Simultaneous Equation Penalized Likelihood Estimation of Vehicle Accident Injury Severity

Francesco Donat and Giampiero Marra*

Abstract

A bivariate system of equations is developed to model ordinal polychotomous dependent variables within an additive regression framework. The functional form of covariate effects is assumed fairly flexible with appropriate smoothers included in the model representation to account for non-linearities and spatial variability in the data. Non-Gaussian error dependence structures are dealt with using Archimedean copulae, whose association parameter is also specified in terms of an additive predictor. The framework is employed to study the effects of several risk factors on the levels of injury sustained by individuals in two-vehicle accidents in France. The use of a bivariate model is motivated by the presence of common unobservables that may affect the inter-relationships between the various parties involved in the same crash. In this way, more reliable and precise estimates are obtained and mis-specification reduced via an enhanced model specification.

Key-words: Archimedean Copulae; Bivariate System of Equations; Ordinal Polychotomous Responses; Penalized Regression Splines; Road Safety.

1 Introduction

Vehicle-related injuries are a source of major concern for national governments and international organizations as they impact the life of millions of individuals around the world. In a recent report, the World Health Organization estimated that around 1.24 million of people die in road accidents every year, whereas approximately 20/50 million are involved in non-fatal injuries (**?**). This makes car crashes the eighth leading cause of death and the prominent one for young people aged 15-29 years. A global awareness campaign on this issue has been launched by the United Nations General Assembly with resolution 64/255 which proclaimed the Decade of Action for Road Safety for the period 2011-2020. Its goal is to stabilise, and possibly reduce, the trend in road traffic fatalities and save around 5 million lives over the foreseen action time.

To tackle this preventable major source of injury, several developed countries have created ad-hoc agencies funded from their national budgets. In France, for instance, a national task force (Comité Interministériel à la Sécurité Routière) was established in 1972 with the aim of defining governmental policies on the matters of road safety and ensuring their proper and timely enforcement. Legislations and information are commonly recognised to play a significant role in prevention and, at least in high-income countries, they showed a generalised reduction in fatal injuries. Despite these encouraging results, the annual costs of crash injuries for society are still high and have been estimated to exceed EUR 180 billion in the European Union alone. Deaths too constitute a non-negligible figure in national statistics (**?**). The French Observatoire National Interministériel de la Sécurité Routière (ONISR), for instance, counted

---

*F. Donat (*corresponding author*): Directorate General Statistics, European Central Bank; G. Marra: Department of Statistical Science, University College London. This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB. E-mail for correspondence: francesco.donat.12@ucl.ac.uk.

in its "Baromètre du mois de juillet 2015" that as much as 3,384 people died within 30 days from any road injury in the country during the past year. A deeper understanding of crash causations and injury severity is therefore fundamental to improve roadway safety, and hence contribute to a transport system that is sustainable in terms of its economic and social costs.

The study of injury severity in vehicle crashes may present several difficulties because of the intrinsic complexity of the problem. Severity levels are often the result of many observed factors (e.g., road geometry, vehicle standards, behaviour of road users) and some others that can be hardly measured by data collectors. For example, speed before impact, presence of moving obstacles on the pavement, or even sudden environmental-related factors are not typically recorded properly by police officers at the time they arrive. Notwithstanding they constitute important factors of accident dynamics. As pointed out by ?, ideally the injuries of all the parties involved in the same car accident should be modelled simultaneously as they are affected by identical crash conditions and occurrences. As such, it is likely that injuries sustained by these individuals are inter-related through their connection to the single crash event. Ignoring this issue, for instance by pooling individuals together across all crashes and estimating individual-level injury severity, may lead to inefficient model parameter estimates. A further limitation in applied research is the modelling of the injury severity sustained by the most severely injured person in a vehicle accident. The problem with this approach is that it does not offer a comprehensive treatment of the nature and level of injuries observed in the crash under investigation (e.g. in ? and ?). The use of multivariate statistical models is therefore advisable in these circumstances.

By acknowledging these concerns, the primary focus of the present paper is on the joint modelling of injury severity sustained by drivers in two-vehicle collisions. Specifically, we develop a class of models that accounts for the role that unobserved factors play in the determination of injury severity in vehicle crashes. Given that injury severity is commonly recorded on an ordinal scale by the police personnel, the proposed modelling strategy builds on a Cumulative Link Model (CLM, ?). This structure is extended to a bivariate semi-parametric framework, in which the effects of the continuous covariates on the responses of interest are estimated using penalized regression splines. In this way, non-linearities are handled flexibly without introducing, for instance, arbitrary categorisations of the relevant regressors into groups based on intervals or frequencies. In line with some recent methodological advances (e.g. ?), we extend the bivariate ordered probit regression framework of ? to incorporate several dependence structures of the responses induced by the class of Archimedean copulae. This development constitutes an advantage in empirical studies. Copulae, in fact, allow us to specify models beyond the classical Gaussian distribution and employ different marginals irrespectively of the particular association linking them. Most importantly, researchers are provided with new tools to assess the sensitivity of their results under different model specifications and assumptions. To enhance flexibility, the proposed framework can also accommodate the copula association (or dependence) parameter as specified in terms of an additive predictor. The model is then effectively allowing the magnitude of the association parameter to be different for each observation in the sample: heterogeneity from individuals' characteristics and accidents' dynamics is therefore mitigated.

Copulae have been previously considered in the general transportation literature by ?, ?, and ?, the latter in the context of road safety. However, to the best of our knowledge, their treatment in the setting of ordinal polychotomous responses with non-parametric covariate effects has not been analysed yet. This paper aims therefore at filling the gap. Our estimation approach takes advantage of a generic multivariate penalized Generalized Linear Model (GLM, ?) in which a number of penalisation terms are used to enforce certain desired characteristics of the functional form of the covariate effects. Among them, non-linearities and spatial variation within the data are shown to be all representable within the same generic framework through the careful choice of smoothers appropriately defined. These features are made operative and automatically estimable by the function `CopulaCLM`, which implements the ideas discussed in this

paper for the R computational environment (**?**). In the Bayesian literature, an analogous model for ordinal dependent variables has been introduced by **?** for a neuroscience application. Their paper, however, investigates only the effects of the Gaussian distribution on the classification ability of seemingly unrelated regression (SUR) equations without assessing the impacts of different dependence structures.

Semi-parametric models have received scarce attention in accident research, despite their well-known superiority over traditional parametric regressions in terms of model specification and efficiency. For example, it seems that the class of Generalized Additive Models (GAMs, **??**) has been only applied twice in the relevant literature by **?** and **?**. We suspect that this practice is due to the predominant empirical interest in ordinal response models, for which extra considerations and carefulness are required to harmonise their structures with those of univariate GAMs (e.g. **?**). In this vein, our contribution is to introduce flexible tools that researchers can use to better assess the effects of observed covariates on the responses of interest, and to allow for a more cautious judgment of the results obtained. The importance of flexibility is demonstrated in practice by the current article through the study of injury severity in various contexts where the joint modelling of the responses is pivotal for the risk factors' estimation and the correctness of the policy actions that could be drawn for the analysis. In the words of **?**, the "use of methodological approaches with known deficiencies [...] has the potential to lead to erroneous and ineffective safety policies that may result in unnecessary injuries and loss of life" (p. 16).

The remainder of the paper is structured as follows. In Section 2 we introduce the statistical model and discuss its representation and main features. We then devote Section 3 to some estimation issues concerning the optimisation of the penalized log-likelihood function and the related automatic smoothness selection. Theoretical considerations about the consistency of our estimator are also provided. The methods are finally illustrated by fitting a bivariate system of SUR equations to the levels of injury sustained by various parties involved in vehicle crashes (Section 4). Using data from the French ONISR, we compare two alternative scenarios and show: (i) how risk factors can have a peculiar influence on the same types of responses if different collision settings are considered; (ii) the various degrees of non-linearities characterising the effects of the continuous regressors; and (iii) the differences in the effects that risk factors have on the probability to sustain a certain injury severity level under several existing model specifications. Conclusions and lines for future research are drawn in Section 5.

2   Statistical Methods

We consider a pair of random variables $\boldsymbol{Y} := (Y_1, Y_2)^\top$ defined on the finite lattice $\mathcal{K}$ generated by the Cartesian product $\mathcal{K}_1 \times \mathcal{K}_2$, where $(\mathcal{K}_j, \preceq)$ is a totally ordered set for every $j \in \{1, 2\}$, and $\mathcal{K}_j := \{1, \ldots, K_j\}$ represents the levels of the categorical variable Y_j. The totality assumption implies that, under the binary relation $\preceq$, every element $k_j \in \mathcal{K}_j$ is comparable amongst all the others in the set. In the real data situations considered in this work, this excludes the possibility that a certain injury level cannot be appropriately recorded by the police officers. For example, this may happen whenever a driver fails to be assigned to a pre-specified severity category after a vehicle accident had occurred.

Let us denote by $\mathrm{y}_{j,i}$ a realisation of the random variable Y_j, for $i = 1, \ldots, n$, where n represents the sample size. Variable $\mathrm{y}_{j,i}$ indicates the level of injury sustained by individual i in vehicle j, with levels in the set $\mathcal{K}_j = \{\mathsf{no\ injury}, \mathsf{non\ hospitalised}, \mathsf{hospitalised}, \mathsf{fatal}\}$ for $j = 1, 2$. In road safety studies, the interest often lies in quantifying the contribution that risk factors have on the injury severity levels sustained by the parties involved in collisions. Mathematically, this is achieved by investigating the effects that a given set of covariates, as encoded in the array of realisations $\mathbf{x}_i := \mathrm{vec}(\mathbf{x}_{1,i}, \mathbf{x}_{2,i})$, have on a meaningful function of the conditional joint mass of the random vector $\boldsymbol{Y}$. The probability of event $\{\mathrm{y}_{1,i} = k_1, \mathrm{y}_{2,i} = k_2\}$ is denoted by $\pi_{k_1,k_2,i}$.

Name	$\mathcal{C}(u,v)$	$\mathcal{D}_\gamma$: support of γ	γ^*
Gaussian	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v))$	$[-1,1]$	$\tanh^{-1}(\gamma)$
Clayton	$(u^{-\gamma} + v^{-\gamma} - 1)^{-1/\gamma}$	$(0,\infty)$	$\log(\gamma - \varepsilon)$
Frank	$-\gamma^{-1} \log[1 + (e^{-\gamma u} - 1)(e^{-\gamma v} - 1)/(e^{-\gamma} - 1)]$	$\mathbb{R} \setminus \{0\}$	$\gamma - \varepsilon$
Gumbel	$\exp\left\{-[(-\log u) + (-\log v)]^{1/\gamma}\right\}$	$[1,\infty)$	$\log(\gamma - 1)$
Joe	$1 - [(1-u)^\gamma + (1-v)^\gamma - (1-u)^\gamma (1-v)^\gamma]^{1/\gamma}$	$(1,\infty)$	$\log(\gamma - 1 - \varepsilon)$

Table 1: Families of some bivariate Archimedean copula functions with association parameter γ. For optimisation and modelling purposes, an appropriate transformation γ^* is used in the estimation algorithm; this is given in the last column of the table. The quantity ε denotes the machine smallest floating point multiplied by 10^6 and is introduced to force the transformed association parameters to lie in their respective supports throughout estimation. Finally, we have introduced u and v to denote the marginals $\Phi(\eta_{j,k_j,i})$ for $j = 1, 2$, respectively.

We then consider the function $r : [0,1] \longrightarrow [0,1]$ such that

$$r(\pi_{k_1,k_2,i}) := \mathbb{P}(\mathrm{y}_{1,i} \preceq k_1, \mathrm{y}_{2,i} \preceq k_2 | \mathbf{x}_i) = \sum_{\tilde{k}_1 \preceq k_1} \sum_{\tilde{k}_2 \preceq k_2} \pi_{\tilde{k}_1,\tilde{k}_2,i}(\mathbf{x}_i), \tag{1}$$

where the vector $\mathbf{x}_{j,i}$ is assumed to collect the M_j explanatory variables of $\mathrm{y}_{j,i}$ for each j. Define now the map $\mathcal{C} : [0,1]^2 \longrightarrow [0,1]$ to be a two-place copula function – a joint distribution with uniformly distributed marginals – in which the dependence between the marginals is measured by the association parameter γ_i. Upon extending the approach introduced by **?** to a multivariate setting, we define a copula regression for bivariate ordinal polychotomous responses as

$$r(\pi_{k_1,k_2,i}) = g^{-1}(\boldsymbol{\eta}_{k_1,k_2,i}) := (\mathcal{C} \circ \mathcal{F})(\boldsymbol{\eta}_{k_1,k_2,i}) \in [0,1], \tag{2}$$

where

$$(\mathcal{C} \circ \mathcal{F})(\boldsymbol{\eta}_{k_1,k_2,i}) := \mathcal{C}(F_{1,1}(\eta_{1,k_1,i}), F_{1,2}(\eta_{2,k_2,i}); \gamma_i)$$

is the vector of the evaluations of link function $g^{-1} : \mathbb{R}^2 \longrightarrow [0,1]$ at the linear predictors $\boldsymbol{\eta}_{k_1,k_2,i} := (\eta_{1,k_1,i}, \eta_{2,k_2,i})^\top \in \mathbb{R}^2$. This array is made up of regression coefficients and covariates. The map $F_{1,j}$ is a univariate cumulative distribution function (cdf) for which we abide the convention of labelling the corresponding dimensionality with its first subscript. For notational convenience we have indicated the dependence of the linear predictor on the observed levels of injury severity by the subscript k_j. In particular, by defining $k := (k_1, k_2) \in \mathcal{K}$, the notation $\boldsymbol{\eta}_{k,i} := \boldsymbol{\eta}_{k_1,k_2,i}$ is used *in lieu* of the more precise $\boldsymbol{\eta}_i(k_1, k_2)$ and represent the bivariate vector collecting the linear predictors of the two parties involved in the i-th accident. This convention is applied analogously to all the other quantities in the model.

