(\hat{\boldsymbol{\delta}})$, for such terms.

5. Application to AREDS data

The proposed approach is applied to a dataset from the AREDS available through the `R` package `CopulaCenR` (Sun & Ding, 2021b), which includes 629 Caucasian participants. The event of interest is the progression to late-AMD disease, which is the most common cause of blindness in developed countries (Swaroop et al., 2009). Due to intermittent assessment times (every 6 months up to the first 6 years and every 1 year thereafter), the exact time

when each eye progressed to late-AMD is only known to lie in a certain interval. More specifically, less than half of the subjects developed late-AMD in both eyes (bivariate interval-censored); around 20% of the subjects developed late-AMD in one eye and did not develop late-AMD in the other eye before the end of the study (mixed interval- and right-censored); more than one third of the subjects did not develop late-AMD in both eyes (bivariate right-censored).

The dataset contains three covariates potentially related with AMD progression: `SevScaleBL` for baseline AMD severity score (a factor variable with values between 4 and 8 with a higher value indicating more severe AMD), `ENROLLAGE` for baseline age (a numeric variable), and `rs2284665` for a genetic variant (a factor variable with levels 0, 1 and 2 which represent GG, GT and TT, respectively).

For the marginal equations, the smooth functions of `ENROLLAGE` and the time variables were represented using penalised thin plate regression splines with second order penalty (Wood, 2017) and monotonic penalised B-splines (see Section 2.1), respectively. The number of bases used for each smooth was 10; increasing this value did not lead to visible changes in the estimated curves. The remaining variables entered the predictors of the marginals linearly. All link functions shown in Table 2 were considered in the modelling. For both margins, `PO` was found to yield the smallest AIC and BIC. As for the copula, we started off with the Gaussian and then, based on the (negative or positive) sign of the dependence, we tried out alternative specifications that were consistent with this initial finding. Using a 2.60-GHz Intel(R) Core(TM) computer running Windows 10, the average computing time to fit a model was about 9 seconds and the length of the model parameter vector was 43. Using the AIC and BIC, where, in their construction, the model edf was used in place of the number of model parameters, the chosen model is based on the Plackett copula with `PO` margins. The `R` code used to fit the models, and to produce all the numerical and visual summaries commented below can be found in Supplementary Material-Section E. Using the second and third best copulae did not change the conclusions of the analysis.

Table 3: AREDS data. Parameters estimates, standard errors and p-values obtained from fitting the model using `gjrm()`.

	Left Eye		Right Eye	
Parametric Eff.	Estimate (Std.error)	$Pr(>\|z\|)$	Estimate (Std.error)	$Pr(>\|z\|)$
`(Intercept)`	-18.0368(4.39)	4.09e-05	-33.2811 (10.89)	0.002246
`ENROLLAGE`	-	-	0.0364 (0.01)	0.011592
`SevScale5`	0.6707 (0.24)	0.00556	0.8187 (0.25)	0.001365
`SevScale6`	1.0049 (0.22)	6.90e-06	1.2957 (0.23)	4.81e-07
`SevScale7`	1.9255 (0.23)	< 2e-16	2.4270 (0.25)	< 2e-16
`SevScale8`	2.8208 (0.31)	< 2e-16	3.2793 (0.32)	< 2e-16
`rs22846651`	0.3269 (0.16)	0.04966	0.4589 (0.16)	0.006467
`rs22846652`	0.6058 (0.23)	0.00927	0.7874 (0.22)	0.000481

All coefficients in the two model marginal equations as well as the dependence parameter are significant (see Table 3). The estimated regression coefficients of `SevScaleBL`, which are 0.67, 1.00, 1.93, 2.82 in the equation for the left eye and 0.82, 1.21, 2.43, 3.28 in that for the right eye, imply, as expected, that the subjects with higher baseline AMD severity score have a higher risk than the subjects with lower baseline AMD severity score. As for the genetic variant, `rs2284665`, the estimated parameters are 0.33 and 0.61 for the left eye equation, and 0.46 and 0.79 for the

right one. This is consistent with the interpretation that participants with TT genotype group have the highest risk of developing the disease, followed by participants with GT genotype group.

Figure 1 shows the estimated functional forms for the effect of ENROLLAGE and times of the selected model. Note that the smooth function for ENROLLAGE in the second equation has not been reported as the effect was linear ($edf = 1$), which indeed indicates that there is a constantly increasing risk associated with age. As for the first equation, the estimated smooth function confirms this increasing trend. Also, since there are few subjects who are younger than 60 and older than 80, the point-wise intervals are larger at lower and higher age values. The plots for the time variables exhibit increasing monotonic trends, suggesting again that the risk increases with time.

