Bivariate distributional copula regression for mixed non-time-to-event and time-to-event responses

Guillermo Briseño Sanchez[1] and Andreas Groll[2]

[1]Methods for Big Data, Scientific Computing Centre, Karlsruhe Institute of Technology, Karlsruhe, Germany
[2]Statistical Methods for Big Data, Department of Statistics, TU Dortmund University, Dortmund, Germany

Abstract: We propose a distributional copula regression modelling approach for bivariate responses comprised of non-commensurate (i.e. mixed) variables. In our case, the margins are a right-censored time-to-event outcome and a non-time-to-event variable. The underlying hazard rate of the time-to-event margin is modelled using discrete-time-to-event (DT) or piecewise-exponential (PW) methods. A flexible statistical model is achieved by relying on the correspondence of the likelihood of the aforementioned time-to-event approaches with well-known univariate distributions. We construct joint bivariate distributions for these mixed responses by means of parametric bivariate copulas. This allows for separate specification of the dependence structure between the margins and their individual distribution functions. All coefficients of the distributional copula regression models considered here are estimated simultaneously via penalized maximum likelihood. We showcase the versatility of our proposed approach in an analysis of red-light running behaviour of E-cyclists by modelling the joint distribution of a mixed response comprised of a binary response and a time-to-event outcome that indicates the time of red traffic light running.

1 Introduction

In many applications one may be interested in modelling the joint statistical behaviour of two random variables, such as Y and T, instead of just one of them. Separate analyses of the individual outcomes are prone to miss important characteristics of their interrelationship. Parametric bivariate distributions are attractive candidates since they provide an interpretable closed form of the joint behaviour of interest. For example, consider continuous variables Y and T quantifying children's malnutrition via stunting and wasting scores. In this case, the bivariate Gaussian distribution with

---

Address for correspondence: Guillermo Briseño Sanchez, Methods for Big Data, Scientific Computing Centre, Karlsruhe Institute of Technology,
Zirkel 2, 76131 Karlsruhe, Germany.
E-mail: guillermo.briseno-sanchez@kit.edu

correlation parameter becomes a natural choice, see Klein et al. (2014). For discrete Y and T, as in the number of goals in football matches, the bivariate Poisson distribution with strictly positive covariance parameter could be appropriate, see Karlis and Ntzoufras (2003) as well as Groll et al. (2018). While there exists a vast literature on bivariate distributions and their applications, see, for example, Lai and Balakrishnan (2009) and Kocherlakota and Kocherlakota (2017), cases where they lose their appeal occur regularly. If the margins follow univariate distribution families (e.g. Y Gaussian and T Gamma) and there is also interest in more complex dependence structures, finding a suitable bivariate distribution can become a challenging task. This problem is exacerbated when the marginal variables have different supports, for example, when Y is continuous and T is binary or discrete. Such non-commensurate or mixed outcomes are ubiquitous in health and medicine related applications (de Leon and Wu, 2011), but they are likely to be prevalent in other areas of research too. For example in transportation (Spissu et al., 2009), social sciences (Wagner and Tüchler, 2013), economics (Hohberg et al., 2021) and environmental sciences (Najib et al., 2022). Last, if one of the margins is subject to censoring (i.e. the actual response realisation is not observed), then the aforementioned bivariate distributions cease to be suitable tools for constructing a statistical model. To sidestep these limitations, we resort to parametric bivariate copulas to construct joint bivariate distributions with arbitrary (but suitable) margins and a separate specification of their dependence structure. Therefore, we aim to provide a tool to analyse complex phenomena via a single statistical model of the multiple dimensions that constitute it. Our main interest is to model jointly bivariate outcomes that are non-commensurable or mixed, that is Y and T which have different supports and in addition, the latter is subject to censoring.

Without loss of generality, for the remainder of this article we denote the first margin Y as *non-time-to-event* and the second margin T as *time-to-event* variables, respectively. We propose to model the underlying hazard rate of the time-to-event margin T using discrete time-to-event (DT, see Tutz and Schmid, 2016) and piecewise-exponential (PW, Bender et al., 2018) methods within a distributional copula regression framework, such that a highly flexible statistical model can be fitted to the response vector. This is achieved by exploiting the likelihood of the aforementioned time-to-event approaches, which correspond to well-known likelihood functions of univariate Bernoulli (DT) and Poisson (PW) random variables, respectively.

Statistical models that are most closely related to this approach are *joint models* for longitudinal and time-to-event outcomes (Rizopoulos, 2012) and *landmarking* models (Suresh et al., 2019). On one hand, joint models aim to study the distribution of a longitudinal marker and a (potentially censored) time-to-event variable. On the other hand, landmarking does not model the entire (longitudinal) history of the non-time-to-event margin and instead conditions the joint distribution on subjects still not having experienced the event in a current observation time period. The two aforementioned methods differ from our contribution in some aspects. One concerns the non-time-to-event response in the sense that we do not work with longitudinal data. Another aspect is that for joint models the longitudinal margin, broadly speaking, receives most of the attention, whereas our focus is equally on both time-to-event and non-time-to-event margins. Nevertheless, our contribution could be employed for example in landmarking in the presence of markers that may be continuous, binary or discrete, see Suresh et al. (2019) and Suresh et al. (2021) for continuous and binary markers, respectively. Most recently, the likelihood of DT models has been used in joint models with a discrete-time component (Medina-Olivares et al., 2023), whereas Rappl et al. (2023) used the likelihood of a PW approach for joint models as well. Related work in the realm of bivariate distributional copula regression models is Marra and Radice (2017, bivariate continuous margins).

Moreover, for mixed outcomes see Klein et al. (2018, binary and continuous margins) and Marra et al. (2020, binary and discrete margins), respectively.

The remainder of this article is structured as follows: Section 2 reviews bivariate distributional copula regression and the considered time-to-event analysis methods. Our contribution is presented in Section 3. In Section 4, we apply our proposed method to analyse red-light running behaviour of E-cyclists using data from China (Gao et al., 2020). This is accomplished by modelling the joint distribution of a bivariate mixed outcome that consists of a binary non-time-to-event and a time-to-event margin, as well as their dependence. Last, a discussion is given in Section 5. Supplementary material containing methodological details, additional diagnostics of the application and simulation studies can be found at https://journals.sagepub.com/home/smja.

2 Methodology

In this section we introduce distributional regression for cases in which the joint distribution of a bivariate response is constructed by means of parametric copulas. Afterwards, we present discrete time and PW time-to-event models from a distributional regression perspective. The methods shown here apply for the i-th observation in a sample of size n, that is, $i = 1, \ldots, n$. For clarity, this index is omitted to avoid clutter in the notation.