If the $F_{1,j}$'s are univariate distributions, the Sklar's theorem ensures that the composition map $(\mathcal{C} \circ \mathcal{F})$ is a 2-dimensional cdf with margins $F_{1,1}$ and $F_{1,2}$ (**?**). Table 1 lists the various copula functions in the Archimedean class with Standard Normal marginals implemented for this work. For completeness, we also consider rotated versions of the Clayton, Gumbel and Joe, which allow us to model negative dependences otherwise not implied by the respective canonical definitions (see Figure 1). Rotations are computed analytically using the definitions of **?**.

Since the strength of the association between the two equations may vary across observations (specifically across categories of vehicles, age and regions in the current study), our framework allows the copula dependence parameter to be specified as a function of an additive predictor (e.g., **?**, **?**, **?**). To this end, for a strictly increasing function h_γ that maps the predictors into the space of γ, we set

$$\gamma_i = h_\gamma(\eta_{\gamma,i}(\mathbf{x}_{\gamma,i})) \iff \gamma_i^* := h_\gamma^{-1}(\gamma_i) = \eta_{\gamma,i}(\mathbf{x}_{\gamma,i}).$$

The characterisation of the form of γ^* for every copula implemented is given in the last column of Table 1. For a complete account of copulae and their theoretical properties we refer the reader to the monograph of **?**, whereas an excellent practical guide to copula modelling is offered by **?**.
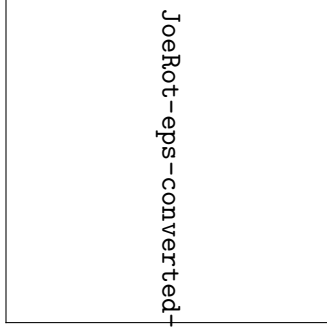
Figure 1: Random samples of 1,000 observations obtained from the two-place Joe copula function with both Standard Normal marginals and different degrees of rotation. The association parameter has been fixed such to corresponds to a Kendall's τ of 0.5 (-0.5 in case of 90 and 270 degrees). Explicit correspondences between γ and Kendall's τ in the context of Archimedean copulae are standard and can be found in **?**, for example.

Remark 1. *Equations (1) and (2) are the equivalent counterparts of a bivariate CLM, whose various specifications can be incorporated in (2) by characterising the distributions $F_{1,j}$'s. For instance, by fixing $F_{1,j} \equiv \Phi$, the Standard Normal cdf for each j, and letting $\mathcal{C}$ be the bivariate Gaussian copula with correlation coefficient ρ, equation (2) defines a bivariate ordered probit regression as*

$$r(\pi_{k,i}) = \Phi_2(\Phi^{-1}(\Phi(\eta_{1,k_1,i})), \Phi^{-1}(\Phi(\eta_{2,k_2,i})); \rho_i) = \Phi_2(\eta_{1,k_1,i}, \eta_{2,k_2,i}; \rho_i).$$

*The proposed model specification, although less intuitive that other existing ones (e.g. those invoking the concept of continuous latent response variables, e.g. **?**), pays off in terms of easiness and logic in the construction of score, Hessian and other quantities needed for inference.*

2.1 Further Considerations

Any model for ordinal responses is complemented by two restrictions. Let $\bar{k}$ and k be any two elements of $\mathcal{K}$ such that $\bar{k} \preceq k$ under a lexicographic order, then we need: (i) $r(\pi_{\bar{k},i}) \leq r(\pi_{k,i})$, and (ii) $\boldsymbol{\eta}_{\bar{k},i} \leq \boldsymbol{\eta}_{k,i}$ element-wise. The definition of some further parameters, termed cut points and labelled c_{j,k_j}'s, allows us to account for requirement (ii) simply by imposing $\{c_{j,k_j}\}_{k_j}$ to be an increasing sequence in k_j for every j. We also set $c_{j,K_j} = +\infty$ and $c_{j,1-1} =: c_{j,0} = -\infty$ to unbound the support of the linear predictors. In practice, for each individual i in vehicle j we observe a certain level k_j of injury severity sustained: the corresponding cut point is denoted by $c_{j,k_j,i}$ from which the vector $\boldsymbol{c}_j := (c_{j,k_j,i})_i \in \mathbb{R}^n$ is constructed.

These requirements are standard in ordinal response modelling and trace back at least to the work of **?** with an explicit reference also in **?**. The probability of the event $\{y_{1,i} = k_1, y_{2,i} = k_2\}$ is finally recovered by inverting the right-hand side of (1) and using (2):

$$\pi_{k,i} = r^{-1}(r(\pi_{k,i})) = \sum_{l,m \in \{0,1\}} (-1)^{l+m} \underbrace{\mathcal{C}(F_{1,1}(\eta_{1,k_1-l,i}), F_{1,2}(\eta_{2,k_2-m,i}); \gamma_i)}_{\equiv \mathcal{C}(\Phi(\eta_{1,k_1-l,i}), \Phi(\eta_{2,k_2-m,i}); \gamma_i)}. \qquad (3)$$

Upon defining the quantities

$$\begin{array}{lllll} \boldsymbol{c} & := & (\boldsymbol{c}_1, \boldsymbol{c}_2) & \in & \mathbb{R}^{n \times 2} & \text{cut points} \\ \mathbf{X} & := & (\mathbf{X}_1, \mathbf{X}_2) & \in & \mathbb{R}^{n \times (M_1+M_2)} & \text{covariates} \\ \boldsymbol{\beta} & := & \mathrm{diag}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) & \in & \mathbb{R}^{(M_1+M_2) \times 2} & \text{regression coefficients} \end{array}$$

and denoting by $\mathbf{X}_\gamma \in \mathbb{R}^{n \times M_\gamma}$ and $\boldsymbol{\beta}_\gamma \in \mathbb{R}^{M_\gamma}$ the components of the linear predictor of the copula association parameter $\boldsymbol{\gamma}^* := (\gamma_i^*)_i = \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma$, we re-write model (2) as

$$r(\pi_{k,i}) = (\mathcal{C} \circ \mathcal{F})(\boldsymbol{\eta}_{k,i}) = (\mathcal{C} \circ \mathcal{F})(\boldsymbol{c}_i - \boldsymbol{\beta}^\top \mathbf{x}_i), \qquad (4)$$

where $\boldsymbol{c}_i^\top$ and $\mathbf{x}_i^\top$ are the i-th rows of $\boldsymbol{c}$ and $\mathbf{X}$, respectively.

2.2 Additive Predictors and Penalized Regression Splines

Different covariate types are included in the model specification in an additive fashion and collected in the regression matrices $\mathbf{X}_j$ and $\mathbf{X}_\gamma$'s of dimensions $n \times M_j$. In particular, for any continuous regressor $\mathrm{v}_{j,l_j} \in \mathbb{R}$, $l_j = 1, \ldots, L_j$, like drivers' age or time of the accident, we advocate a non-parametric approach to curve estimation using penalized regression splines. We first assume that the functional form of v_{j,l_j} is representable via a smooth curve, $s_{j,l_j} : \mathbb{R} \longrightarrow \mathbb{R}$, then we seek to represent this effect without imposing any pre-determined parametric structure. This is achieved by choosing $H_j + 1$ knot points in the interior of $[\mathrm{v}_{j,l_j,(1)}, \mathrm{v}_{j,l_j,(n)}]$, with $H_j < n$ and $\mathrm{v}_{j,l_j,(i)} \leq \mathrm{v}_{j,l_j,(i+1)}$ for any $i = 1, \ldots, n-1$, and approximating the generic s_{j,l_j}-th curve by a linear combination of known basis spline functions, $\boldsymbol{b}_{j,l_j}$, and corresponding coefficients, $\boldsymbol{\delta}_{j,l_j}$. In other words we set

$$s_{j,l_j}(\mathrm{v}_{j,l_j,i}) \approx \boldsymbol{\delta}_{j,l_j}^\top \boldsymbol{b}_{j,l_j}(\mathrm{v}_{j,l_j,i}), \quad l_j = 1, \ldots, L_j,$$

where the above vectors are H_j-dimensional. Since the estimates so obtained are identified only up to an intercept term, a centering constraint of the form $\mathbf{1}_n^\top \boldsymbol{s}_{j,l_j} = 0$ has to be imposed, with $\boldsymbol{s}_{j,l_j} := (s_{j,l_j}(\mathrm{v}_{j,l_j,i}))_i$ (**?**).

Basis functions are usually chosen for their mathematical tractability and numerical stability. Among the most widely used in applications, we mention the cubic, penalized B-splines (**?**) and thin-plate regression splines (**?**), which are all supported by the computational routine attached to this article. Notice that the bases can be included in the design matrix by specifying the sub-matrix $\mathbf{X}_{[j,l_j]} := (\boldsymbol{b}_{j,l_j}(\mathrm{v}_{j,l_j,1})| \cdots |\boldsymbol{b}_{j,l_j}(\mathrm{v}_{j,l_j,n}))^\top \in \mathbb{R}^{n \times H_j}$ and, accordingly, the sub-vector $\boldsymbol{\beta}_{[j,l_j]} := \boldsymbol{\delta}_{j,l_j} \in \mathbb{R}^{H_j}$, for $j \in \{1, 2, \gamma\}$. Finally we write the linear predictors of the model as

$$
\begin{aligned}
\boldsymbol{\eta}_j &= \boldsymbol{c}_j - \mathbf{X}_j \boldsymbol{\beta}_j := \mathbf{Z}_j \boldsymbol{\theta}_j & \in \mathbb{R}^n & \quad j = 1, 2 \\
\boldsymbol{\eta}_\gamma &= \mathbf{X}_\gamma \boldsymbol{\beta}_\gamma & \in \mathbb{R}^n &
\end{aligned}
,
$$

with $\mathbf{Z}_j := (\mathbb{I}_j, -\mathbf{X}_j^p, \ldots, -\mathbf{X}_j^s)$ and $\boldsymbol{\theta}_j := \mathrm{vec}(\boldsymbol{c}_{j,k_j}, \boldsymbol{\beta}_j^p, \ldots, \boldsymbol{\beta}_j^s)$, where $\mathbb{I}_j := \mathrm{diag}(\mathbb{1}_{\mathrm{y}_{j,i}=k_j})_{i,k_j} \in \{0, 1\}^{n \times K_j - 1}$ and $\boldsymbol{c}_{j,k_j} := (c_{j,k_j})_{k_j} \in \mathbb{R}^{K_j - 1}$. The above representation is pivotal in applied research as it includes at the same time both non- and purely parametric covariate effects. This form is commonly termed semi-parametric in the statistical literature (e.g. **??**) and, once it is employed in equation (2), the additive extension of a CLM emerges. We consequently label it a copula *Cumulative Link Additive Model*. For the purposes of our empirical study, we implement the following structure for the i-th (pair of) individual(s)

$$r(\pi_{k,i}) = (\mathcal{C} \circ \mathcal{F})(\boldsymbol{c}_i - \boldsymbol{\beta}^\top \mathbf{x}_i) = \mathcal{C}(\Phi(\eta_{1,k_1,i}), \Phi(\eta_{2,k_2,i}), \gamma_i(\eta_{\gamma,i})) \in [0, 1],$$

where $\eta_{j,i}$ is the i-th element of $\boldsymbol{\eta}_j$. The dependence of γ_i on γ_i^* has been omitted to keep the notation more friendly.

3 Parameter Estimation

Under the usual *i.i.d.* conditions of the data generating process, the log-likelihood function corresponding to any bivariate model for ordinal responses is given by

$$\ell(\boldsymbol{\vartheta}|\mathbf{y}_1, \mathbf{y}_2, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_\gamma) = \sum_{i=1}^n \sum_{k \in \mathcal{K}} \mathbb{1}_{\mathrm{y}_{1,i}=k_1} \mathbb{1}_{\mathrm{y}_{2,i}=k_2} \log \pi_{k,i}(\mathbf{x}_{1,i}, \mathbf{x}_{2,i}, \mathbf{x}_{\gamma,i}) \in \mathbb{R},$$

where $\boldsymbol{\vartheta} := \mathrm{vec}(\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_\gamma)$ is the p-dimensional parameter vector and the expression of the joint probability mass is given by (3).

In order to make the model coherent in terms of its definition and parametrisation (cf. Section 2.1), we transform some of the parameters in $\boldsymbol{\vartheta}$ and perform estimation over a modified vector $\widetilde{\boldsymbol{\vartheta}} \in \mathbb{R}^p$. However, to avoid clutter in the notation, from now on we will use $\boldsymbol{\vartheta}$ to denote both the transformed and the original parameter vector as the distinction will be clearly inferred from the context. In particular, the transformation γ_i^* is copula-specific and ensures that an unconstrained optimisation algorithm can be employed in the derivation of the Maximum Likelihood Estimator. The cut points are instead normalised through a squared polynomial transform: $\tilde{c}_{j,1} = c_{j,1}$ and $\tilde{c}_{j,k_j} := \sqrt{c_{j,k_j} - c_{j,k_j-1}}$ for any $k_j \in \mathcal{K}_j \setminus \{1\}$ and all j, so that it is ensured the inequality $c_{j,k_j} = c_{j,k_j-1} + \tilde{c}_{j,k_j}^2 \geq c_{j,k_j-1}$.