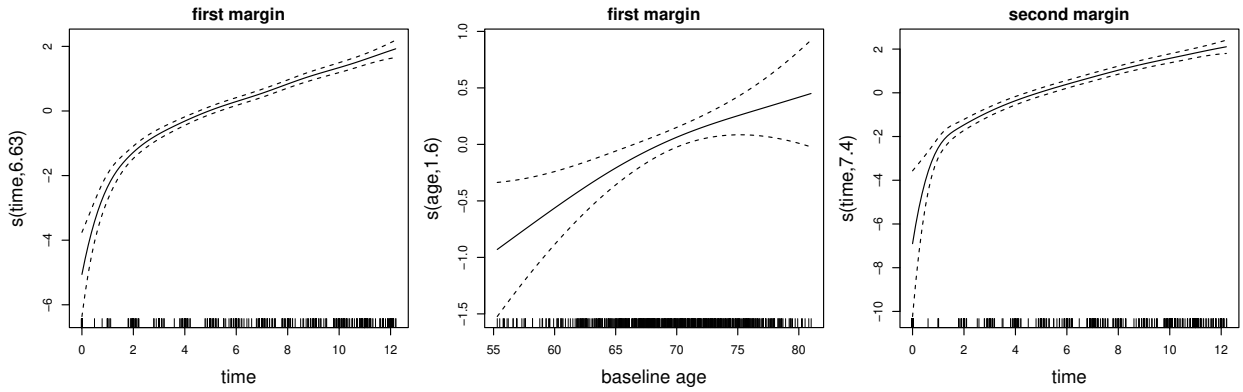

Figure 1: AREDS data. Baseline risks and smoothed effect of baseline age (ENROLLAGE), for the first equation only. 95% point-wise intervals are based on the result mentioned in Section 3. The rug plot, at the bottom of each graph, shows the values of the considered variable. The number in brackets in the y-axis caption of each plot represents the edf of the respective estimated smooth function.

The estimated Kendall's τ is 0.36 which implies moderate dependence in AMD progression between the two eyes. Given the capabilities of the proposed modelling framework, we also specified a model where the dependence parameter is expressed as a flexible function of the covariates. This feature can help understand how and which covariates modify the strength of the dependence across observations. In this case, however, the coefficients were found not to be significant (see Supplementary Material-Section E). It is worth noting that such specifications are likely to be more successful in finding covariate patterns when the number of observations is higher than that available for this study.

Using the chosen model, we produced joint survival functions under several scenarios. The left panel of Figure 2 displays the joint progression-free probability contours for subjects who are 69 years old, with AMD severity score equal to 6 for both eyes, but with different rs2284665 genotypes. The middle panel of Figure 2 shows the joint progression-free probability contours for subjects who are 69 year old, with GT genotype, but with different severity scores (4, 6 and 8). Finally, the right panel of the figure plots the joint progression-free probability contours for GT genotype subjects, with AMD severity score equal to 6 in both eyes, but different ages (56, 69 and 81). In the left panel, it can be clearly seen that the three genotype groups are separated, with the GG group having the largest progression-free probabilities. In the middle panel, the difference between the three AMD severity groups is rather pronounced, with the highest AMD severity group having the smallest progression-free probabilities. Finally, the right panel shows

how the progression-free probabilities are higher for younger subjects as compared to older subjects. The scenarios considered here illustrate how valuable the proposed modelling framework is in characterising and identifying AMD patients at a higher risk of developing late-AMD. Of course, several other scenarios can be considered and other quantities of interest worked out. For example, one could be interested in visualising conditional and marginal survival probabilities.