2.1 Bivariate distributional copula regression

The response is assumed to be an independent random draw from a parametric distribution. Our goal is to construct a model for the joint probability $P(Y \leq y, T > t)$ of the non-time-to-event variable and the time-to-event margin (occurring after t). In the context of bivariate outcomes with distributions constructed by means of copulas, this may be written as

$$P(Y \leq y,\ T > t \mid \boldsymbol{\vartheta}) = C\left(F_1(y;\ \boldsymbol{\vartheta}^{(1)}),\ S_2(t;\ \boldsymbol{\vartheta}^{(2)}) \mid \vartheta^{(c)}\right), \tag{2.1}$$

where $C : (0, 1)^2 \to (0, 1)$ is a parametric copula function that connects or binds the margins, but does not depend on them (Nelsen, 2006), $F_1(y; \cdot)$ denotes the cumulative distribution function (CDF) of Y and $S_2(t; \cdot)$ indicates the survival function of T. We do not assume that the true copula function is necessarily known in advance. The scalar quantity $\vartheta^{(c)}$, commonly known as 'dependence parameter', determines the dependence structure between the margins. The range of the dependence parameter typically depends on the assumed or chosen copula function. However, it is possible to express the dependence between the margins in terms of Kendall's τ based on values of $\vartheta^{(c)}$, see Table D1 in Supplementary Material D for the corresponding transformations. The margin-specific parameter vectors $\boldsymbol{\vartheta}^{(\bullet)} = (\vartheta_1^{(\bullet)}, \ldots, \vartheta_{K_{(\bullet)}}^{(\bullet)})^\top$ contain the $k = 1, \ldots, K_{(\bullet)}$, with $\bullet \in \{1, 2\}$, parameters that completely specify the distribution of the margins. In turn, the bivariate distribution depends on the vector $\boldsymbol{\vartheta} = \left(\left(\boldsymbol{\vartheta}^{(1)}\right)^\top, \left(\boldsymbol{\vartheta}^{(2)}\right)^\top, \vartheta^{(c)}\right)^\top$, which contains all parameters of the margins and the dependence parameter. Following the Generalized Additive Models for Location, Scale and Shape (GAMLSS; Stasinopoulos et al., 2017) approach, potentially all parameters of the joint distribution are modelled as functions of the covariates at hand $\mathbf{x} = (x_1, x_2, \ldots, x_P)^\top \in \mathbb{R}^P$ by means

of a structured additive predictor $\eta_k^{(\bullet)}$ in the following fashion:

$$\vartheta_k^{(\bullet)} = g_{(\bullet)k}^{-1}\left(\eta_k^{(\bullet)}\right) \;\Leftrightarrow\; g_{(\bullet)k}\left(\vartheta_k^{(\bullet)}\right) = \eta_k^{(\bullet)} = \beta_{0k}^{(\bullet)} + \sum_{r=1}^{P_k^{(\bullet)}} s_r^{(\bullet)}(x_r), \;\bullet \in \{1, 2, c\}, \qquad (2.2)$$

where $g^{-1}(\cdot)$ is a response function with corresponding inverse or 'link' $g(\cdot)$, guaranteeing that the distribution parameters lie in their respective range. The range of the dependence parameter $\vartheta^{(c)}$ depends on the copula. The summation limit $P_k^{(\bullet)}$ emphasizes that the individual parameters must not be necessarily modelled using the same subset of covariates. The coefficient $\beta_{0k}^{(\bullet)}$ denotes a parameter-specific intercept and $s_r^{(\bullet)}(\cdot)$ are smooth functions that can accommodate a wide range of functional forms of the covariates such as linear, non-linear (e.g. P-splines, Eilers and Marx, 1996), random effects or spatial. The smooth functions can be represented as linear combinations of basis functions and unknown regression coefficients to be estimated: $\eta_k^{(\bullet)} = \mathbf{Z}_k^{(\bullet)}\boldsymbol{\beta}_k^{(\bullet)}$, where $\mathbf{Z}_k^{(\bullet)}$ is a design matrix of suitable dimensions, $\boldsymbol{\beta}_k^{(\bullet)}$ is a vector of coefficients and $\bullet \in \{1, 2, c\}$. Hence, the model ultimately depends on the vector $\boldsymbol{\beta} = \left(\left(\boldsymbol{\beta}^{(1)}\right)^\top, \left(\boldsymbol{\beta}^{(2)}\right)^\top, \left(\boldsymbol{\beta}^{(c)}\right)^\top\right)^\top$, which contains the coefficients corresponding to the models of the margins' parameters and the copula dependence parameter, respectively. Note that for some covariate effects the coefficients may be associated with a quadratic penalty and a smoothing parameter to enforce smoothness of the estimated functions, see Wood (2017) for more details. Distributional copula regression models as in Equations (2.1)–(2.2) can be fitted using either penalized maximum likelihood (Marra and Radice, 2017) (as in our case), fully Bayesian inference (Umlauf and Kneib, 2018) or boosting (Hans et al., 2022). Here we estimate all unknown coefficients in $\boldsymbol{\beta}$ simultaneously using penalized maximum likelihood and base our software implementation **entirely** on the R package Generalised Joint Regression Modelling (GJRM) (Marra and Radice, 2023).

2.2 Considered time-to-event analysis approaches

Recall that T denotes the event time and let T^{cens} denote a random censoring time. In the case of right-censored data one observes $\tilde{T} = \min\{T, T^{cens}\}$ and its censoring indicator $\delta = 1\{T \leq T^{cens}\}$. We assume that the censoring time T^{cens} is non-informative and independent of the event time T. Note that it is possible to relax the independent censoring assumption and allow for dependence between T and T^{cens}, although this will result in a more complex model than the one we present here, see Czado and Van Keilegom (2022) for more details. In this article, we consider DT and PW approaches to model the underlying hazard function of the right-censored time-to-event variable. Let $\mathcal{I}_j$ with $j = 1, \ldots, J$ denote an interval of finite length with respective left and right bounds κ_{j-1}, κ_j and $\kappa_0 = 0 < \kappa_1 < \cdots < \kappa_J < \infty$. In the DT framework, T takes values in the set $\{1, 2, \ldots, J\}$. One may assume that the process generating the observed times is truly discrete or that there are J intervals $\mathcal{I}_j$ that 'cut' or discretize the underlying continuous follow-up. The conditional discrete hazard is defined as the following conditional probability (Tutz and Schmid, 2016): $\lambda(t \mid \mathbf{x}) = P(T = t \mid T \geq t, \mathbf{x})$.

When working with a PW approach, see Holford (1980) and Laird and Olivier (1981), the follow-up is decomposed into J intervals. The hazard rate, defined in this case as $\lambda(t \mid \mathbf{x}) =$

$\lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \le T < t + \Delta t \mid T \ge t, \mathbf{x})$, is assumed to be constant within each interval. An illustration of the aforementioned discretized follow-up and hazard can be found in Figure 1 of Bender et al. (2018). To achieve correspondence with the Bernoulli or the Poisson log-likelihoods, we require an augmentation of the original dataset into a 'long' format. See Bender et al. (2018) (Table 1 and Table 2) as well as Berger and Schmid (2018) (Equations [3.9] and [3.10]) for an illustration of the augmented data for PW and DT approaches, respectively. In practice, we transform the data using the R packages discSurv (Welchowski et al., 2022 DT) and pammtools (Bender and Scheipl, 2018, PW). This produces the following auxiliary variable:

$$\delta_j = \begin{cases} 1, & \text{if subject has an event in } \mathcal{I}_j, \\ 0, & \text{else,} \end{cases} \quad \text{with} \quad \mathcal{I}_j = \begin{cases} [\kappa_{j-1}, \ \kappa_j) & \text{if DT,} \\ (\kappa_{j-1}, \ \kappa_j] & \text{if PW,} \end{cases}$$