3.1 Penalized GLM Representation

Following the terminology of **?**, model (4) gives the $(r, F_2, \mathbf{Z})$ representation of a GLM for categorical responses, with the caveat that the foreseen bivariate distribution has to be replaced here with the copula function, namely $F_2 \equiv \mathcal{C}(F_{1,1}, F_{1,2})$. Differently from a pure GLM, however, the specification of the functional form of covariate effects in terms of non-parametric and spatial components is likely to give raise to over-fitting unless a suitable regularisation is introduced. To this end, a ridge-type penalisation acting on the elements of $\boldsymbol{\vartheta}$ is defined. For each equation $j \in \{1, 2, \gamma\}$, we construct the quadratic form $\mathcal{P}_j := \boldsymbol{\beta}_j^\top \boldsymbol{\lambda}_j \mathbf{Q}_j \boldsymbol{\beta}_j$, where $\mathbf{Q}_j$ is a penalty matrix whose role is to enforce specific properties of the (j, l_j)-th function and $\boldsymbol{\lambda}_j \in [0, +\infty)^{M_j \times M_j}$ is a diagonal matrix of tuning parameters. They control for the trade-off between smoothness and fit: for instance, as $\lambda_{j,l_j} \to 0$ the estimated effects become wiggler and the fit perfect; vice versa wherever $\lambda_{j,l_j} \to \infty$. The selection of the "right" amount of smoothness is important in applied penalized regression splines modelling and a suitable method to deal with it is discussed in Section 3.3.

Upon setting $\mathbf{Q}_{\boldsymbol{\lambda}} := \text{diag}(\boldsymbol{\lambda}_j \mathbf{Q}_j)_{j \in \{1, 2, \gamma\}}$, we define an overall penalty $\mathbf{S}_{\boldsymbol{\lambda}}$ as $\mathbf{Q}_{\boldsymbol{\lambda}}$ padded with zeros such that $\mathcal{P}_{\boldsymbol{\lambda}} := \boldsymbol{\vartheta}^\top \mathbf{S}_{\boldsymbol{\lambda}} \boldsymbol{\vartheta} = \boldsymbol{\beta}^\top \mathbf{Q}_{\boldsymbol{\lambda}} \boldsymbol{\beta}$. In the following paragraphs, we illustrate the two types of penalty matrices used to adapt our generic representation to the specific model components employed in the analysis of injury severity in vehicle accidents. In general, no penalisation is attached to the fully parametric model components.

Continuous Covariates Regression splines account for non-linear smooth effects with varying degrees of complexity. The corresponding elements in $\boldsymbol{\lambda}_j$ are associated to a conventional measure of curvature typically defined through an integrated square second derivative spline penalty. Namely, we set $\mathbf{Q}_{j,l_j} := \int \boldsymbol{b}_{j,l_j}'' (\boldsymbol{b}_{j,l_j}'')^\top \mathrm{dv}_{j,l_j}$ with the integration conducted over the whole range of v_{j,l_j} (e.g. **?**). Examples comprise drivers' age and the time of the day (expressed in hours and minutes) in which the accident occurred.

Spatial Effects A location variable can be included in the model to control for the influence that geographical-specific factors may have on the phenomenon under investigation. A Markov random field smoother is implemented to exploit the spatial information in the data, and is suitable whenever a given area is made up of discrete contiguous units, as the 96 Departments of continental France.

Let us assume that we have R regions indexed by r, so that the spatial covariate effect of the i-th regressor is given by $\mathbf{x}_{j,i}^\top \boldsymbol{\beta}_{j,r_j}$, with $\boldsymbol{\beta}_{j,r_j} := (\beta_{r_j,1}, \ldots, \beta_{r_j,R_j})^\top$. The design matrix is constructed such that its (i, r_j)-th element equals 1 if observation i belongs to r_j, and 0 otherwise. The corresponding penalty matrix is based on the idea that spatially adjacent regions share similar effects. Hence, for any two regions r_j and s_j, the penalty matrix associated to the

spatial covariate is given by the adjacent matrix with elements

$$(\mathbf{Q}_j^s)_{[r_j, s_j]} := \begin{cases} -1 & r_j \neq s_j \ \wedge \ s_j \in \delta_{r_j} \\ 0 & r_j \neq s_j \ \wedge \ s_j \notin \delta_{r_j} \\ N_{r_j} & r_j = s_j \end{cases}$$

where δ_{r_j} is the set of regions adjacent to r_j, and $N_{r_j} := \#(\delta_{r_j})$ its cardinality (?).

3.2   Penalized Log-likelihood Function

Models in the $(r, F_2, \mathbf{Z})$ form augmented by the regularisation term $\mathcal{P}_{\boldsymbol{\lambda}}$ can be estimated within a Penalized Likelihood framework. The corresponding Maximum Penalized Likelihood Estimator (MPLE) is defined as the solution of the problem

$$\widehat{\boldsymbol{\vartheta}} := \underset{\boldsymbol{\vartheta} \in \mathbb{R}^p}{\arg\max} \left\{ \ell(\boldsymbol{\vartheta}|\cdot) - \frac{1}{2}\boldsymbol{\vartheta}^\top \mathbf{S}_{\boldsymbol{\lambda}} \boldsymbol{\vartheta} \right\}, \tag{5}$$

which is optimised for any given value of the smoothing parameter vector $\boldsymbol{\lambda}$. In practice, (5) is maximised using a trust-region algorithm which is generally more stable and faster than its line-search counterparts, especially for functions that exhibit non-linearities or regions that are close to flat (?, Ch. 4). The approach we have employed follows the one presented in ?, to which we refer the interested reader for practical details.

Analytic derivations of the score and Hessian employed in the computation of (5) are obtained by exploiting the multivariate GLM-type structure of model (4) and implemented in `CopulaCLM`. In particular, by letting $\ell_{\mathrm{p},i}(\boldsymbol{\vartheta}|\cdot)$ be the penalized log-likelihood contribution of the i-th observation, we obtain expressions

$$\boldsymbol{g}_{\mathrm{p},i} := \nabla_{\boldsymbol{\vartheta}} \ell_{\mathrm{p},i}(\boldsymbol{\vartheta}) = \frac{\partial \boldsymbol{\eta}_{k,i}}{\partial \boldsymbol{\vartheta}} \left(\frac{1}{\pi_{k,i}} \frac{\partial \mathcal{F}_{k,i}}{\partial \boldsymbol{\eta}_{k,i}} \frac{\partial \mathcal{C}_{k,i}}{\partial \mathcal{F}_{k,i}} \frac{\partial \pi_{k,i}}{\partial \boldsymbol{r}_{k,i}} \right) - \mathbf{S}_{\boldsymbol{\lambda},i}\boldsymbol{\vartheta} = \mathbf{D}_i^\top \mathbf{u}_i - \mathbf{S}_{\boldsymbol{\lambda},i}\boldsymbol{\vartheta},$$

with $\mathbf{D}_i^\top := (\partial \boldsymbol{\eta}_{k,i}/\partial \boldsymbol{\vartheta})$, and

$$\mathcal{H}_{\mathrm{p},i} := \nabla_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}^\top} \ell_{\mathrm{p},i}(\boldsymbol{\vartheta}) =$$
$$\mathbf{D}_i^\top \left[\frac{1}{\pi_{k,i}} \left\{ \frac{\partial \mathcal{F}_{k,i}}{\partial \boldsymbol{\eta}_{k,i}} \frac{\partial^2 \mathcal{C}_{k,i}}{\partial \mathcal{F}_{k,i} \partial \mathcal{F}_{k,i}^\top} \left(\frac{\partial \mathcal{F}_{k,i}}{\partial \boldsymbol{\eta}_{k,i}} \right)^\top + \frac{\partial^2 \mathcal{F}_{k,i}}{\partial \boldsymbol{\eta}_{k,i} \partial \boldsymbol{\eta}_{k,i}^\top} \frac{\partial \mathcal{C}_{k,i}}{\partial \mathcal{F}_{k,i}} \right\} \frac{\partial \pi_{k,i}}{\partial \boldsymbol{r}_{k,i}} - \mathbf{u}_i \mathbf{u}_i^\top \right] \mathbf{D}_i + \mathbf{K}_i - \mathbf{S}_{\boldsymbol{\lambda},i}.$$

The term $\mathcal{C}_{k,i}$ is the 4-dimensional that collects all the possible configurations of the joint distribution implied by (3), whereas $\mathcal{F}_{k,i} := (\Phi(\eta_{1,k_1-1,i}), \Phi(\eta_{2,k_2-1,i}), \Phi(\eta_{1,k_1,i}), \Phi(\eta_{2,k_2,i}), \gamma_i)^\top \in [0,1]^4 \times \mathcal{D}_\gamma$, with $\mathcal{D}_\gamma$ being the support of the dependence parameter. Expression in square brackets $[\cdot]$ above is commonly labelled $\mathbf{W}_i \in \mathbb{R}^{5\times5}$ and is the multivariate analogous of the weight matrix in classical iterative GLM estimation. Matrix $\mathbf{K}_i$ accounts instead for the transformed parameters discussed at the beginning of Section 3. Explicit forms for the components of these expressions are given in the Appendix (Lemma 1) for an equivalent representation. Notice that, upon defining

$$\boldsymbol{r}_{k,i} := (r(\pi_{k_1-1,k_2-1,i}), r(\pi_{k_1-1,k_2,i}), r(\pi_{k_1,k_2-1,i}), r(\pi_{k_1,k_2-1,i}))^\top \in [0,1]^4,$$

it holds that $\partial \pi_{k,i}/\partial \boldsymbol{r}_{k,i} = (1, -1, -1, 1)^\top$ from equivalence (3), whilst $\mathbf{D}_i$ is of dimension $(5 \times p)$ and includes the derivatives of the cut points and the covariate vector. Finally we set the quantities $\mathbf{D} := (\mathbf{D}_1^\top | \cdots | \mathbf{D}_n^\top | \mathbf{I}_p)^\top$, $\mathbf{u} := \mathrm{vec}(\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{0}_p)$ and $\mathbf{W} := -\mathrm{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_n, \mathbf{K})$, with $\mathbf{K} := \sum_i \mathbf{K}_i$, to obtain the global expressions needed for the algorithm's development.

3.3 Estimating the Smoothing Parameters

Given the multidimensional nature of our framework, computations may become burdensome if a direct grid search optimisation of, say, a prediction error criterion is used to smoothness selection. It is therefore essential to be able to estimate $\boldsymbol{\lambda}$ in an automatic way. To this end, we adopt a modified version of the Un-biased Risk Estimator recently applied by **?** to the context of bivariate dichotomous response modelling. The key idea is to base the derivation of a penalized iterative re-weighted least squares estimator on the Hessian matrix and score vector considered globally, rather than on their single components. Traditional methods, in fact, involve the computation of the square root and inversion of $\mathbf{W}$, which are typically more likely to fail to be positive definite for a subset of observations.

To overcome this inconvenience, let us define $\mathcal{I} := -\mathcal{H}$, with $\mathcal{H}$ being the Hessian of the un-penalized log-likelihood. By computing a first-order Taylor expansion of $\boldsymbol{g}_{\mathrm{p}}$ about the MPLE, and appropriately re-arranging its terms, we get an iterative algorithm of the form

$$\boldsymbol{\vartheta}^{[\alpha+1]} = (\mathcal{I}^{[\alpha]} + \mathbf{S}_{\boldsymbol{\lambda}}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[\alpha]}})^{-1} \sqrt{\mathcal{I}^{[\alpha]}} \mathbf{z}^{[\alpha]},$$

where $\mathbb{R}^p \ni \mathbf{z}^{[\alpha]} := \sqrt{\mathcal{I}^{[\alpha]}} \boldsymbol{\vartheta}^{[\alpha]} + \sqrt{\mathcal{I}^{[\alpha]}}^{-1} \boldsymbol{g}^{[\alpha]}$ is the pseudo-data vector associated to the penalized GLM model. It holds asymptotically that $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_p)$, where $\boldsymbol{\mu} := \sqrt{\mathcal{I}} \boldsymbol{\vartheta}_0$ is evaluated at the true parameter vector. Define now $\widehat{\boldsymbol{\mu}}$ the plug-in estimator obtained from the Generalized Least Squares estimate $\widehat{\boldsymbol{\vartheta}}_{\mathrm{GLS}}$ of the regression of $\mathbf{z}$ onto the columns of $\mathcal{I}$ and ridge penalty $\mathcal{P}_{\boldsymbol{\lambda}}$. Namely, $\widehat{\boldsymbol{\mu}} := \sqrt{\mathcal{I}} \widehat{\boldsymbol{\vartheta}}_{\mathrm{GLS}} = \mathbf{P}_{\boldsymbol{\lambda}} \mathbf{z}$, with $\mathbf{P}_{\boldsymbol{\lambda}} := \sqrt{\mathcal{I}} (\mathcal{I} + \mathbf{S}_{\boldsymbol{\lambda}})^{-1} \sqrt{\mathcal{I}}$ being the corresponding model influence matrix. Then we seek to estimate $\boldsymbol{\lambda}$ in such a way that the resulting non-parametric covariate effects are as close as possible to the real ones, that is by suppressing any complex structure which is not supported by the available data. We then compute

$$\mathbb{E}\|\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}\|^2 = \mathbb{E}\|\mathbf{z} - \mathbf{P}_{\boldsymbol{\lambda}} \mathbf{z}\|^2 - \widetilde{n} + 2\mathrm{tr}(\mathbf{P}_{\boldsymbol{\lambda}}), \tag{6}$$

where $\widetilde{n} := 5n$ and $\mathrm{tr}(\mathbf{P}_{\boldsymbol{\lambda}})$ defines the effective degrees of freedom (edf) of the penalized model. Hence an estimator for the smoothing parameters is defined iteratively as

$$
\begin{aligned}
\boldsymbol{\lambda}^{[\alpha+1]}|\boldsymbol{\vartheta}^{[\alpha+1]} &:= \underset{\boldsymbol{\lambda}}{\arg\min}\, \mathcal{V}(\boldsymbol{\lambda}) \\
&:= \underset{\boldsymbol{\lambda}}{\arg\min} \left\{ \|\mathbf{z}^{[\alpha+1]} - \mathbf{P}_{\boldsymbol{\lambda}}|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[\alpha]}} \mathbf{z}^{[\alpha+1]}\|^2 - \widetilde{n} + 2\mathrm{tr}(\mathbf{P}_{\boldsymbol{\lambda}})|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^{[\alpha]}} \right\},
\end{aligned}
$$

which is performed using the stable and efficient routine of **?**. The two steps detailed in this and previous sections are iterated until convergence.