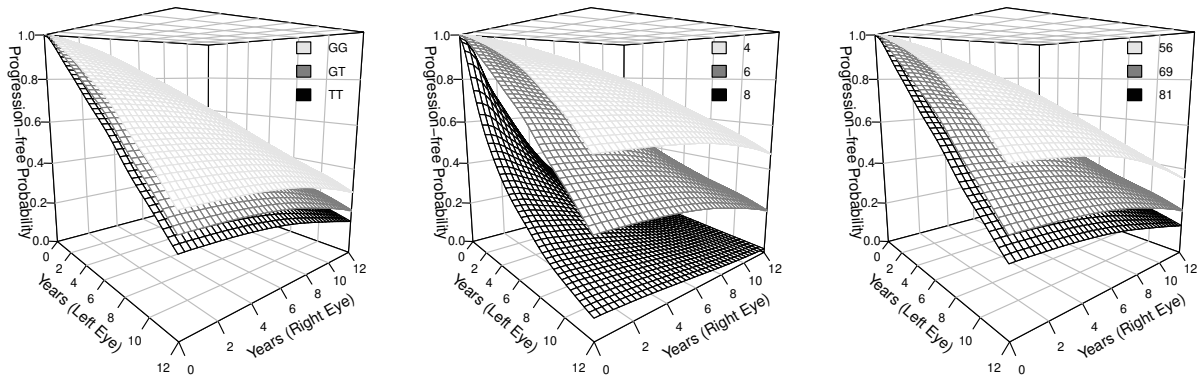

Figure 2: AREDS data. Joint progression-free probability contours for progression to late-AMD disease (in years) in the left and right eyes, under different scenarios. In left panel, age is set to 69, and AMD severity score to 6 for both eyes. In the middle panel, age is set to 69, and genotype to GT. In the right panel, genotype is set to GT, and AMD severity score to 6 in both eyes.

6. Discussion

We have introduced a copula link-based additive model for bivariate time-to-event outcomes under various types of censoring mechanisms. Model fitting is based on the simultaneous estimation of all model parameters and relies on a penalised maximum likelihood approach with integrated stable and efficient automatic multiple smoothing parameter selection. Inferential results are also readily available. All developments have been integrated within the `R` package `GJRM` whose modularity allows for easy inclusion of potentially any parametric link marginal function and copula. The proposed approach makes a significant contribution in applied statistics as it is methodologically flexible, computationally sound and practically usable.

Although the literature in this area is reasonably ample, to the best of our knowledge, only Sun & Ding (2021a) provided a methodological framework together with software for modelling bivariate censored data. Unlike their cop-

ula approach, which allows the margins to be specified through semi-parametric transformation models, the baseline survival functions to be modelled using Bernstein polynomials and the dependence between events to be captured via one-parameter and two-parameter copulae, our proposal permits to specify all model parameters (including the dependence parameter) as flexible functions of covariate effects, model the baseline survival functions by means of monotonic P-splines which are theoretically and computationally advantageous, and conveniently characterise the marginals via links of the survival functions. Methodologically speaking, both approaches have been conceived to handle any combination of censoring mechanisms as well as have two different sets of regression coefficients for the marginal survival functions. However, from a computational point of view, the implementation provided by Sun & Ding (2021a) does not simultaneously support all possible bivariate combinations of censoring types and forces the two set of regression parameters to be the same.

Future research will focus on extending the approach to more than two event times (e.g., multi-morbidity) exploring, for instance, the use of multivariate Archimedean copulae, mixtures of powers, pair-copulae constructions, the multivariate Gaussian and Student's t distributions, and the composite likelihood approach (see, e.g., the supplementary material of Filippou et al., 2019, and references therein, which illustrates succinctly these ideas in a different context). Other potentially interesting extensions would be to account for informative and/or dependent censoring (e.g., Dettoni et al., 2020) as well as consider the case of excess hazard modelling (Eletti et al., 2022).

Acknowledgments

We would like to thank the Associate Editor and two anonymous reviewers for providing very helpful suggestions which improved and clarified the message of the article. The last two authors have been supported by the Engineering and Physical Sciences Research Council [EP/T033061/1].

References

Barthel, N., Geerdens, C., Killiches, M., Janssen, P., & Czado, C. (2018). Vine copula based likelihood estimation of dependence patterns in multivariate event time data. *Computational Statistics & Data Analysis*, 117, 109–127.

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.

Chen, M., Chen, L., Lin, K., & Tong, X. (2014). Analysis of multivariate interval censoring by diabetic retinopathy study. *Communications in Statistics-Simulation and Computation*, 43(7), 1825–1835.

Chen, M., Tong, X., & Sun, J. (2009). A frailty model approach for regression analysis of multivariate current status data. *Statistics in Medicine*, 28(27), 3424–3436.

Cook, R. & Tolusso, D. (2009). Second-order estimating equations for the analysis of clustered current status data. *Biostatistics*, 10(4), 756–772.

Dettoni, R., Marra, G., & Radice, R. (2020). Generalized link-based additive survival models with informative censoring. *Journal of Computational and Graphical Statistics*, 29(3), 503–512.

Eletti, A., Marra, G., Quaresma, M., Radice, R., & Rubio, F. (2022). A unifying framework for flexible excess hazard modeling with applications in cancer epidemiology. *Journal of the Royal Statistical Society Series C, in press*.