Using the augmented data, the log-likelihood of a DT model coincides with the Bernoulli log-likelihood (Tutz and Schmid, 2016). The discrete hazard is then $\lambda(j \mid \mathbf{x}) = \vartheta_j = g^{-1}(\eta_j)$, where $\vartheta_j \in [0, 1]$ is the parameter of the Bernoulli distribution. If one follows a PW approach, the log-likelihood contribution coincides with the Poisson log-likelihood with offset $\ln(t_j) = o_j$, where t_j is the time spent in the j-th interval (Bender et al., 2018). In this case, the hazard is $\lambda(j \mid \mathbf{x}) = \exp(\eta_j) = \vartheta_j / \exp(o_j)$, where $\vartheta_j = \exp(\eta_j) \exp(o_j)$ is the parameter of the Poisson distribution. Note that both DT and PW approaches can accommodate time-dependent covariates, that is, $\mathbf{x}_j$. For either approach, a structured additive predictor is set as foundation for a regression model for the hazard rate:

$$\lambda(t \mid \mathbf{x}) = g^{-1} \left(\underbrace{s_0(t)}_{\text{link-scale baseline hazard}} + \sum_{r=1}^{P} s_r(t, x_r) \right) \quad \forall\, t \in \mathcal{I}_j,$$

where the baseline hazard is modelled as a smooth function of time and the functional form of the covariates can be any of those described in Section 2.1. For PW models $g(\cdot)$ is the natural logarithm function. In the DT framework, $g(\cdot)$ can be any suitable link function for for parameters with range $[0, 1]$, such as *logit*, *probit* and *clog-log*. We remark that our implementation currently features only the clog-log link but we plan to add others in the future. From a practical perspective, the clog-log link produces the 'grouped proportional hazards model', which can be seen as a discretized version of Cox's proportional hazards model (Tutz and Schmid, 2016).

3 Embedding DT and PW approaches into distributional copula regression

Using the DT or PW approach together with the augmented data, the i-th observation of the time-to-event margin is now represented by $j = 1, \ldots, j(i)$ auxiliary variables, where $j(i)$ denotes the length of the sequence emanating from the i-th observational unit in the sample. We denote these as $\boldsymbol{\delta} = \left(\delta_1, \ldots, \delta_{j(i)} \right)^{\top}$, which can have two possible expressions: $\boldsymbol{\delta} = 0_{j(i)}^{\top}$ in case of censoring or $\boldsymbol{\delta} = \left(0_{j(i)-1}^{\top}, \ 1 \right)$ in case of an event, that is, a non-censored observation, where $0_{j(i)}^{\top}$ and $0_{j(i)-1}^{\top}$ denote zero vectors of length $j(i)$ and $j(i) - 1$, respectively.

We propose to use the following function based on the likelihood contributions of either DT or PW techniques:

$$F(\boldsymbol{\delta}) = \begin{cases} f\left(0_{j(i)}\right), & \text{if } \boldsymbol{\delta} = 0_{j(i)}^{\top} \\ f\left(0_{j(i)}\right) + f\left(\left(0_{j(i)-1}^{\top},\ 1\right)^{\top}\right), & \text{if } \boldsymbol{\delta} = \left(0_{j(i)-1}^{\top},\ 1\right), \end{cases} \tag{3.1}$$

where $f(\boldsymbol{\delta})$ is the likelihood contribution for a right-censored or an observed time, respectively. Here we cumulate over the expressions $f\left(0_{j(i)}\right)$ and $f\left(\left(0_{j(i)-1}^{\top},\ 1\right)^{\top}\right)$ in case of an event, that is, $\boldsymbol{\delta} = (0_{j(i)-1}^{\top},\ 1)$, to account for the auxiliary variables $\boldsymbol{\delta}$ being discrete. The concrete expression of $f(\boldsymbol{\delta})$ depends on the adopted approach. For instance, choosing a DT model and setting the link function to the clog-log, that is, $\lambda(j|\mathbf{x}) = \vartheta_j = 1 - \exp(-\exp(\eta_j))$, results in the following function based on the Bernoulli likelihood:

$$f(\boldsymbol{\delta}) = \begin{cases} \exp\left(-\sum_{j=1}^{j(i)} \exp(\eta_j)\right), & \text{if } \boldsymbol{\delta} = 0_{j(i)}^{\top}, \\ \exp\left(-\sum_{j=1}^{j(i)-1} \exp(\eta_j)\right) - \exp\left(-\sum_{j=1}^{j(i)} \exp(\eta_j)\right), & \text{if } \boldsymbol{\delta} = \left(0_{j(i)-1}^{\top},\ 1\right). \end{cases} \tag{3.2}$$

Using a PW approach, the conditional hazard corresponds to $\lambda(j|\mathbf{x}) = \exp(\eta_j) = \vartheta_j/\exp(o_j)$, leading to the following function based on the Poisson likelihood:

$$f(\boldsymbol{\delta}) = \begin{cases} \exp\left(-\sum_{j=1}^{j(i)} \vartheta_j\right), & \text{if } \boldsymbol{\delta} = 0_{j(i)}^{\top}, \\ \exp\left(-\sum_{j=1}^{j(i)} \vartheta_j\right) \vartheta_{j(i)}, & \text{if } \boldsymbol{\delta} = \left(0_{j(i)-1}^{\top},\ 1\right). \end{cases} \tag{3.3}$$

We provide analytical first and second order partial derivatives of the proposed functions $f(\boldsymbol{\delta})$ and $F(\boldsymbol{\delta})$ based on DT and PW approaches with respect to a generic coefficient vector $\boldsymbol{\beta}$ in Supplementary Material B. These are required for the simultaneous estimation algorithm of our software implementation, which is based entirely on GJRM (Marra and Radice, 2023).

3.1 Bivariate additive copula regression for mixed non-time-to-event and time-to-event responses

Recall that the bivariate non-commensurable outcome of interest is comprised of a non-time-to-event and a potentially right-censored time-to-event variable. We now replace the survival function of the potentially right-censored time-to-event margin with the proposed function $F(\boldsymbol{\delta})$ in Equation (3.1). Note that in this case Sklar's theorem is no longer used in its original form. We focus on cases when the non-time-to-event margin is binary, therefore the

log-likelihood of the *i*-th observation has the following form (Marra et al., 2020):

$$\ell(\boldsymbol{\beta}) = (1-y)\ln\left[C\Big(F_1(0), F(\boldsymbol{\delta}); \vartheta^{(c)}\Big) - C\Big(F_1(0), F(\boldsymbol{\delta}) - f(\boldsymbol{\delta}); \vartheta^{(c)}\Big)\right]$$
$$+ y\ln\left[f(\boldsymbol{\delta}) - C\Big(F_1(0), F(\boldsymbol{\delta}); \vartheta^{(c)}\Big) + C\Big(F_1(0), F(\boldsymbol{\delta}) - f(\boldsymbol{\delta}); \vartheta^{(c)}\Big)\right], \qquad (3.4)$$

where the copula $C(\cdot)$ is evaluated using the CDF of y at zero, that is, $F_1(0)$ and $F(\boldsymbol{\delta})$.