It is worthwhile remarking that our scheme is implemented only from the knowledge of score and Hessian. In principle, therefore, any likelihood model regularised by a ridge-type penalisation term can be naturally extended to the various types of covariate effects described in this article in a relatively straightforward way. Potentially, also beyond the GLM family.

In the Appendix we establish the asymptotic behaviour of the proposed Maximum Penalized Likelihood Estimator for a model with non-parametric effects represented using penalized B-splines. The dimension of the spline basis, H_n, is allowed to growth with the sample size at a given rate. Theorem 1 shows the consistency of the MPLE at a rate of $o_P(\sqrt{H_n/n})$: this is in line with previous theoretical investigations and extends them to our bivariate setting.

3.4 Monte Carlo Evidence

A small set of Monte Carlo experiments has been conducted to investigate the finite-sample properties of the MPLE. For the sake of space, the exact definitions of the data generating processes (DGPs) employed are given in the Supplementary Material, so we just comment here on the results obtained. Figure 2 shows the behaviour of the estimator of the copula
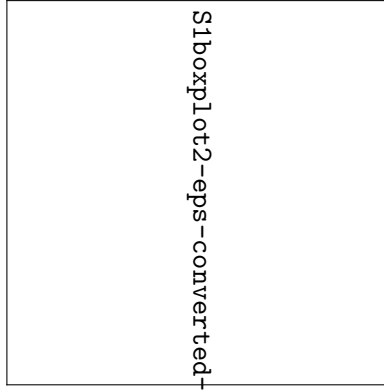
Figure 2: Box plots corresponding to the estimates of the copula association parameter γ for different sample sizes and copulae employed. γ has been reported under its corresponding Kendall's τ correlation coefficient, whose true simulated value ($\tau = 0.1$) is depicted as the red line in each panel. Results are obtained from 100 replications of the DGP detailed in the Supplementary Material.

association parameter under different settings and sample sizes. Specifically, the association parameter has been set to an intercept term for all the individuals, and is reported using the equivalent Kendall's τ metric to facilitate the comparison between the various copula scenarios. Its magnitude matches the one found in the real-data application. In line with expectations, we report that, as the sample size increases, the MPLE approaches the true value with a lower standard deviation and the parameter of interest is recovered reasonably well also at "modest" n of around $3,000$. This is a remarkable finding considering that a low magnitude of the association between the two equations locates γ close to the lower bound of its support (or to zero in the case of the Frank copula), which may result in numerical instabilities as well as make it hard to detect a dependence (**?**).

Some evidence of the ability of our model to recover the non-parametric covariate effects is provided in Figure 3, in which three smooth functions were included in the DGP. The first two refer to the equation for Y_1, and the last one to that for Y_2. The curves recovered at each replications illustrate graphically the degree on uncertainty attached to smooth function estimation, a concept formalised in **?** for the construction of point-wise Bayesian credible intervals in GAMs. Of course, less precise results are expected when fewer observations are used.

4 Data Analysis

Our empirical study uses data from the "Bulletins d'Analyse des Accidents Corporels" (BAAC) 2014. This dataset collects information about all vehicle accidents occurred in France that required the intervention of the police personnel. Agents were responsible for recording crash details, which were then centrally administrated by the ONIRS and subsequently published in the BAAC. At present, it comprises 4 headings referring to different accident features, and labelled accordingly "caractéristique", "lieux", "véhicules" and "usagers". Every accident is identified by a unique progressive serial number, which is identical for each vehicle and individual involved in the same crash.

Since the original dataset contains details on every kind of accident with at least one vehicle affected, we consider only those instances conforming with the features of primary interest for the present study. Accordingly, we select single-car crashes of four-wheels motor vehicles with two occupants (Scenario I), and two-vehicle collisions in which the injury severity of the two drivers is modelled jointly (Scenario II). The resulting datasets include $1,232$ and $20,079$ observations, respectively. Although some insights into the factors that influence injury outcomes can be
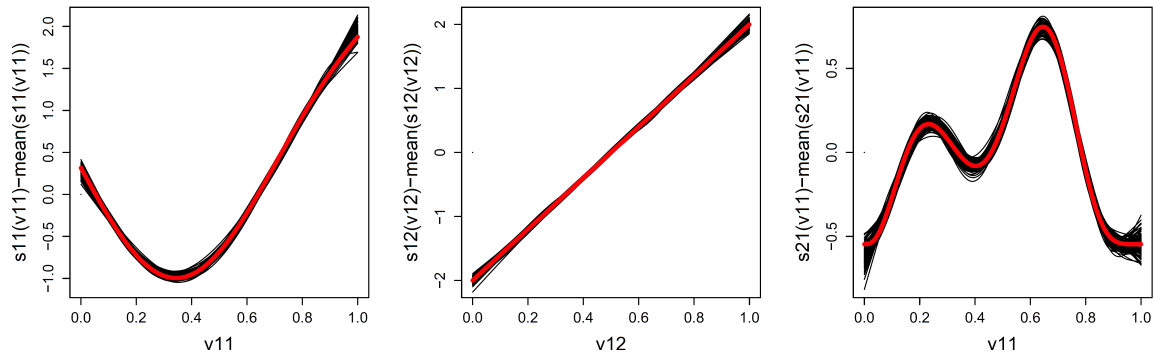
Figure 3: Estimated smooth curves obtained from 100 replicates of a Monte Carlo experiment comprising 10,000 simulated observations of a Joe copula model (true curves reported in red). The smooth components were represented using penalized thin-plate regression splines with basis dimensions equal to 10 and penalties based on second-order derivatives. Results are plotted on the scale of the linear predictors.

drawn from univariate models for crash, vehicle or roadway types, the results obtained in this way may not be directly applicable to all traffic crash scenarios (**?**). It may be the case, in fact, that even identical risk factors can affect the same response variables peculiarly once different crash dynamics are considered. For example, the presence of a roundabout may reduce the injury severity in two-vehicle collisions because roundabouts force drivers to a reduced speed and smooth the angle of the impact. However, they have a more complex geometry than, say, a straight roadway and that may result in a higher level of injury in case of single-car accident. Hence it is important to discern these issues in various settings. In what follows, we report all the details obtained from Scenario II in the main body of the article, whereas those referring to Scenario I are given in the Supplementary Material to the sake of space. The main findings are nonetheless commented and compared to those of Scenario II.

The aim of this analysis is to quantify the influence that some risk factors of interest have on the probability that vehicles' occupants sustain a certain level of injury severity, while accounting for the presence of unobserved variables affecting their inter-relationships. A bivariate copula Cumulative Link Additive Model is estimated to this purpose, with response variables being the injury severity sustained by two parties of the same crash. They are recorded by data collectors into four ordered categories: "no injury" (level 1), "non hospitalised" and "hospitalised" injuries (levels 2 and 3), and "fatal" (level 4) ones: a summary of the empirical distributions is reported in Table 2.

4.1 Models and Results

The proposed model specification follows closely previous published works in order to facilitate the comparison of our results with the relevant literature (e.g. **?**, **?** for Scenario I; or **?** for Scenario II). In line with the BAAC organisation, we group the explanatory variables into four macro-areas referring to occupant, vehicle, motorway and accident characteristics: they are all listed in the first column of Table 4. The continuous covariates included in the analysis, age of the occupants and time of collision, are estimated non-parametrically and approximated using row-rank penalized thin plate regression splines with basis dimension equal to 10 and penalties as described in Section 3.1. Since both of these variables are usually categorised to achieve a different fit for each ordinal level (e.g. in **?** and **?**), our methodology improves on the road safety literature by providing the researcher with an automatic data-driven way to model covariate effects flexibly. A better fit is therefore expected in this case. A regional variable identifying the location of each crash has also been included in the model specification

11

SCENARIO I					
Injury Severity Driver	Injury Severity Passenger				Marginals
	no injury	non hospitalised	hospitalised	fatal	
no injury	0 (0.00%)	188 (15.26%)	166 (13.47%)	17 (1.38%)	371 (30.11%)
non hospitalised	59 (4.79%)	251 (20.37%)	75 (6.09%)	10 (0.81%)	395 (32.07%)
hospitalised	69 (5.60%)	68 (5.52%)	213 (17.29%)	42 (3.41%)	392 (31.82%)
fatal	12 (0.97%)	11 (0.89%)	36 (2.92%)	15 (1.22%)	74 (6.00%)
Marginals	140 (11.36%)	518 (42.05%)	490 (39.77%)	84 (6.82%)	1,232 (100.00%)
SCENARIO II					
Injury Severity Driver A	Injury Severity Driver B				Marginals
	no injury	non hospitalised	hospitalised	fatal	
no injury	1,389 (6.92%)	6,190 (30.83%)	2,809 (13.99%)	138 (0.69%)	10,526 (52.43%)
non hospitalised	3,825 (19.05%)	1,348 (6.71%)	473 (2.36%)	49 (0.24%)	5,695 (28.36%)
hospitalised	2,206 (10.99%)	582 (2.90%)	613 (3.05%)	68 (0.34%)	3,469 (17.28%)
fatal	165 (0.82%)	90 (0.45%)	107 (0.53%)	27 (0.13%)	389 (1.93%)
Marginals	7,585 (37.78%)	8,210 (40.89%)	4,002 (19.93%)	282 (1.40%)	20,079 (100.00%)

Table 2: Distributions of injury severity sustained by driver and passenger (Scenario I) and by the two drivers (Scenario II) in vehicle-related accidents in the BAAC 2014 dataset. The categorisation follows the information recorded by the police personnels on the place of crash. The identification of a party as Driver A or B is based on the attribute `Num_Veh` in the heading `usagers`: it is an alphanumeric code where the initial letter refers to the vehicle involved in the accident and the subsequent number tot he location of the individual in the car. For example, the two drivers carry codes `A01` and `B01`. The letter A is conventionally assigned to the vehicle deemed to be responsible for the accident.

to control for Department-specific factors at a lower geographical level than continental France as a whole. This is illustrated, for example, by Figures 5 and 6.

As anticipated in the model description, in Scenario II the copula association parameter has been expressed as a function of known covariates to account for the possible heterogeneity in the distribution of risk factors among the categories of vehicle, individuals and Regions (e.g., **?**). This equation has been implemented in R by calling the line

```
eta3.dep <- ~ catv1 + catv2 + s(ans1) + s(ans2) + s(reg, bs = "mrf"),
```

where `catv{j}` denotes the vehicle type for each driver $j = \{1, 2\}$, `ans{j}` their respective age, and `reg` the Region in which the accident occurred. The syntax `s()` is used to define the non-parametric estimation of the argument in brackets, whilst `bs = "mrf"` specifies that a Markov random field smoother has to be employed.

Estimation has been performed for all the copula specifications described in Table 1, alongside with their corresponding degrees of rotations. A total of 15 different models were therefore fitted, including the one referring to the Independent copula obtained by pooling together all the observations and estimating a univariate ordered probit regression. Based on the Akaike Information Criterion (AIC, Table 3), the preferred models were those based on the Joe_0 in both Scenarios I and II. All the results reported in the proceeding discussion refer to this distribution.

Copula	Scenario I			Scenario II		
	$\widehat{\gamma}$	95% CI	AIC	$\widehat{\gamma}$	95% CI	AIC
Independent	–	–	$6,253.07$	–	–	–
Gaussian	0.0118	$[-0.0543, 0.0757]$	$6,262.80$	0.0476	$[0.0259, 0.0698]$	$72,468.21$
Clayton$_{180}$	0.2281	$[\ 0.1525, 0.3479]$	$6,224.60$	0.4303	$[0.3920, 0.4838]$	$71,991.01$
Frank	0.1945	$[-0.2152, 0.5975]$	$6,261.00$	0.2162	$[0.0025, 0.4194]$	$72,140.29$
Gumbel$_0$	1.0780	$[\ 1.0425, 1.1375]$	$6,238.46$	1.2220	$[1.1979, 1.2565]$	$72,109.84$
Joe$_0$	1.2130	$[\ 1.1412, 1.3176]$	$\mathbf{6,219.17}$	1.3607	$[1.3301, 1.4059]$	$\mathbf{71,882.27}$

Table 3: Estimated association parameters for the converging copula models considered in the article, with corresponding 95% confidence interval. Confidence intervals have been obtained via simulation from the posterior distribution of the MPLE following the procedure described in ?. The last column of each scenario shows the associated AIC: the ones of the selected models are in bold. Since the penalty matrix in the estimation algorithm can suppress some dimensions of the parameter space, we have: AIC $= -2\ell_{\mathsf{p}}(\widehat{\boldsymbol{\vartheta}}|\cdot) + 2edf$, where the edf are the estimated degrees of freedom mentioned in Section 3.3. The AIC for the independent case in Scenario I has one parameter less than the others (i.e. the association parameter), while the one of Scenario II is not given because it is based on double the number of observations, and so is misleading (remember that the independent case corresponds to a model in which all the observations are pooled together). Similar results were obtained when the Bayesian Information Criterion was used. Regarding Scenario II, the association parameter reported is the average of $\hat{\gamma}_i$ for all the data points; confidence intervals have been adjusted accordingly.