Emura, T., Sofeu, C. L., & Rondeau, V. (2021). Conditional copula models for correlated survival endpoints: Individual patient data metaanalysis of randomized controlled trials. *Statistical methods in medical research*, 12(30), 2634–2650.

Filippou, P., Kneib, T., Marra, G., & Radice, R. (2019). A trivariate additive regression model with arbitrary link functions and varying correlation matrix. *Journal of Statistical Planning and Inference*, 199, 236–248.

Geerdens, C., Acar, E. F., & Janssen, P. (2018). Conditional copula models for right-censored clustered event time data. *Biostatistics*, 19(2), 247–262.

Group, A. (1999). The age-related eye disease study (areds): Design implications. *AREDS report no. 1. Controlled Clinical Trials*, 20(6), 573–600.

Hu, T., Zhou, Q., & Sun, J. (2017). Regression analysis of bivariate current status data under the proportional hazards model. *Canadian Journal of Statistics*, 45(4), 410–424.

Kwon, S., Ha, I. D., Shih, J.-H., & Emura, T. (2021). Flexible parametric copula modeling approaches for clustered survival data. *Pharmaceutical Statistics*, 21(1), 69 – 88.

Leitenstorfer, F. & Tutz, G. (2006). Generalized monotonic regression based on B-splines with an application to air pollution data. *Biostatistics*, 8(3), 654–673.

Liu, X.-R., Pawitan, Y., & Clements, M. (2018). Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5), 1531–1546.

Lo, S. M. S., Mammen, E., & Wilke, R. A. (2020). A nested copula duration model for competing risks with multiple spells. *Computational Statistics & Data Analysis*, 150, 106986.

Marra, G. & Radice, R. (2020). Copula link-based additive models for right-censored event time data. *Journal of the American Statistical Association*, 115(530), 886–895.

Marra, G. & Radice, R. (2022). *GJRM: Generalised Joint Regression Modelling*. R package version 0.2-6.

Martins, A., Aerts, M., Hens, N., Wienke, A., & Abrams, S. (2019). Correlated gamma frailty models for bivariate survival time data. *Statistical Methods in Medical Research*, 28(10-11), 3437–3450.

Pya, N. & Wood, S. (2015). Shape constrained additive models. *Statistics and Computing*, 25(3), 543–559.

R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Reid, N. (1994). A conversation with sir david cox. *Statistical Science*, 9(3), 439–455.

Romeo, J., Meyer, R., & Gallardo, D. (2018). Bayesian bivariate survival snalysis using the power variance function copula. *Lifetime Data Analysis*, 24(2), 355–383.

Sujica, A. & Van Keilegom, I. (2018). The copula-graphic estimator in censored nonparametric location-scale regression models. *Econometrics and Statistics*, 7, 89–114.

Sun, T. & Ding, Y. (2021a). Copula-based semiparametric regression method for bivariate data under general interval censoring. *Biostatistics*, 22(2), 315–330.

Sun, T. & Ding, Y. (2021b). *CopulaCenR: Copula-Based Regression Models for Bivariate Censored Data*. R package version 1.1.3.

Swaroop, A., Chew, E., Rickman, C., & Abecasis, G. (2009). Unraveling a multifactorial late-onset disease: from genetic susceptibility to disease mechanisms for age-related macular degeneration. *Annual Review of Genomics and Human Genetics*, 10(1), 19–43.

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

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society. Series B*, 45(1), 133–150.

Wang, L., Sun, J., & Tong, X. (2008). Efficient estimation for the proportional hazards model with bivariate current status data. *Lifetime Data Analysis*, 14(2), 134–153.

Wang, N., Wang, L., & McMahan, C. (2015). Regression analysis of bivariate current status data under the gamma-frailty proportional hazards model using the em algorithm. *Computational Statistics & Data Analysis*, 83(C), 140–150.

Wen, C. & Chen, Y. (2013). A frailty model approach for regression analysis of bivariate interval-censored survival data. *Statistica Sinica*, 23(1), 383–408.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction With R*. Second Edition, Chapman & Hall/CRC, London.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111(516), 1548–1563.

Zeng, D., Gao, F., & Lin, D. (2017). Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika*, 104(3), 505–525.

Zhou, Q., Hu, T., & Sun, J. (2017). A sieve semiparametric maximum likelihood approach for regression analysis of bivariate interval-censored failure time data. *Journal of the American Statistical Association*, 112(518), 664–672.