Considerations for copula modelling: A crucial aspect of copula modelling is the fact that the dependence structure between the margins is preserved if a monotonic transformation is applied to them (Klement et al., 2002; Nelsen, 2006). If the proposed function $F(\boldsymbol{\delta})$ is a monotonic decreasing function of the time-to-event variable T, then the original dependence structure will be preserved and the use of our proposed approach would be justified. Otherwise, $F(\boldsymbol{\delta})$ will alter the original dependence structure, rendering our proposal not suitable for copula modelling. Supplementary Material A contains a proof that the proposed function $F(\boldsymbol{\delta})$ is a monotonic decreasing function of the original variable for both DT and PW approaches. Hence, replacing the survival function $S_2(t)$ with $F(\boldsymbol{\delta})$ in a bivariate copula model preserves the original dependence structure of interest. Additionally, we have conducted various simulation studies to assess the performance of the proposed approach using DT and PW likelihoods, these can be found in Supplementary Material D, available at https://journals.sagepub.com/home/smja. The R code used to conduct the simulations can be found in the following repository: https://github.com/GuilleBriseno/DTPW_DistCopulaReg.

4   Analysis of red-light running behaviour of E-cyclists

We study the behaviour of E-bike cyclists (E-cyclists) regarding traffic red-light running or traffic red light violation using a sample of $n = 2\,173$ subjects from Shanghai, China collected and previously analyzed by Gao et al. (2020). The margins of the bivariate mixed outcome consist of a binary non-time-to-event and a time-to-event variable, respectively. The binary response VisualSearch indicates whether a cyclist quickly turns their head in both directions while waiting at the crossing (0 = no, 1 = yes). This variable is considered a relevant factor in describing cyclists' red-light running and driving behaviour (Fraboni et al., 2018; Rupi and Krizek, 2019). The time-to-event variable WaitingTime gives the time in seconds until an E-cyclist crosses an intersection when the traffic light is red. The censoring indicator RedLight is equal to one if an E-cyclist runs the red-light and zero if they wait until the traffic light turns green. This means that those E-cyclists that arrived at the intersection during a red-light and only cross it until the signal turns green are treated as censored observations (Gao et al., 2020). This yields a censoring rate of 47.78%. It could be argued that those individuals that do not run the red traffic light in the data, that is, censored observations, will never commit such an infringement and should not be analyzed jointly with those that do so. However, we believe it is more reasonable to assume that any E-cyclist is capable of running a red traffic light *under certain circumstances*. Hence, for the censored observations in the sample, the conditions for violating the red traffic light were simply not met at the time of collecting the data and we assume that under different conditions such cyclists would cross an intersection given a red-light.

Table 1 Description of the variables in the data. Summary column shows median survival time for WaitingTime and percentages of the remaining variables.

Variable	Description	Summary
WaitingTime	Time (in seconds) waiting at crossing for green traffic light.	79
RedLight	Red-light running (0 = no, 1 = yes).	47.8/52.2
VisualSearch	Cyclist turns head quickly in both directions (0 = no, 1 = yes).	41.1/58.9
Gender	0 = Female, 1 = Male.	29.8/70.2
RiderType	Type of cyclist riding the E-bike (0 = Normal, 1 = Takeaway).	51.9/48.1
GroupSize	Number of cyclists waiting at intersection at arrival of cyclist. ($\in$ {'none', '1 to 4' and '5 and more'}).	33.5/55.4/11.1
ConformBehaviour	Another cyclist runs the red light (0 = no, 1 = yes).	50.3/49.7
Position	Position at crossing where cyclist waits. ($\in$ { 'behind stop line', 'middle', 'close to motorized lane' }).	29.5/34.4/36.1
ComingDirection	Direction from where the cyclist comes into intersection. ($\in$ {'through', 'left', 'right'}).	63.6/20.6/15.7

Sample size: $n = 2\,173$. Censoring rate: 47.78%.

Table 2 Akaike and Bayesian information criteria of copula models with clog-log link for VisualSearch as well as independent margins. Bold numbers indicate the best-fitting model.

Copula	DT		PW	
	AIC	BIC	AIC	BIC
Independence	9 266.31	9 479.82	19 544.90	19 730.12
Gaussian	9 159.60	9 342.00	19 825.64	20 014.59
Frank	9 157.01	9 339.28	19 442.84	19 625.98
Clayton	**9 155.67**	**9 338.02**	**19 441.87**	**19 625.11**
Clayton (90°)	9 272.28	9 454.43	19 550.75	19 733.74
Gumbel	9 166.21	9 348.63	19 491.04	19 674.13
Gumbel (90°)	9 272.12	9 454.26	19 548.36	19 731.35

The covariates in the data describe three aspects of the E-cyclists: Individual characteristics, social influence and cycling information (Gao et al., 2020). Among the individual characteristics are the gender (female = 0, male = 1) of the cyclist and the factor variable RiderType which indicates whether the cyclist is a takeaway delivery driver or not. The social influence aspect is represented by two variables.

The categorical variable GroupSize, which says how many other cyclists were waiting at the intersection at the arrival of the subject ('none', '1 to 4' and '5 and more'), and ConformBehaviour which is equal to one if another cyclist runs the red light and zero otherwise. Social influence characteristics have been found to play an important role in decreasing the propensity of red-light running (Bai and Sze, 2020). Last the cycling information is described by the categorical variables Position and ComingDirection. Previous studies have found that Position plays an important role in determining driving behaviour (Tang et al., 2020). See Table 1 for a description of the variables.

Model configuration: We aim to model the joint distribution of VisualSearch and WaitingTime using our proposed distributional copula regression approach. The binary response VisualSearch is modelled using a Bernoulli distribution with parameter $\vartheta^{(1)} \in [0, 1]$. The hazard rate of the time-to-event variable WaitingTime is modelled using the proposed functions based on the

auxiliary variables $\boldsymbol{\delta}$ using DT and PW approaches:

$$(\textsf{VisualSearch},\ \textsf{WaitingTime})^{\top} \sim C\left(F_1(\textsf{VisualSearch};\ \vartheta^{(1)}),\ F(\boldsymbol{\delta};\ \boldsymbol{\vartheta}) \mid \vartheta^{(c)}\right),$$

where $F_1(\cdot)$ is the CDF of the Bernoulli distribution and $F(\cdot)$ is the proposed function from Equation (3.1) with $\boldsymbol{\vartheta}$ denoting the hazard rate built using either DT (clog-log link) or PW (natural logarithm link) methods. We omit the index k in the distribution parameters for the remainder of the article due to both margins depending only on one parameter. To construct the model of the hazard for $\textsf{WaitingTime}$, the follow-up is split using 20 intervals, resulting in intervals that are 6.5 seconds of length. We investigated alternative configurations with 130, 65, 35 and 13 intervals, but these did not change the results in a relevant way. The structured additive predictors of the model are:

$$\eta^{(1)} = \beta_0^{(1)} + \beta_1^{(1)}\textsf{Gender} + \beta_2^{(1)}\textsf{RiderType} + \beta_3^{(1)}\textsf{Position},$$
$$\eta^{(2)} = s_0^{(2)}(t,\ \textsf{RiderType}) + \beta_1^{(2)}\textsf{Gender} + \beta_2^{(2)}\textsf{RiderType} + \beta_3^{(2)}\textsf{Position} + \beta_4^{(2)}\textsf{GroupSize} +$$
$$\beta_5^{(2)}\textsf{ConformBehaviour},$$
$$\eta^{(c)} = \beta_0^{(c)} + \beta_1^{(c)}\textsf{RiderType} + \beta_2^{(c)}\textsf{ConformBehaviour},$$

where the link-scale baseline hazard $s_0^{(2)}(\cdot)$ is modelled as a smooth function of time using P-splines with second order differences and is stratified depending on $\textsf{RiderType}$. Hence, we fit a separate baseline hazard for $\textsf{Normal}$ and $\textsf{TakeAway}$ E-cyclists. The remaining categorical covariates enter the model as linear effects with dummy coding. The dependence between the margins modelled via $\eta^{(c)}$ is allowed to change based on $\textsf{RiderType}$ and $\textsf{ConformBehaviour}$. The variable $\textsf{ComingDirection}$ is not included in the model due having no significant effect on any additive predictor. Note that the covariates were selected into the different additive predictors based on hypothesis testing (Marra and Radice, 2017). The copula function is selected out of the Gaussian, Frank, Clayton and Gumbel (with 90° rotations) copulas by means of the Akaike and Bayesian Information Criteria (AIC, BIC), see Table 2. We remark that $\textsf{GJRM}$ supports more copula functions aside from the aforementioned, however we do not consider the entire catalogue in this article. All model coefficients are estimated simultaneously. Confidence intervals are obtained as described in Radice et al. (2016). The goodness-of-fit of $\boldsymbol{\delta}$ is assessed using Cox–Snell residuals (Klein and Moeschberger, 2003) in Figure 2. Additional diagnostics of the fitted copula function can be found in Supplementary Material C2. The $\textsf{R}$ code used to fit the models to the data can be found in the following repository: $\textsf{https://github.com/GuilleBriseno/DTPW\_DistCopulaReg}$.

Results using DT approach: The best fitting model is a Clayton copula with clog-log link for $\textsf{VisualSearch}$, the estimated coefficients are shown in the DT column of Table 3. The column corresponding to the DT models in Table 2 shows that the Clayton copula fits the data better compared to using independent margins. We find that on average there exists a significant 'moderate' positive dependence between $\textsf{VisualSearch}$ and $\textsf{WaitingTime}$ with Kendall's τ estimated at $\hat{\tau} = 0.236\ [0.181,\ 0.311]$ with $\alpha = 0.05$. The additive predictor of the dependence parameter $\eta^{(c)}$ decreases for takeaway relative to normal cyclists. A less pronounced decrease in the predictor $\eta^{(c)}$ is also observed when another cyclist at the intersection commits a red traffic light violation ($\textsf{ConformBehaviour} = 1$). These estimates indicate that the behaviour of E-cyclists is influenced by social or group factors while waiting at an intersection. Table 4 shows

Table 3 Estimated coefficients with 95% confidence intervals of the Clayton copula models using DT and PW approaches. All coefficients are estimated simultaneously.

	DT		PW	
Binary margin: VisualSearch				
$\beta_0^{(1)}$	−0.675	[−0.827; −0.528]	−0.674	[−0.819; −0.529]
GenderMale	0.162	[0.023; 0.304]	0.164	[0.017; 0.309]
RiderTypeTakeAway	0.510	[0.386; 0.634]	0.504	[0.376; 0.628]
PositionMiddle	0.256	[0.109; 0.404]	0.257	[0.110; 0.400]
PositionCloseMotorLane	0.291	[0.146; 0.436]	0.297	[0.151; 0.442]
Time-to-event margin: WaitingTime				
GenderMale	0.300	[0.134; 0.464]	0.290	[0.125; 0.454]
RiderTypeTakeAway	1.460	[1.294; 1.626]	1.496	[1.329; 1.663]
PositionMiddle	0.428	[0.258; 0.601]	0.420	[0.251; 0.589]
PositionCloseMotorLane	0.688	[0.525; 0.853]	0.683	[0.521; 0.845]
GroupSize1to4	−0.332	[−0.452; −0.215]	−0.356	[−0.476; −0.237]
GroupSize5AndMore	−0.758	[−0.996; −0.513]	−0.811	[−1.049; −0.580]
ConformBehaviourYes	0.276	[0.156; 0.397]	0.259	[0.136; 0.381]
Dependence				
$\beta_0^{(c)}$	0.546	[0.222; 0.873]	0.495	[0.160; 0.824]
RiderTypeTakeAway	−2.137	[−2.950; −1.315]	−2.560	[−3.576; −1.547]
ConformBehaviourYes	−0.711	[−1.244; −0.182]	−0.719	[−1.261; −0.160]
Kendall's $\hat{\tau}$	0.236	[0.181; 0.311]	0.217	[0.162; 0.292]

Sample size: $n = 2{,}173$. Censoring rate: 47.78%.

the values of the estimated Kendall's τ for the combinations of the binary variables RiderType and ConformBehaviour. It can be seen that the dependence between the margins decreases for takeaway cyclists whenever there is conforming behaviour. Comparatively, the estimated dependence between VisualSearch and ConformBehaviour is much stronger for normal cyclists and when no other driver crosses the junction. Regarding the model for VisualSearch, takeaway cyclists have on average a significantly higher probability of checking for traffic in both directions relative to normal cyclists, *ceteris paribus*. On average, male cyclists also have a significantly higher change of conducting VisualSearch relative to female cyclists, *ceteris paribus*. Figure 1 shows the estimated hazard rate and corresponding estimated survival function from the Clayton model based on different values of GroupSize and RiderType. Overall, a noticeable shift upwards in the hazard rate can be seen for takeaway cyclists. Conversely, the estimated survival function for takeaway cyclists lies below the one corresponding to normal cyclists. This phenomenon depicts the riskier behaviour that delivery-service cyclists exhibit to meet demand in the delivery-service industry (Gao et al., 2020; Zhang and Liu, 2023).