Estimated Association Between the Responses   It is interesting to note that all the copulae supported by the data are all those defined for a positive, albeit low, association parameter γ. In particular, we report that these models exhibit an upper tail dependence, which indicates that high values of the propensity of injury severity for one individual tend to be associated with high values recorded for the other person in the study.

Before illustrating the results obtained in more details, some comments on the estimated association parameter in Scenario II are in order. Our analysis shows that the estimated $\hat{\gamma}$'s averaged by Department are fairly constant, with possibly some exemptions in the south of the country (refer to the bottom-right map of Figure 4). However, a naive look at the average dependence parameter hinders heterogeneity among accidents' characteristics. In fact, if we average instead the association parameter for collisions between two motorcycles ($< 125cm^3$) or between two cars, we obtain 3.35 and 1.09, respectively. This means that the role of underlying factors in the propensity to sustain higher injury severity in an accident is stronger in case of crashes between motorcycles than between cars. This result highlights the increasing need for flexible methodologies in road safety studies. Previous research has generally reported that a positive association exists between the two equations in the system. In other words, unobserved factors tend, on average, to drive in the same direction the injury severity sustained by the occupants of the same motor vehicle in single-car accidents, and by both drivers in two-car crashes.

Estimated Covariate Effects   In Table 4, the estimates corresponding to the coefficients of the parametric covariate effects are reported alongside with those arising from the independent model (refer to Table S1 in the Supplementary Material for those of Scenario I). Notice that, by pooling all the individuals together, the number of observations doubled compared to the employment of a bivariate system of equations. Scenario I uses instead a bivariate Gaussian model with γ structurally constrained to zero as independence benchmark. In fact, we believe that drivers and passengers are often subject to different effects for the same risk factor, and this would in turn distort any analysis conducted on a single pooled model. Consider, for instance, the different manners of collision. Qualitatively, the estimates obtained show that a sideswipe collision to the left would result in a higher injury severity propensity for the driver; however,

whenever it occurs to the right, the front passenger is likely to be the most affected party. This distinction is not captured when a single-equation model is used. Table S2 in the Supplementary Material reports the full set of estimates referring to the pooled independent model for Scenario I.

Overall, the sign of many of the estimated coefficients confirms the results previously reported in the accident literature and is consistent with expectations. In particular, for Scenario II we find that females show a higher propensity to sustain severe injuries when compared to males, regardless of the vehicle they are sit in. This gender difference might be related to weight, body mass and other factors, and is in line with other authors' findings (see, for instance, **?**). In addition, travelling in motor vehicles with four-wheels is generally associated with lower injuries than is travelling on motorcycles, with larger ones ($> 125\text{cm}^3$) being unsafer compared to smaller motorised two-wheelers ($< 125\text{cm}^3$). These results are intuitive since a better protection from severe injuries can be expected in cars, whereas small motorcycles may be constrained to reduced speed limits and restricted circulation regimes on faster roadways. Among environmental factors, lighting in force at night tends to lower the propensity of a severe injury. We note a contrasting effect between driver and passenger in Scenario I when we compare adverse weather conditions against normal ones in both scenarios. These estimates are quite surprising: if on the one hand the presence of water or ice on street pavement may reduce vehicles' friction, hence fostering the likelihood of an accident, on the other hand cars may proceed at a reduced speed and drivers be more cautious (**?**). Analogous arguments can be made in those instances of reduced visibility, like foggy weather conditions. In any case, systematic differences between driver and passenger, as well as among Scenarios I and II, strengthen the assertion that risk factors act differently on sustained injury severity in various crash settings. Some implications for roadway design can be drawn from this study. Specifically, our analysis shows that the presence of roundabouts is likely to be associated with a reduction in injury propensity whenever two vehicles hit each other (Scenario II), whereas a corresponding increase emerges for both driver and passenger if only one car is involved in the crash (Scenario I). This is presumably a result of the implicit accident dynamics leading to the two scenarios, where the latter is likely to be mostly caused by a loss of control of the vehicle at entry, on the circulatory roadway or its exit. This suggests undertaking suitable actions in terms of design and safety countermeasures of roundabouts to make them an even more effective tool for prevention.

The estimated non-parametric model components for Scenario II are depicted in Figure 4. In line with the literature, our analysis highlights an almost steady effect of age on injury severity up to around 40-45 years and it increases rapidly for people older than 60 years. This evidence deserves some attention: with an increasing number of elderly people in Europe, the implementation of ad-hoc actions and/or legislations seem to us of growing importance to foster road safety measures. The maps in Figure 4 depict the estimated contribution of each Department to the propensity of injury severity for the two drivers. Value greater than zero imply that two-vehicle accidents incurred in these Departments tend to result in higher levels of injury.

Quantitative Interpretation of the Results Parametric and non-parametric estimates have to be interpreted qualitatively since, by model construction, they only affect directly the propensity of injury rather than the responses measured on the ordinal scale. This is in line with the interpretation of a model for ordinal responses as a coarse version of a latent continuous variable, i.e. the propensity of injury. This practical limitation is accounted here by computing the model (pseudo-)elasticities, roughly interpretable as the percent change in the probability that the average individual sustains a certain injury level for a 1% increase in a measured continuous covariate (e.g. **?**). Alternatively, for any categorical regressor and under the maintained assumptions, the pseudo-elasticity of the (j, m_j)-th covariate on the j-th

	SCENARIO II: ESTIMATES					
Variables	Driver A		Driver B		Independent model	
	estimates	(se)	estimates	(se)	estimates	(se)

	estimates	(se)	estimates	(se)	estimates	(se)
Occupant Characteristics						
Gender (male)						
female	0.2426	(0.0193)	0.3243	(0.0183)	0.3030	(0.0136)
Vehicle Characteristics						
Type (motorcycle $< 125cm^3$)						
Motorcycle $> 125cm^3$	0.0254	(0.0210)	0.0960	(0.0256)	0.1235	(0.0198)
Vehicle M1	−1.8511	(0.0216)	−1.5011	(0.0210)	−1.5058	(0.0151)
Vehicle N1	−2.1256	(0.0443)	−1.8402	(0.0427)	−1.7876	(0.0307)
Motorway Characteristics						
Intersection (off intersection)						
X	−0.0100	(0.0221)	−0.0475	(0.0214)	−0.0418	(0.0158)
T	−0.2312	(0.0262)	−0.1654	(0.0252)	−0.2053	(0.0188)
Y	−0.2046	(0.0628)	−0.2148	(0.0611)	−0.2274	(0.0455)
> 4 branches	−0.2191	(0.0767)	−0.3022	(0.0731)	−0.2753	(0.0550)
roundabout	−0.4087	(0.0530)	−0.3370	(0.0498)	−0.3970	(0.0374)
circus/square	−0.3129	(0.0859)	−0.1856	(0.0738)	−0.2299	(0.0579)
level crossing	0.5300	(0.4254)	0.2461	(0.4585)	0.3719	(0.3259)
other	0.0474	(0.0751)	0.0458	(0.0734)	0.0537	(0.0539)
Type (motorway)						
Route Nationale	0.1173	(0.0443)	0.2620	(0.0442)	0.2096	(0.0321)
Route Départementale	0.1179	(0.0299)	0.3300	(0.0297)	0.2385	(0.0216)
Voie Communale	−0.4107	(0.0296)	−0.1225	(0.0289)	−0.2626	(0.0213)
off public road network	−0.2205	(0.3013)	−0.8125	(0.3500)	−0.5129	(0.2399)
parking	−0.4365	(0.1595)	−0.1506	(0.1502)	−0.3182	(0.1128)
other	−0.1280	(0.0950)	−0.0730	(0.0948)	−0.0993	(0.0702)
Circulation regime (missing)						
one-way	−0.1245	(0.0391)	−0.2032	(0.0379)	−0.1695	(0.0280)
two-way	0.0532	(0.0334)	−0.0399	(0.0324)	0.0068	(0.0240)
presence of median	−0.1443	(0.0418)	−0.1672	(0.0407)	−0.1764	(0.0299)
other	−0.0781	(0.1095)	−0.1848	(0.1084)	−0.1293	(0.0796)
Horizontal alignment (straight)						
left curve	0.2338	(0.0320)	0.1654	(0.0319)	0.2037	(0.0232)
right curve	0.3645	(0.0318)	0.1438	(0.0322)	0.2734	(0.0232)
S	0.4188	(0.0780)	0.2089	(0.0789)	0.3394	(0.0568)
Accident Characteristics						
Lighting (daylight)						
sunrise/sunset	0.0310	(0.0362)	0.0661	(0.0351)	0.0281	(0.0268)
night without street lights	0.4592	(0.0364)	0.4537	(0.0354)	0.4198	(0.0272)
night, street lights in force	−0.1067	(0.0288)	−0.0150	(0.0260)	−0.1143	(0.0221)
Atmospheric conditions (normal)						
light rain	0.0785	(0.0262)	0.0336	(0.0254)	0.0665	(0.0187)
heavy rain	0.2363	(0.0530)	0.0872	(0.0525)	0.1804	(0.0382)
snow	0.4928	(0.2016)	0.7005	(0.1975)	0.6305	(0.1417)
fog	0.2811	(0.1106)	0.2730	(0.1107)	0.2836	(0.0791)
heavy wind/storm	0.2155	(0.2223)	−0.0090	(0.2219)	0.1117	(0.1613)
clear	0.0999	(0.0800)	0.2677	(0.0788)	0.1842	(0.0573)
clouds	0.2124	(0.0447)	0.1243	(0.0447)	0.1770	(0.0324)
Manner of collision (missing/other)						
head-on	0.0732	(0.0323)	0.0576	(0.0316)	0.0861	(0.0236)
rear-end	−0.0780	(0.0452)	−0.0965	(0.0359)	−0.0363	(0.0287)
sideswipe, right	0.0739	(0.0456)	−0.0978	(0.0464)	−0.0076	(0.0341)
sideswipe, left	0.1060	(0.0463)	0.0006	(0.0418)	0.0618	(0.0326)
Passenger (no)						
yes	0.0517	(0.0237)	−0.0098	(0.0207)	0.0702	(0.0159)
Security device (not put on)						
put on	−0.0451	(0.0225)	−0.0715	(0.0235)	−0.0294	(0.0168)
$c_{j,1}$	−1.4236	(0.0641)	−1.3804	(0.0627)	−1.1815	(0.0462)
$c_{j,2}$	−0.3453	(0.0056)	0.0345	(0.0051)	0.0746	(0.0038)
$c_{j,3}$	1.0458	(0.0095)	1.6872	(0.0100)	1.6557	(0.0071)
No. observations	20, 079		20, 079		40, 158	

Table 4: Estimates and associated standard errors (in brackets) obtained by applying `CopulaCLM` to the BAAC 2014 data. Corresponding confidence intervals are available form the authors upon request. Results are for the parametric model components of Scenario II when the Joe$_0$ copula function is used. The last columns report the results corresponding to the independent model. The reference categories are given in round brackets next to the variable names to which they refer.
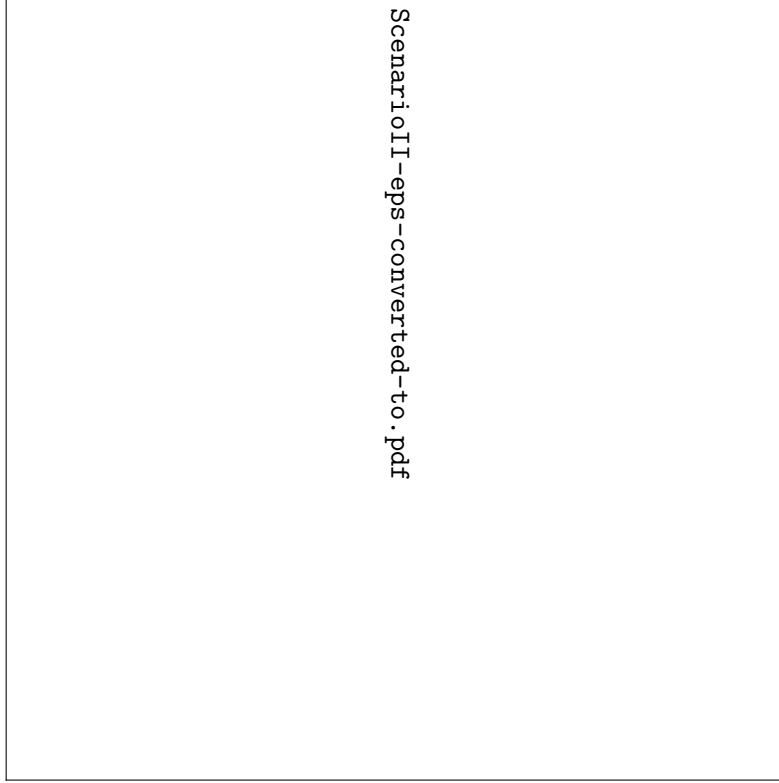
Figure 4: Smooth functions estimates and associated 95% point-wise confidence intervals corresponding to the three equations of the bivariate model applied to the BAAC 2014 data under Scenario II, when the Joe_0 copula function is used. The first two equations have as response variable the injury severity sustained by Drivers A and B, respectively, whereas the third one describes the copula dependence parameter in terms of a linear predictor. The curves relate to the effects of `age` and `time` (expressed in hours and minutes, `hrmn`) on the propensity of injury severity of drivers in two-car collisions. Confidence intervals are based on the results of **?** for GAMs, which can be accommodated in bivariate penalized GLMs admitting a $(r, F_2, \mathbf{Z})$ representation as explained by **?**. The effective degrees of freedom are reported into brackets in the y-axis caption, with a value of one corresponding to a straight line estimate. The covariate values are represented by a jittered rug plot at the bottom of each graph. The maps, instead, quantify the estimates obtained for the regional variable in each of the 96 Department of continental France. The third map shows the averaged association parameter in each Department.

response for individual i is defined as

$$\widehat{\mathcal{E}}_{\mathrm{x}_{j,m_j,i}}^{\mathbb{P}[\mathrm{y}_{j,i}=k_j]} :=$$
$$\sum_{l\in\{0,1\}} (-1)^l \left[\frac{\Phi(\eta_{j,k_j,i} - \beta_{j,m_j}\{\mathbb{1}_{l=0} - \mathrm{x}_{j,m_j,i}\}) - \Phi(\eta_{j,k_j-1,i} - \beta_{j,m_j}\{\mathbb{1}_{l=0} - \mathrm{x}_{j,m_j,i}\})}{\Phi(\eta_{j,k_j,i}) - \Phi(\eta_{j,k_j-1,i})} \right]_{\boldsymbol{\vartheta}=\widehat{\boldsymbol{\vartheta}}},$$

which is averaged to obtain

$$\widehat{\mathcal{E}}_{\mathrm{x}_{j,m_j}}^{\mathbb{P}[\mathrm{y}_j=k_j]}(\widehat{\boldsymbol{\vartheta}}) = \left[\frac{1}{n} \sum_{i=1}^{n} \widehat{\mathcal{E}}_{\mathrm{x}_{j,m_j,i}}^{\mathbb{P}[\mathrm{y}_{j,i}=k_j]}(\boldsymbol{\vartheta})|_{\boldsymbol{\vartheta}=\widehat{\boldsymbol{\vartheta}}} \right] \cdot 100.$$

The statistic above estimates the percent change corresponding to the observation of a certain level of the (j, m_j)-th categorical covariate. For instance, it measures by how much the probability of an individual being hospitalised changes (on average) when the crash occurs at roundabouts, with respect to all the other intersection types. Table 5 reports all pseudo-
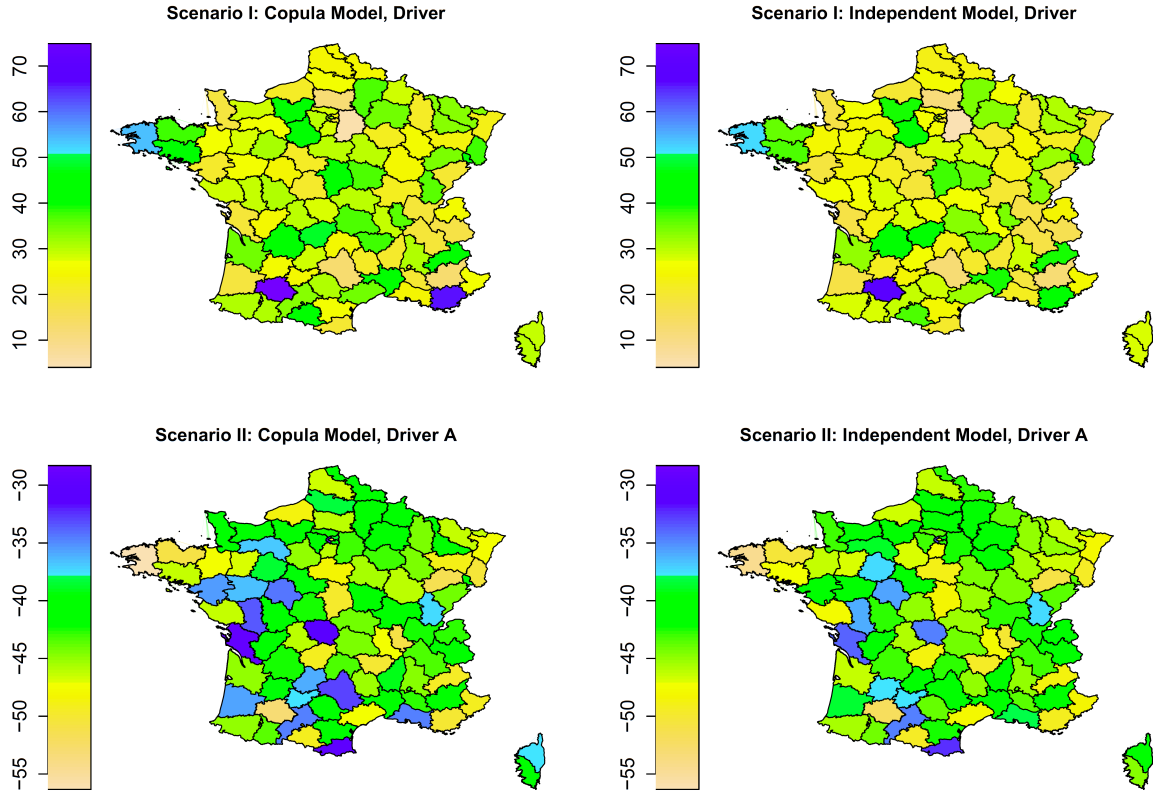
Figure 5: Pseudo-elasticities of the presence of roundabouts on the probability of the average occupant to sustain a hospitalised injury in each of the 96 French Departments. The comparison between copula and independent models is presented for Scenario I, top row, and Scenario II, bottom row. Notice: (i) the qualitative analysis of the coefficients' signs is enhanced by the formal computation of the (pseudo-)elasticities; and (ii) the difference in the estimates obtained when a pooled univariate model is employed rather than a bivariate one (Scenario II). In this case, only results for Driver A have been reported for no good reasons.

elasticities for a number of competing models: they all confirm the previous considerations based on the models' estimates.

Figure 5 illustrates the effect of the presence of roundabouts computed at the geographical location of each collision, and we compare them against the pseudo-elasticities obtained from the corresponding independent models. Not accounting for unobservables in the study may lead to underestimated elasticity effects, which may in turn distort the information transmitted to policymakers throughout the decision process in place. This is seen in the bottom row for Scenario II, where a *univariate* model is compared against the preferred copula specification. In the corresponding choropleth map we observe higher (i.e. less negative) elasticities for a number of Departments, including Orne, Morbihan, Loire Atlantique, Maine et Loire, Indre et Loire, Deux Sèvres, Creuse, Charente Martime, Landes, Lot and Aveyron, roughly corresponding to the Regions of Pays de la Loire and Poitou-Charentes. The same figure referring to Scenario I shows instead that this difference is qualitatively almost indistinguishable. This may be explained by the fact that, in the latter, a *bivariate* independent model is employed. In fact, since this case assumes the product of two Normals as reference joint distribution, as γ tends to the infimum of its support the copula model converges to this independence benchmark.

The gains of applying a bivariate copula Cumulative Link Additive Model are perhaps better summarised in Figure 6, where we compare the pseudo-elasticities of S curves on hospitalised injuries obtained under alternative model specifications. The Gaussian copula reported for Scenario I (top-right map) corresponds to a semi-parametric bivariate probit regression model,

Variables	Joe$_0$: Semi-parametric Driver A	Driver B	Independent Drivers	Joe$_0$: Parametric Driver A	Driver B
Occupant Characteristics					
Gender (male)					
female	44.3759	58.4245	57.2809	42.6745	59.6290
Vehicle Characteristics					
Type (motorcycle < 125cm^3)					
Motorcycle > 125cm^3	4.0750	15.1656	22.4733	−1.6936	13.9652
Vehicle M1	−96.0412	−93.3907	−93.9468	−96.3122	−93.9779
Vehicle N1	−97.7507	−96.9027	−96.7766	−98.0451	−97.2620
Motorway Characteristics					
Intersection (off intersection)					
X	−1.5411	−6.8827	−6.3512	0.3398	−5.5226
T	−30.6038	−22.3151	−28.6553	−28.6553	−21.0711
Y	−27.5807	−28.1332	−30.6469	−21.5933	−23.9078
> 4 branches	−37.5156	−37.9463	−35.9644	−17.1633	−27.2364
roundabout	−48.9606	−39.0754	−47.9295	−49.3973	−43.4329
circus/square	−24.7456	−24.4390	−30.9286	−28.8199	−9.8975
level crossing	42.4098	58.0468	73.3202	88.0083	32.8606
other	7.5959	7.0168	8.6798	3.0991	2.3340
Type (motorway)					
Route Nationale	19.7095	45.5678	37.3954	13.1281	44.5684
Route Départementale	19.8178	59.6455	43.3151	14.2249	55.9591
Voie Communale	−48.1706	−16.9727	−34.5895	−48.6248	−15.6858
off public road network	−29.4003	−74.0948	−57.4700	−25.5407	−72.5099
parking	−50.3304	−20.5108	−40.4252	−54.8696	−25.0024
other	−18.1880	−10.4255	−14.5495	−24.8368	−14.6871
Circulation regime (missing)					
one-way	−17.7225	−26.7986	−23.7186	−8.0744	−18.8303
two-way	8.5552	−5.8169	1.0663	9.7344	−4.9204
presence of median	−22.5343	−22.2966	−24.5684	−11.9607	−16.4462
other	−11.4739	−24.6431	−18.5694	−9.8600	−24.3300
Horizontal alignment (straight)					
left curve	42.5274	27.2082	36.1979	37.5154	22.2455
right curve	72.3288	23.3504	50.7559	66.8536	18.2799
S	117.6723	40.3920	83.0723	103.3627	31.1941
Accident Characteristics					
Lighting (daylight)					
sunrise/sunset	4.9171	10.2574	4.4718	7.8686	10.7904
night without street lights	96.9460	87.7765	85.1674	104.2143	84.8416
night, street lights in force	−14.5202	−2.1912	−15.6351	−4.1686	2.8148
Atmospheric conditions (normal)					
light rain	12.8390	5.1190	10.8377	12.9079	6.9491
heavy rain	43.0424	13.7026	31.5915	39.5674	10.8913
snow	106.3239	153.8087	144.4313	72.0278	120.7264
fog	52.7815	47.7781	52.9707	36.1097	38.0877
heavy wind/storm	38.7047	−1.3342	18.7279	24.7814	−5.1406
clear	16.5813	46.7119	32.3513	11.1678	38.7617
clouds	38.0796	19.9676	30.9299	37.1391	21.8370
Manner of collision (missing/other)					
heads-on	9.0503	3.4770	14.9423	7.3696	13.5057
rear-end	−15.2807	−12.4695	−5.4242	−16.4820	−9.6178
sideswipe, right	10.3250	−11.2449	−1.1826	6.9893	−10.1128
sideswipe, left	15.7586	2.1764	10.3965	13.2137	3.5122
Passenger (no)					
yes	12.7516	0.7857	11.4579	11.6577	−2.6750
Security device (not put on)					
put on	−7.6349	−9.3241	−4.4262	−10.0084	−9.5552
No. observations	20,079		40,158	20,079	

Table 5: Pseudo-elasticities for Scenario II obtained by applying the preferred Joe$_0$ copula, independent and the fully parametric models. Quantities are computed with respect to the hospitalised injuries.
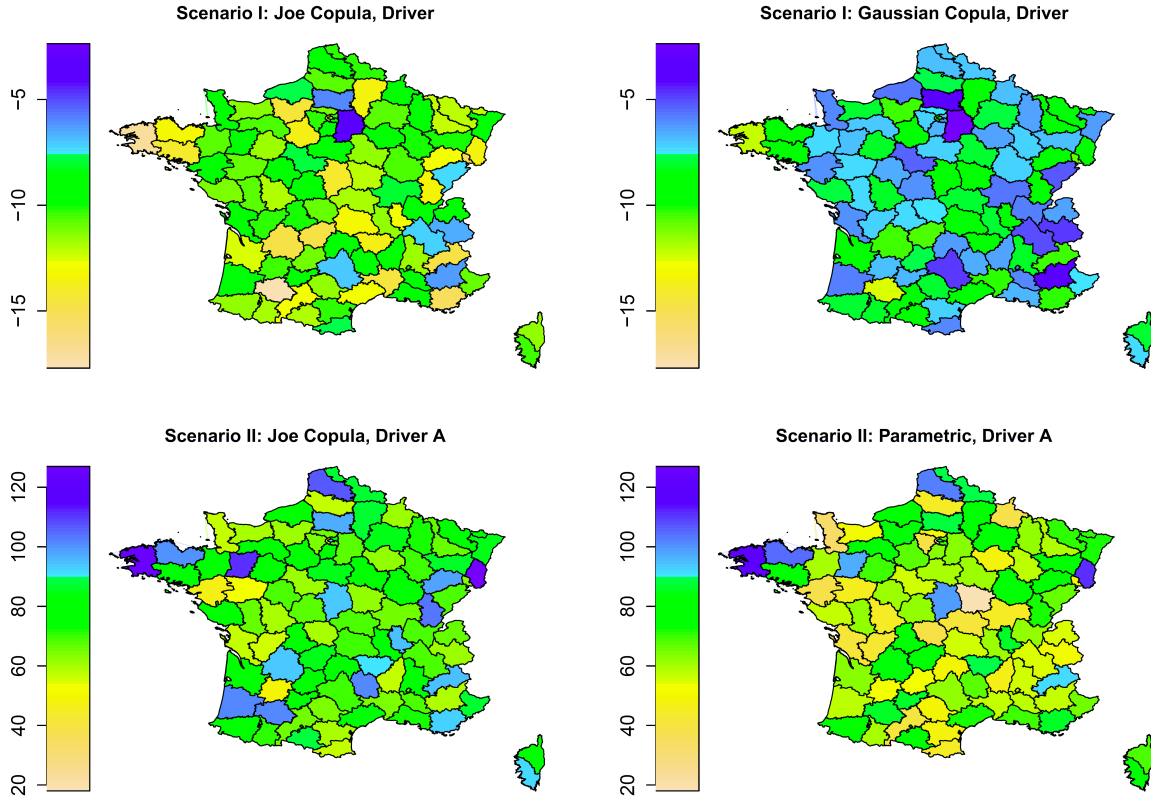
Figure 6: Some potential risks of model mis-specification: comparison between pseudo-elasticities of S curves on hospitalised injuries. The Gaussian copula (not preferred based on the AIC) corresponds to the use of a semi-parametric bivariate ordered probit regression. The parametric model, instead, neglects both non-linearities and smoothed variation in the regional variable. This result highlights the need of using flexible models to reduce the risk of mis-specification. Notice that less vivid conclusions may be drawn when the effects of different covariates are computed: the full set of results are listed in Table 5.

analogous to that described in **?**. This assumption is precisely what we aimed at extending with the introduction of Archimedean copulae (top-left map). The Gaussian model overestimates the effects at departmental level by an average of -3.26 (with minimal and maximal difference of -5.10 and -1.47, respectively). That is, under a bivariate Gaussian assumption, the presence of an S curve reduces the probability of facing a hospitalised injury (resulting from a single-car collision in the average Department) of 3.26% more than what is estimated under a Joe_0. Also, in the fully parametric model specification, drivers' age and time of collision are assumed to affect the responses linearly, as it is typically done in the applied road safety literature (bottom-right map, to be compared with bottom-left map). In this instance, we find that a fully parametric model estimates that S curves increase the probability of hospitalisations by 14.21 percentage points less than a semi-parametric model does. In both cases, these traditional models tend to distort the effects computed by the copula model. Although these considerations may be less vivid for different covariates, we think that the proposed methodology can constitute a valid way for researchers to test their results against different assumptions and scenarios.