The GroupSize covariate describes the social behaviour of the cyclist while waiting for green traffic light. The magnitude of the estimated effect on the link-scale hazard rate increases based on the level of GroupSize. Hence, the more individuals waiting at the crossing or the more crowded the crossing becomes, the less likely that the cyclist will violate the traffic red-light. Figure 1 column (a) depicts the estimated hazard rates and survival functions based on the DT approach. The effect of GroupSize can be seen to produce a shift downwards in the estimated hazards and upwards in the estimated survival functions. The estimated survival functions shown in Figure 1
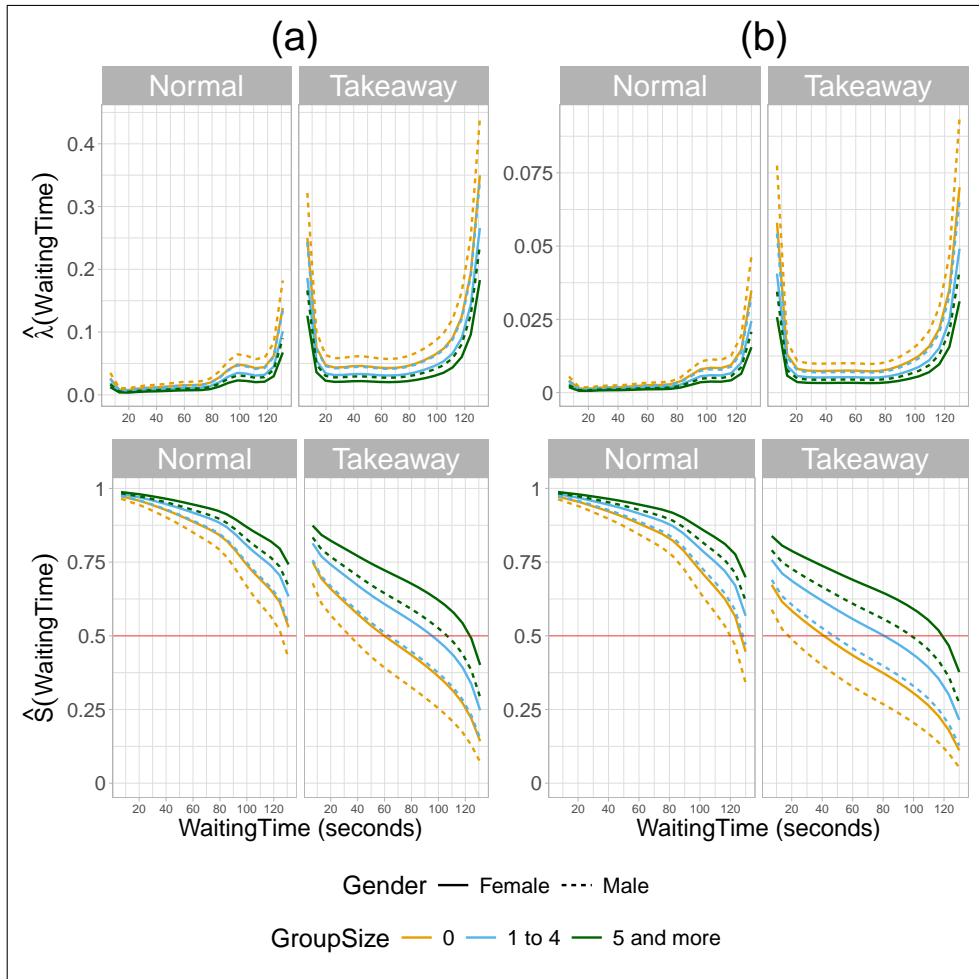
Figure 1 Estimated hazard rate and survival function across RiderType, GroupSize and Gender for DT (a) and PW (b) approaches. Red line indicates the median.

column (a) exhibit different patters for normal and takeaway cyclists, respectively. It can be seen that the median survival time is dramatically lower for takeaway cyclists. The estimated hazard rate shown in Figure 1 column (a) shows that the hazard for takeaway E-cyclists has a 'U' or bathtub shape. This means that the risk of violating the red traffic light is higher at the cyclist's arrival at the intersection as well as close to the end of the red signal. The Cox–Snell residuals shown in Figure 2 (a) corresponding to the DT approach indicate that the fitted model with a stratified baseline hazard for RiderType fits the data well. Table C1 and Figure C1 in Supplementary Material C show the estimated coefficients as well as the estimated hazard and survival functions, respectively, when assuming independence between the margins. Overall, the estimated coefficients from univariate independent models exhibit some differences in their magnitude, but a similar

Table 4 Estimated dependence expressed as Kendall's τ for the combination of RiderType and ConformBehaviour obtained using DT and PW approaches with 95% confidence intervals shown underneath the estimates.

RiderType ConformBehaviour	DT		PW	
	Normal	TakeAway	Normal	TakeAway
No	0.463	0.092	0.451	0.060
	[0.382; 0.545]	[0.042; 0.186]	[0.370; 0.532]	[0.022; 0.149]
Yes	0.298	0.048	0.286	0.030
	[0.213; 0.402]	[0.023; 0.097]	[0.203; 0.386]	[0.012; 0.074]

trend can be seen for all of the considered covariates. The diagnostics regarding the fitted Clayton copula function in Supplementary Material C indicate that although the Clayton copula performs best among the considered parametric copula families, asymmetry can be seen in the joint tail dependence. Figure C2 in Supplementary Material C indicates that a parametric copula family likely struggles to adequately model the dependence between the margins. Histograms of the margins $F_1(\mathsf{VisualSearch})$ and $F(\delta)$ shown in Figure C4 (a) show that the margin $F(\delta)$ does not resemble a uniform distribution, indicating that the fit of the lower tail of the aforementioned margin could be improved.

Results using PW approach: According to Table 2, the best-fitting copula is once again the Clayton copula. In addition, the column PW in Table 2 shows that the Clayton copula model delivers a better fit compared to fitting independent univariate margins. The estimated coefficients for the model based on the PW approach closely resemble those of the DT approach. The 95% confidence intervals shown in Table 3 are also very similar for both approaches. The estimated dependence quantified as Kendall's τ is also close to the DT approach-based model, sitting at $\hat{\tau} = 0.217 \in [0.162; 0.292]$. See Table 4 for the estimated values of Kendall's τ for every combination of RiderType and ConformBehaviour. Once again, the estimated values resemble that of the model based on the DT approach. The estimated hazard rates and survival functions shown in Figure 1 (b) exhibit the same pattern as those based on the DT approach. The Cox-Snell residuals from Figure 2 (b) also suggest that the PW-based model fits the data well apart from some outliers in the upper tail of the data. The results from fitting independent univariate models shown in Table C1 and Figure C1 in Supplementary Material C exhibit some changes in the magnitude of the estimated coefficients. Similar trends to those found in the copula models can be observed. The diagnostics regarding the Clayton copula from Supplementary Material C point towards similar conclusions as those drawn from the DT approach, where asymmetry can be seen in the joint tail dependence. In addition, Figure C4 (b) in Supplementary Material C shows that the fit of lower tail of the margin $F(\delta)$ can be improved. One possible solution to accommodate the aforementioned asymmetry while keeping a parametric model for the dependence would be to implement Khoudraji's device (Genest et al., 1998). This approach modifies parametric copula functions by introducing one or two scalar asymmetry parameters. Another option would be to pursue a non-parametric approach, such as using kernel density-based bivariate copulas, as proposed in Nagler and Czado (2016). This approach would have the advantage of being more flexible than the parametric copula families considered so far. However, these are outside of the scope of this article.
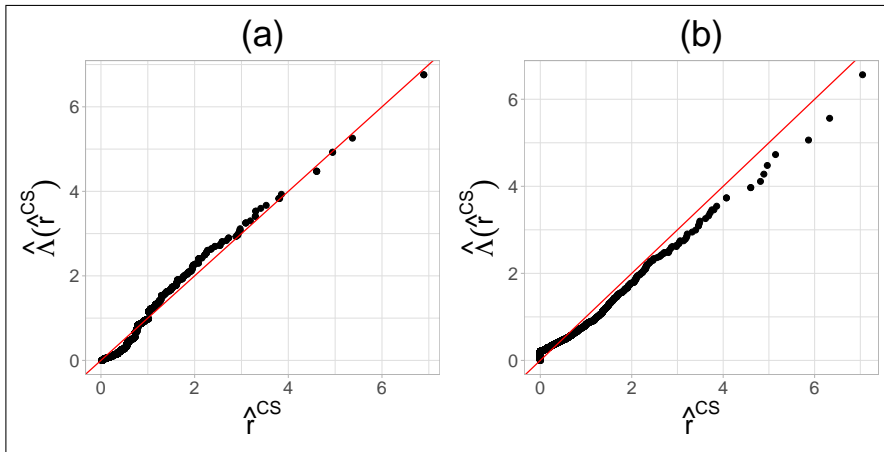
Figure 2 Cox-Snell residuals corresponding to the model of WaitingTime using the DT (a) and PW (b) approaches.