5 Discussion

This paper introduced a flexible bivariate regression model for ordinal responses. The proposed framework can account for the presence of common unobserved variables influencing the inter-

relationships between the responses, non-linear covariate effects and non-Gaussian dependences. The model has been described in terms of a copula-based additive extension of the classical Cumulative Link Model, whose representation has taken advantage of a handy penalized GLM form. The estimation algorithm was discussed and all the necessary computational procedures incorporated in the freely-available routine `CopulaCLM` for the `R` environment.

The proposed methodology was motivated by the study of two different scenarios within the road safety literature, in which a bivariate specification has been used to model jointly the injury severity sustained by individuals in vehicle accidents. This methodology is useful in instances were the presence of unobserved factors is likely to affect the injury severity in the parties involved in the accident. Specifically, we dealt with single-car crashes of four-wheel vehicles with one passenger, and collisions of two motorised vehicles. Using information criteria, we selected the Joe_0 copula to represent the dependence structure between the two equations in the model: a positive association between them has been documented. Insights on the role of specific risk factors in the determination of this association have been exploited through the specification of a third equation in the system explaining the behaviour of γ. Some heterogeneity stemming from the people driving specific vehicle types have been documented. The effects that the induced dependence has on the model interpretation have been summarised in Table 5 and Figure 5. In the bottom row of the figure, in particular, we reported the differences a researcher would have incurred if all the individuals involved in the same crash were not modelled separately, but polled together and analysed in the same univariate model. The pseudo-elasticities estimated from the selected Joe_0 were also mapped and compared against those from the Gaussian copula and a fully parametric model specification (Figure 6).

In summary, the proposed application study argued on the importance in accident research to be able to discriminate between different dynamics and occurrences in order to investigate the role of several risk factors in the injury severity sustained by people in vehicle collisions. The need of enhancing model flexibility has been discussed throughout the paper, and the detrimental effects of using standard models *in lieu* of more comprehensive ones have been illustrated (namely, bivariate Gaussian, univariate pooled and purely parametric models).

A possible shortcoming of the copulae employed in this article is that they are exchangeable (**?**; **?**). In the context of two-car accidents, for example, exchangeability implies that the probability of the two drivers to sustain a certain level of injury severity is invariant to whether $\{y_{1,i} = k_1, y_{2,i} = k_2\}$ or $\{y_{2,i} = k_2, y_{1,i} = k_1\}$, conditioned to the regressors, because they both give rise to the same bivariate distribution. Although this seems not to be an issue here, because we expect a symmetric behaviour in the injury severity of the two drivers, it is important to further assess the extent to which this constitutes a limitation in applied research. In addition, since dichotomous responses can be thought as a special case of discrete variables with only two levels, the possibility to allow one dependent variable to be binary will be investigated. This will not change substantially the essence of our theoretical and computational framework. Moreover, the practical limitation of having just one association parameter linking the responses can be relaxed by considering a different class of copulae within our estimation setting. This would in turn allow the study of the dependence structure between the outcomes in a more complete manner.

Supplementary Material   The Supplementary Material is made up of four sections containing more detailed information on the proposed copula setting (Section S.1), the DGP used in simulations (Section S.2), all the results of the analysis of Scenario I (Section S.3) and further details on the asymptotic properties of the proposed estimator described in the Appendix (Section S.4). Datasets employed, codes for replication and a folder with all `CopulaCLM` files are attached to this paper.

A   Some Asymptotic Results

This appendix provides some considerations regarding the behaviour of the proposed Maximum Penalized Likelihood Estimator as the sample size increases. As such, we extend to a bivariate setting of ordinal responses previous works of **?** and, more recently, **?**. The theoretical framework is developed using penalized B-splines to curve approximation (**?**). Their precise formulation is given in the Supplementary Material for the sake of completeness. It worth reminding that each linear predictor is made up of L_j smooth functions of degree $q_j = q$ for all j, no other parametric nor spatial components are included. In this scenario, the j-th linear predictor becomes

$$\boldsymbol{\eta}_j = \boldsymbol{c}_j - s_{j,1}(\mathbf{v}_{j,1}) - \cdots - s_{j,L_j}(\mathbf{v}_{j,L_j}) = \mathbf{Z}_j \boldsymbol{\beta}_j \in \mathbb{R}^n.$$

Assumptions:

(A.1) Every $v_{j,l_j,i}$ is a realisation of the random variable v_{j,l_j}, for $i = 1, \ldots, n$ and, without loss of generality, the vector $\boldsymbol{v}_j := (v_{j,l_j})_{l_j=1,\ldots,L_j}$ is assumed to be distributed on $[0,1]^{L_j}$, $j \in \{1,2,\gamma\}$.

(A.2) The knots for the B-spline basis are equidistantly located so that $\kappa_h - \kappa_{h-1} = H_n^{-1}$ for $h = 1, \ldots, H_n$; moreover, the dimension of the spline basis satisfies $H_n = O(n^{1/(2q+3)})$.

(A.3) The number of knots are chosen so that $[(L_1 + L_2 + L_3)(H_n + q) + K_1 + K_2 - 2] < n$.

(A.4) The smoothing parameters are taken to grow with the sample size with order

$$\max_{j \in \{1,2,\gamma\}; l_j = 1,\ldots,L_j} \lambda_{[j,l_j]} = O(n^\zeta) \quad \text{and} \quad \zeta \leq \frac{2}{2q+3}.$$

It should also be noted that Archimedean copulae are differentiable by construction, thus we only need their partial derivatives to be bounded too. At the outset, we show that expressions in Section 3.2 can be re-stated as:

Lemma 1. *The unpenalized Hessian matrix for a bivariate copula Cumulative Link Additive Model in the form of (4) is equivalent to*

$$\nabla_{\boldsymbol{\vartheta}\boldsymbol{\vartheta}^\top} \ell(\boldsymbol{\vartheta}) = \tilde{\mathbf{D}}^\top \tilde{\mathbf{W}} \tilde{\mathbf{D}},$$

where $\tilde{\mathbf{D}} := \text{diag}(\tilde{\mathbf{D}}_{k_1}, \tilde{\mathbf{D}}_{k_2}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_\gamma) \in \mathbb{R}^{5n \times p}$, $\tilde{\mathbf{D}}_{k_j} := \mathbf{1} \otimes \boldsymbol{d}_{k_j}^\top$, $j = 1, 2$, *and*

$$\tilde{\mathbf{W}} := \begin{bmatrix} \tilde{\mathbf{W}}_{k_1} & \tilde{\mathbf{W}}_{k_1,k_2} & \tilde{\mathbf{W}}_{k_1,1} & \tilde{\mathbf{W}}_{k_1,2} & \tilde{\mathbf{W}}_{k_1,\gamma} \\ \bullet & \tilde{\mathbf{W}}_{k_2} & \tilde{\mathbf{W}}_{k_2,1} & \tilde{\mathbf{W}}_{k_2,2} & \tilde{\mathbf{W}}_{k_2,\gamma} \\ \bullet & \bullet & \tilde{\mathbf{W}}_1 & \tilde{\mathbf{W}}_{1,2} & \tilde{\mathbf{W}}_{1,\gamma} \\ \bullet & \bullet & \bullet & \tilde{\mathbf{W}}_2 & \tilde{\mathbf{W}}_{2,\gamma} \\ \bullet & \bullet & \bullet & \bullet & \tilde{\mathbf{W}}_\gamma \end{bmatrix} \in \mathbb{R}^{5n \times 5n}$$

is the symmetric matrix whose generic element is given by $\tilde{\mathbf{W}}_{\bar{h}} := \text{diag}(w_{\bar{h},i})_i \in \mathbb{R}^{n \times n}$, *for* $\bar{h}$ *encompassing all the subscripts in* $\tilde{\mathbf{W}}$. *Namely,* $\bar{h} := \{k_1, (k_1, k_2), \ldots, \gamma\}$. *The remaining quantities are computed as follows:*

$$\boldsymbol{d}_{k_j} := ((2\tilde{c}_{k_j}^2)^{\mathbb{1}_{k_j \neq 1}})_{k_j = 1, \ldots, \mathcal{K}_j \setminus \{K_j\}},$$

- $\bar{h} \in \{k_j\}$, $j \in \{1,2\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}}$$

$$\left[\left(\frac{\partial \Phi(\eta_{j,k_j,i})}{\partial \eta_{j,k_j,i}} \right)^2 \left(\mathcal{C}_{0,0,i}^{(k_j^2)} - \mathcal{C}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i}^{(k_j^2)} \right) + \frac{\partial^2 \Phi(\eta_{j,k_j,i})}{\partial \eta_{j,k_j,i}^2} \left(\mathcal{C}_{0,0,i}^{(k_j^2)} - \mathcal{C}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i}^{(k_j^2)} \right) \right] -$$

$$\boldsymbol{d}_{k_j}^\top (\boldsymbol{d}_{k_j} \boldsymbol{d}_{k_j}^\top)^{-1} \boldsymbol{d}_{k_j,k_j} (\boldsymbol{d}_{k_j} \boldsymbol{d}_{k_j}^\top)^{-\top} \boldsymbol{d}_{k_j} u_{k_j,i};$$

- $\bar{h} \in \{(k_{\tilde{\jmath}}, k_{\bar{\jmath}})\}$, $\tilde{\jmath}, \bar{\jmath} \in \{1, 2\}$, $\tilde{\jmath} \neq \bar{\jmath}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \mathcal{C}_{0,0,i}^{(k_{\tilde{\jmath}}, k_{\bar{\jmath}})} \frac{\partial \Phi(\eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i})}{\partial \eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i}} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}, i}};$$

- $\bar{h} \in \{(k_j, j)\}$, $j \in \{1, 2\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}}$$
$$\left[\left(\frac{\partial \Phi(\eta_{j, k_j, i})}{\partial \eta_{j, k_j, i}} \right)^2 \left(\mathcal{C}_{0,0,i}^{(k_j^2)} - \mathcal{C}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i}^{(k_j^2)} \right) + \frac{\partial^2 \Phi(\eta_{j, k_j, i})}{\partial \eta_{j, k_j, i}^2} \left(\mathcal{C}_{0,0,i}^{(k_j^2)} - \mathcal{C}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i}^{(k_j^2)} \right) \right];$$

- $\bar{h} \in \{(k_{\tilde{\jmath}}, \bar{\jmath})\}$, $\tilde{\jmath}, \bar{\jmath} \in \{1, 2\}$, $\tilde{\jmath} \neq \bar{\jmath}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \frac{\partial \Phi(\eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i})}{\partial \eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i}} \left(\mathcal{C}_{0,0,i}^{(k_{\tilde{\jmath}}, k_{\bar{\jmath}})} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}, i}} - \mathcal{C}_{\mathbb{1}_{\tilde{\jmath}=1}-1, \mathbb{1}_{\tilde{\jmath}=2}-1, i}^{(k_{\tilde{\jmath}}-\mathbb{1}_{\bar{\jmath}=2}, k_{\bar{\jmath}}-\mathbb{1}_{\bar{\jmath}=1})} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i}} \right);$$

- $\bar{h} \in \{(k_j, \gamma)\}$, $j \in \{1, 2\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \frac{\partial \gamma_i}{\partial \gamma_i^*} \frac{\partial \Phi(\eta_{j, k_j, i})}{\partial \eta_{j, k_j, i}} \left(\mathcal{C}_{0,0,i}^{(k_j), \gamma} - \mathcal{C}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i}^{(k_j), \gamma} \right)$$

- $\bar{h} \in \{1, 2\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \sum_{l,m \in \{0,1\}} (-1)^{l+m} \left[\mathcal{C}_{-l,-m,i}^{((k_{\bar{h}}-l)^2)} \left(\frac{\partial \Phi(\eta_{\bar{h}, k_{\bar{h}}-l, i})}{\partial \eta_{\bar{h}, k_{\bar{h}}-l, i}} \right)^2 + \mathcal{C}_{-l,-m,i}^{(k_{\bar{h}}-l)} \frac{\partial^2 \Phi(\eta_{\bar{h}, k_{\bar{h}}-l, i})}{\partial \eta_{\bar{h}, k_{\bar{h}}-l, i}^2} \right] - \mathrm{u}_{\bar{h},i}^2;$$