5 Discussion

We have proposed a distributional regression model that embeds DT as well as PW modelling approaches into a copula regression framework for bivariate mixed outcomes. This allows to cast highly flexible regression models for the hazard rate of the time-to-event margin, which can include smooth functions of time, time-dependent covariates as well as time-varying effects. From a computational point of view, since our regression model for the time-to-event margin is based on the hazard rate, it does not require any constraints on the estimated smooth functions of time. Hence, we can avoid adding penalties or reparameterisations to impose such constraints, for example, that the baseline function of time must be monotonically decreasing, see Liu et al. (2018) and Marra and Radice (2019) for more details. The data transformation into long format depends on the configurations set by the analyst. Instead of treating the number of intervals as a tuning parameter, we suggest to set the number of intervals that cut the follow-up to a rather large value (e.g. 20 intervals) and let the penalty term of the baseline hazard take care of the compromise between overfitting the data and smoothness of the estimated function.

We embedded the DT and PW approaches into a bivariate distributional copula regression framework and implemented them in the convenient open source R library GJRM, allowing practitioners to cast flexible statistical models for mixed outcomes as well as bivariate time-to-event (implemented, but not presented here). Although not shown in this article, the proposed approach can also be used for modelling bivariate right-censored time-to-event responses as in Marra and Radice (2019), using DT or PW methods. This would involve modifying the log-likelihood to one based on bivariate discrete data shown in van der Wurp et al. (2020). We have demonstrated the versatility of the proposed approach by analysing the joint bivariate distribution of a non-commensurable outcome that consisted of binary and time-to-event margins using DT and PW approaches. The analysis showed that takeaway E-cyclists have a higher propensity to conduct a visual search, that is to inspect both directions of an intersection, as well as to run traffic red

lights compared to normal E-cyclists. The model shows a moderately strong estimated dependence between visual search and the time to red-light running for all E-cyclists. Our results also show that the red-light running of another cyclist waiting at the same intersection decreases the dependence between the responses.

Our approach has several potential avenues for future developments. One of them would be to adapt our software implementation to support other non-time-to-event margins such as ordinal outcomes following Hohberg et al. (2021). Another aspect could be the inclusion of a *cure fraction* (i.e. *cure models*, Peng and Yu, 2021) to account for subjects that will never experience the event. This phenomenon appears to be present in the data analyzed in Section 4, where cyclists that did not violate the traffic light are considered censored observations. Instead, one could argue that those cyclists belong to a sub-population that does not commit such infractions. An indication of the presence of such a cure fraction can be seen in the fact that the estimated survival function of normal (non-takeaway) E-cyclists remains far from zero. Thus, there may be E-cyclists that 'never' violate red traffic lights. We may consider relaxing the assumption of non-informative censoring times and model their distribution similar to Dettoni et al. (2020). Alternatively, we could explore relaxing the assumption of independent censoring and allow for dependent censoring as in Czado and Van Keilegom (2022). By doing so we would need to not only account for the dependence between the margins but also between the censoring and event times, which would result in a model that uses two copula functions. Another potential extension is modelling recurrent events based on Ramjith et al. (2022), which could require modifications of the proposed PW functions, but it would greatly extend the applicability of our methods. Data-driven variable selection could be tackled, for example, using a quadratic approximation of LASSO-type penalized models as in van der Wurp and Groll (2021) or boosting for distributional copula regression following Hans et al. (2022) and Strömer et al. (2023). Last, more flexible types of copula functions could be explored. For instance, Khoudraji's device (Genest et al., 1998) allows to model asymmetry in the dependence or alternatively a fully non-parametric approach could be pursued for modelling the dependence between the margins in cases where parametric copula families struggle to accommodate the data. These type of copulas are an attractive candidate for future work due to their parametric nature.

Acknowledgements

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Supplementary material

Supplementary material for this article is available online.

References

Bai L and Sze N (2020) Red light running behavior of bicyclists in urban area: Effects of bicycle type and bicycle group size. *Travel Behaviour and Society*, **21**, 226–34. doi: 10.1016/j.tbs.2020.07.003

Bender A and Scheipl F (2018) pammtools: Piecewise exponential Additive Mixed Modeling tools. URL https://arxiv.org/abs/1806.01042.

Bender A, Groll A and Scheipl F (2018) A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, **18**, 299–321. doi: 10.1177/1471082X17748083

Berger M and Schmid M (2018) Semiparametric regression for discrete time-to-event data. *Statistical Modelling*, **18**, 322–45. doi: 10.1177/1471082X17748084

Czado C and Van Keilegom I (2022) Dependent censoring based on parametric copulas. *Biometrika*, **110**, 721–38. doi: 10.1093/biomet/asac067

de Leon AR and Wu B (2011) Copula-based regression models for a bivariate mixed discrete and continuous outcome. *Statistics in Medicine*, **30**, 175–85. doi: 10.1002/sim.4087

Dettoni R, Marra G and Radice R (2020) Generalized link-based additive survival models with informative censoring. *Journal of Computational and Graphical Statistics*, **29**, 503–12. doi: 10.1080/10618600.2020.1724544

Eilers PHC and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–121. doi: 10.1214/ss/1038425655

Fraboni F, Puchades VM, Angelis MD, Pietrantoni L and Prati G (2018) Red-light running behavior of cyclists in Italy: An observational study. *Accident Analysis and Prevention*, **120**, 219–32. doi: 10.1016/j.aap.2018.08.013

Gao X, Zhao J and Gao H (2020) Red-light running behavior of delivery-service e-cyclists based on survival analysis. *Traffic Injury Prevention*, **21**, 558–62. doi: 10.1080/15389588.2020.1819989

Genest C, Ghoudi K and Rivest L-P (1998) Understanding relationships using copulas, by Edward Frees and Emiliano Valdez, January 1998. *North American Actuarial Journal*, **2**, 143–49. doi: 10.1080/10920277.1998.10595749

Groll A, Kneib T, Mayr A and Schauberger G (2018) On the dependency of soccer scores-A sparse bivariate Poisson model for the UEFA European football championship 2016. *Journal of Quantitative Analysis in Sports*, **14**, 65–79. doi: 10.1515/jqas-2017-0067

Hans N, Klein N, Faschingbauer F, Schneider M and Mayr A (2022) Boosting distributional copula regression. *Biometrics*, **79**, 2298–2310. doi: 10.1111/biom.13765

Hohberg M, Donat F, Marra G and Kneib T (2021) Beyond unidimensional poverty analysis using distributional copula models for mixed ordered-continuous outcomes. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **70**, 1365–90. doi: 10.1111/rssc.12517

Holford TR (1980) The analysis of rates and of survivorship using log-linear models. *Biometrics*, **36**, 299–305. doi: 10.2307/2529982

Karlis D and Ntzoufras I (2003) Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **52**, 381–93. doi: 10.1111/1467-9884.00366

Klein JP and Moeschberger ML (2003) *Survival Analysis*. New York: Springer. doi: 10.1007/b97377

Klein N, Kneib T, Klasen S and Lang S (2014) Bayesian structured additive distributional regression for multivariate responses. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **64**, 569–91. doi: 10.1111/rssc.12090

Klein N, Kneib T, Marra G, Radice R, Rokicki S and McGovern ME (2018) Mixed binary-continuous copula regression models with application to adverse birth outcomes. *Statistics in Medicine*, **38**, 413–36. doi: 10.1002/sim.7985

Klement EP, Mesiar R and Pap E (2002) Invariant copulas. *Kybernetika*, **38**, 275–85. URL http://eudml.org/doc/33582.

Kocherlakota S and Kocherlakota K (2017) *Bivariate Discrete Distributions*. Boca Raton, FL: CRC Press. doi: 10.1201/9781315138480

Lai CD and Balakrishnan N (2009) *Continuous Bivariate Distributions*. New York: Springer. doi: 10.1007/b101765

Laird N and Olivier D (1981) Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, **76**, 231–40. doi:10.2307/2287816

Liu X-R, Pawitan Y and Clements M (2018) Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, **27**, 1531–1546. doi: 10.1177/0962280216664760

Marra G and Radice R (2017) Bivariate copula additive models for location, scale and shape. *Computational Statistics and Data Analysis*, **112**, 99–113. doi: 10.1016/j.csda.2017.03.004

Marra G and Radice R (2019) Copula link-based additive models for right-censored event time data. *Journal of the American Statistical Association*, **115**, 886–95. doi: 10.1080/01621459.2019.1593178

Marra G and Radice R (2023) GJRM: Generalised Joint Regression Modelling. URL https://cran.r-project.org/web/ packages/GJRM/index.html.

Marra G, Radice R and Zimmer DM (2020) Estimating the binary endogenous effect of insurance on doctor visits by copula-based regression additive models. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **69**, 953–71. doi: 10.1111/rssc.12419

Medina-Olivares V, Calabrese R, Crook J and Lindgren F (2023) Joint models for longitudinal and discrete survival data in credit scoring. *European Journal of Operational Research*, **307**, 1457–73. doi: 10.1016/j.ejor.2022.10.022

Nagler T and Czado C (2016) Evading the curse of dimensionality in nonparametric density estimation with simplified vine copulas. *Journal of Multivariate Analysis*, **151**, 69–89. doi: 10.1016/j.jmva.2016.07.003

Najib MK, Nurdiati S and Sopaheluwakan A (2022) Multivariate fire risk models using copula regression in Kalimantan, Indonesia. *Natural Hazards*, **113**, 1263–83. doi: 10.1007/s11069-022-05346-3

Nelsen RB (2006) *An Introduction to Copulas*. New York: Springer. doi:10.1007/0-387-28678-0

Peng Y and Yu B (2021) *Cure Models*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/9780429032301

Radice R, Marra G and Wojty M (2016) Copula regression spline models for binary outcomes. *Statistics and Computing*, **26**, 981–95. URL http://dx.doi.org/10.1007/s11222-015-9581-6

Ramjith J, Bender A, Roes KCB and Jonker MA (2022) Recurrent events analysis with piece-wise exponential additive mixed models. *Statistical Modelling*. doi: 10.1177/1471082X221117612

Rappl A, Kneib T, Lang S and Bergherr E (2023) Spatial joint models through Bayesian structured piecewise additive joint modelling for longitudinal and time-to-event

data. *Statistics and Computing*, **33**, 135. doi: 10.1007/s11222-023-10293-5

Rizopoulos D (2012) *Joint Models for Longitudinal and Time-to-Event Data*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/b12208

Rupi F and Krizek KJ (2019) Visual eye gaze while cycling: Analyzing eye tracking at signalized intersections in urban conditions. *Sustainability*, **11**, 6089. doi: 10.3390/su11216089

Spissu E, Pinjari AR, Pendyala RM and Bhat CR (2009) A copula-based joint multinomial discrete-continuous model of vehicle type choice and miles of travel. *Transportation*, **36**, 403–22. doi: 10.1007/s11116-009-9208-x

Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V and Bastiani FD (2017) *Flexible Regression and Smoothing*. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/b21973

Strömer A, Klein N, Staerk C, Klinkhammer H and Mayr A (2023) Boosting multivariate structured additive distributional regression models. *Statistics in Medicine*, **42**, 1779–1801. doi: 10.1002/sim.9699

Suresh K, Taylor JMG and Tsodikov A (2019) A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker. *Biostatistics*, **22**, 504–21. doi: 10.1093/biostatistics/kxz049

Suresh K, Taylor JM and Tsodikov A (2021) A copula-based approach for dynamic prediction of survival with a binary time-dependent covariate. *Statistics in Medicine*, **40**, 4931–46. doi: 10.1002/sim.9102

Tang T, Wang H, Ma J and Zhou X (2020) Analysis of crossing behavior and violations of electric bikers at signalized intersections. *Journal of Advanced Transportation*, **2020**, 1–14. doi: 10.1155/2020/3594963

Tutz G and Schmid M (2016) *Modeling Discrete Time-to-Event Data*. Cham: Springer International Publishing. doi: 10.1007/978–3-319-28158-2

Umlauf N and Kneib T (2018) A primer on Bayesian distributional regression. *Statistical Modelling*, **18**, 219–47. doi: 10.1177/1471082X18759140

van der Wurp H and Groll A (2021) Introducing LASSO-type penalisation to generalised joint regression modelling for count data. *AStA Advances in Statistical Analysis*, **107**, 127–51. doi: 10.1007/s10182-021-00425-5

van der Wurp H, Groll A, Kneib T, Marra G and Radice R (2020) Generalised joint regression for count data: A penalty extension for competitive settings. *Statistics and Computing*, **30**, 1419–32. doi: 10.1007/s11222-020-09953-7

Wagner H and Tüchler R (2013). Sparse Bayesian modeling of mixed econometric data using data augmentation. In *Analysis of mixed data*, edited by de Leon AR and Chough KC. Pages 173–188. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/b14571

Welchowski T, Berger M, Koehler D and Schmid M (2022) *discSurv: Discrete Time Survival Analysis*. URL https://cran.r-project.org/web/packages/discSurv/index.html.

Wood S (2017) Generalized Additive Models: An Introduction with R. Boca Raton, FL: Chapman and Hall/CRC. doi: 10.1201/9781315370279

Zhang Z and Liu C (2023) Exploration of riding behaviors of food delivery riders: A naturalistic cycling study in Changsha, China. *Sustainability*, **15**, 16227. doi: 10.3390/su152316227