- $\bar{h} \in \{\tilde{\jmath}, \bar{\jmath}\}$, $\tilde{\jmath}, \bar{\jmath} \in \{1, 2\}$, $\tilde{\jmath} \neq \bar{\jmath}$:

$$\begin{aligned} w_{\bar{h},i} = \ & \frac{1}{\pi_{k,i}} \left[\frac{\partial \Phi(\eta_{\tilde{\jmath}, k_{\tilde{\jmath}}-1, i})}{\partial \eta_{\tilde{\jmath}, k_{\tilde{\jmath}}-1, i}} \left(\mathcal{C}_{-1,-1,i}^{(k_{\tilde{\jmath}}-1, k_{\bar{\jmath}}-1)} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i}} - \mathcal{C}_{-1,0,i}^{(k_{\tilde{\jmath}}-1, k_{\bar{\jmath}})} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}, i}} \right) + \right. \\ & \left. \frac{\partial \Phi(\eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i})}{\partial \eta_{\tilde{\jmath}, k_{\tilde{\jmath}}, i}} \left(\mathcal{C}_{0,0,i}^{(k_{\tilde{\jmath}}, k_{\bar{\jmath}})} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}, i}} - \mathcal{C}_{0,-1,i}^{(k_{\tilde{\jmath}}, k_{\bar{\jmath}}-1)} \frac{\partial \Phi(\eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i})}{\partial \eta_{\bar{\jmath}, k_{\bar{\jmath}}-1, i}} \right) \right] - \mathrm{u}_{\tilde{\jmath},i} \mathrm{u}_{\bar{\jmath},i}; \end{aligned}$$

- $\bar{h} \in \{(j, \gamma)\}$, $j \in \{1, 2\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \frac{\partial \gamma_i}{\partial \gamma_i^*}$$
$$\left[\frac{\partial \Phi(\eta_{j, k_j-1, i})}{\partial \eta_{j, k_j-1, i}} \left(\mathcal{C}_{-1-1,i}^{(k_j-1), \gamma} - \mathcal{C}_{-1,0,i}^{(k_j-1), \gamma} \right) + \frac{\partial \Phi(\eta_{j, k_j, i})}{\partial \eta_{j, k_j, i}} \left(\mathcal{C}_{0,0,i}^{(k_j), \gamma} - \mathcal{C}_{0,-1,i}^{(k_j), \gamma} \right) \right] - \mathrm{u}_{j,i} \mathrm{u}_{\gamma,i};$$

- $\bar{h} \in \{\gamma\}$:

$$w_{\bar{h},i} = \frac{1}{\pi_{k,i}} \sum_{l,m \in \{0,1\}} (-1)^{l+m} \left[\mathcal{C}_{-l,-m,i}^{((\bar{h}-l)^2)} \left(\frac{\partial \gamma_i}{\partial \gamma_i^*} \right)^2 + \mathcal{C}_{-l,-m,i}^{(\bar{h}-l)} \frac{\partial^2 \gamma_i}{\partial \gamma_i^{*2}} \right] - \mathrm{u}_{\bar{h},i}^2.$$

Moreover, the next quantities have been used:

$$\boldsymbol{d}_{k_j, k_j} := \frac{\partial \boldsymbol{d}_{k_j}}{\partial \tilde{\boldsymbol{c}}_{k_j}}, \quad \mathcal{C}_{a,b,i}^{(c,d)} := \frac{\partial^2 \mathcal{C}(u-a, v-b; \gamma_i)}{\partial \eta_{c,i} \partial \eta_{d,i}},$$

for u and v being the inputs of the copula function and $a, b \in \{0, 1\}$; also, for $j \in \{1, 2\}$,

$$\mathrm{u}_{k_j,i} := \frac{1}{\pi_{k,i}} \frac{\partial \Phi(\eta_{j,k_j,i})}{\partial \eta_{j,k_j,i}} \left(\mathcal{C}^{(k_j)}_{0,0,i} - \mathcal{C}^{(k_j)}_{\mathbb{1}_{j=1}-1, \mathbb{1}_{j=2}-1, i} \right)$$

$$\mathrm{u}_{j,i} := \frac{1}{\pi_{k,i}} \sum_{l,m \in \{0,1\}} (-1)^{l+m} \mathcal{C}^{(k_j)}_{-l,-m,i} \frac{\partial \Phi(\eta_{j,k_j,i})}{\partial \eta_{j,k_j,i}}$$

$$\mathrm{u}_{\gamma,i} := \frac{1}{\pi_{k,i}} \frac{\partial \gamma_i}{\partial \gamma_i^*} \sum_{l,m \in \{0,1\}} (-1)^{l+m} \mathcal{C}^{(k_j)}_{-l,-m,i}.$$

Proof. The proof is given in the Supplementary Material. ∎

Let us define now $\mathbf{F}_n(\boldsymbol{\vartheta}_0)$ be the expected Fisher Information evaluated at $\boldsymbol{\vartheta}_0$. This denotes the parameter vector that gives the best approximation of the unknown smooth functions s_{j,l_j}'s in terms of the Kullback-Leibler measure. Specifically, we have that $\mathbf{F}_n(\boldsymbol{\vartheta}_0) = \tilde{\mathbf{D}}^\top \mathbb{E}_{\boldsymbol{\vartheta}_0}[\tilde{\mathbf{W}}] \tilde{\mathbf{D}}$ and the following result applies:

Lemma 2 (Lemma 2, **?**)**.** *For a model expressible in the form of (4) and under Assumptions (A.1)-(A.4), the elements of $\mathbf{F}_n(\boldsymbol{\vartheta}_0)$ are of asymptotic order $O(n/H_n)$. Moreover, denoting by $\mathbf{F}_{\mathrm{p},n}(\boldsymbol{\vartheta}_0)$ the penalized Fisher Information, it follows that*

$$\mathbf{F}_{\mathrm{p},n}(\boldsymbol{\vartheta}_0) = O\left(\frac{n}{H_n} + \max_{j \in \{1,2,\gamma\}; l_j=1,\dots,L_j} \lambda_{[j,l_j]} H_n^{2q} \right).$$

Proof. The proof is given in the Supplementary Material. ∎

These preliminary results allow us to characterise the order of consistency of the MPLE with respect to $\boldsymbol{\vartheta}_0$.

Theorem 1. *Consider a bivariate copula Cumulative Link Additive Model with non-parametric components approximated by penalized B-splines. Let $\widehat{\boldsymbol{\vartheta}}$ be the Maximum Penalized Likelihood Estimator and $\boldsymbol{\vartheta}_0$ such that $\mathbb{E}[\nabla_{\boldsymbol{\vartheta}} \ell(\boldsymbol{\vartheta})]|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} = \mathbf{0}$. Then, under Assumptions (A.1)-(A.4), it holds that*

$$\sqrt{n/H_n}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}_0) = o_P(1).$$

Proof. In the proceeding arguments we adopt the Einstein summation convention, and we denote accordingly by ϑ^j the j-th component of the parameter vector $\boldsymbol{\vartheta} := (\vartheta^1, \dots, \vartheta^p)^\top$. To prove the statement, we first expand $\widehat{\ell}_{\mathrm{p},r}$ about $\ell_{\mathrm{p},r}$ to get

$$\widehat{\ell}_{\mathrm{p},r} = \ell_{\mathrm{p},r} + \ell_{\mathrm{p},rs}(\widehat{\vartheta} - \vartheta_0)^s + \frac{1}{2} \ell_{\mathrm{p},rst}(\widehat{\vartheta} - \vartheta_0)^{st} + \dots = 0,$$

where $\widehat{\ell}_{\mathrm{p},r} := \partial \ell_{\mathrm{p}} / \partial \vartheta^r |_{\vartheta^r = \vartheta_0^r}$ and, similarly, $\ell_{\mathrm{p},rs} := \partial^2 \ell_{\mathrm{p}} / \partial \vartheta^r \partial \vartheta^s$. In the expression above we have defined $(\widehat{\vartheta} - \vartheta_0)^s := \widehat{\vartheta}^s - \vartheta_0^s$ and $(\widehat{\vartheta} - \vartheta_0)^{st} := (\widehat{\vartheta} - \vartheta_0)^s (\widehat{\vartheta} - \vartheta_0)^t$. Series inversion yields

$$(\widehat{\vartheta} - \vartheta_0)^r = -\ell_{\mathrm{p}}^{rs} \ell_{\mathrm{p},s} - \frac{1}{2} \ell_{\mathrm{p}}^{rtv} \ell_{\mathrm{p},u} \ell_{\mathrm{p},w} + \dots = -\ell_{\mathrm{p}}^{rs} \ell_{\mathrm{p},s} - \frac{1}{2} (\ell_{\mathrm{p}}^{rs} \ell_{\mathrm{p}}^{tu} \ell_{\mathrm{p}}^{vw} \ell_{\mathrm{p},stv}) \ell_{\mathrm{p},u} \ell_{\mathrm{p},w} + \dots,$$

with ℓ_{p}^{rs} denoting the (r, s)-th element of the inverse of the Hessian matrix given in Lemma 1. Let us define now

$$f_{rs,0} := -\mathbb{E}[\partial^2 \ell / \partial \vartheta^r \partial \vartheta^s]|_{\vartheta^r = \vartheta_0^r; \vartheta^s = \vartheta_0^s},$$
$$f_{\mathrm{p},rs} = f_{rs,0} + q_{rs}$$

be the (r, s)-th element of the (unpenalized) Fisher Information and the corresponding penalized version, respectively. Notice that the latter carries the asymptotic order derived in Lemma 2.

Upon denoting $r_{rs} := \ell_{rs} - f_{rs,0}$, the difference between the given elements of the Hessian and the Fisher Information matrices, we have

$$\ell_{\mathrm{p}}^{rs} = f_{\mathrm{p}}^{rs} - f_{\mathrm{p}}^{rt} f_{\mathrm{p}}^{su} r_{tu}. \tag{7}$$

In particular, applying a similar argument as **?**, and making use of the results of **?** for the inverse of band matrices, it is possible to bound

$$f_{\mathrm{p}}^{rs} = \rho^{|r-s|} O\left(\left(\frac{n}{H_n} + \max_{j \in \{1,2,\gamma\}; l_j = 1, \ldots, L_j} \lambda_{[j,l_j]} H_n^{2q} \right)^{-1} \right) = \rho^{|r-s|} O\left(\frac{H_n}{n} \right)$$

for $\rho \in (0,1)$. The last equality is established under Assumptions (A.2) and (A.4). Furthermore, the direct application of Lemma 4 in **?** results in the matrix with element r_{rs} being block diagonal, with each $r_{rs} = O_P(\sqrt{n/K_n})$. Hence expression (7) becomes

$$\ell_{\mathrm{p}}^{rs} = f_{\mathrm{p}}^{rs} \left[1 + O\left(\frac{H_n}{n} \right) O_P\left(\sqrt{\frac{n}{H_n}} \right) \right] = f_{\mathrm{p}}^{rs} \left[1 + O_P\left(\sqrt{\frac{H_n}{n}} \right) \right].$$

Analogously, $\ell_{\mathrm{p},stv}$ is of diagonal structure, meaning that is it equals 0 if $\max\{|s-t|, |s-v|, |t-v|\} > q+1$ and is of order $O_P(n/H_n)$ otherwise. Hence

$$\ell_{\mathrm{p}}^{rs} \ell_{\mathrm{p}}^{tu} \ell_{\mathrm{p}}^{vw} \ell_{\mathrm{p},stv} = f_{\mathrm{p}}^{rs} f_{\mathrm{p}}^{tu} f_{\mathrm{p}}^{vw} O_P\left(\frac{H_n^{-1}}{n^{-1}} \right) (1 + o_P(1)) = O_P\left(\frac{H_n^2}{n^2} \right) (1 + o_P(1))$$

and $\ell_{\mathrm{p}}^{rtv} \ell_{\mathrm{p},u} \ell_{\mathrm{p},w} = o_P(H_n/n)$. We can finally write

$$(\widehat{\vartheta} - \vartheta_0)^r = -f_{\mathrm{p}}^{rs} \ell_{\mathrm{p},s} (1 + o_P(1)) + o_P\left(\frac{H_n}{n} \right) = o_P\left(\sqrt{\frac{H_n}{n}} \right) + o_P\left(\frac{H_n}{n} \right),$$

and the assertion follows. ∎

With the stated result in hand, we can quantify the error a researcher would commit in estimating the linear predictor $\boldsymbol{\eta}_j$ using the B-splines approximation:

Corollary 1 (Theorem 1, **?**)**.** *Under Assumptions (A.1)-(A.4) it holds that, for each j,*

$$\mathrm{MSE}\{\widehat{\boldsymbol{\eta}}_j\} := \mathbb{E}\|\widehat{\boldsymbol{\eta}}_j - \boldsymbol{\eta}\|^2 = O(n^{-(2q+2)/(2q+3)}),$$

where $\widehat{\boldsymbol{\eta}}_j$ is evaluated at the MPLE $\widehat{\boldsymbol{\vartheta}}$.

Supplementary Material to: "Simultaneous Equation Penalized Likelihood Estimation of Vehicle Accident Severity Injuries"

Francesco Donat and Giampiero Marra

S.1 Copulae

Rotated copulae can be obtained by applying the following transformations:

$$
\begin{aligned}
\mathcal{C}_{90}(u, v) &= u - \mathcal{C}_\gamma(1 - u, v) \\
\mathcal{C}_{180}(u, v) &= u + v - 1 + \mathcal{C}_\gamma(1 - u, 1 - v) \\
\mathcal{C}_{270}(u, v) &= u - \mathcal{C}_\gamma(u, 1 - v),
\end{aligned}
$$

where we have followed the convention of labelling the marginals corresponding to $j = 1, 2$ with u and v, respectively. Contour plots of the copuae implemented in `CopulaCLM` are given in Figure S1.
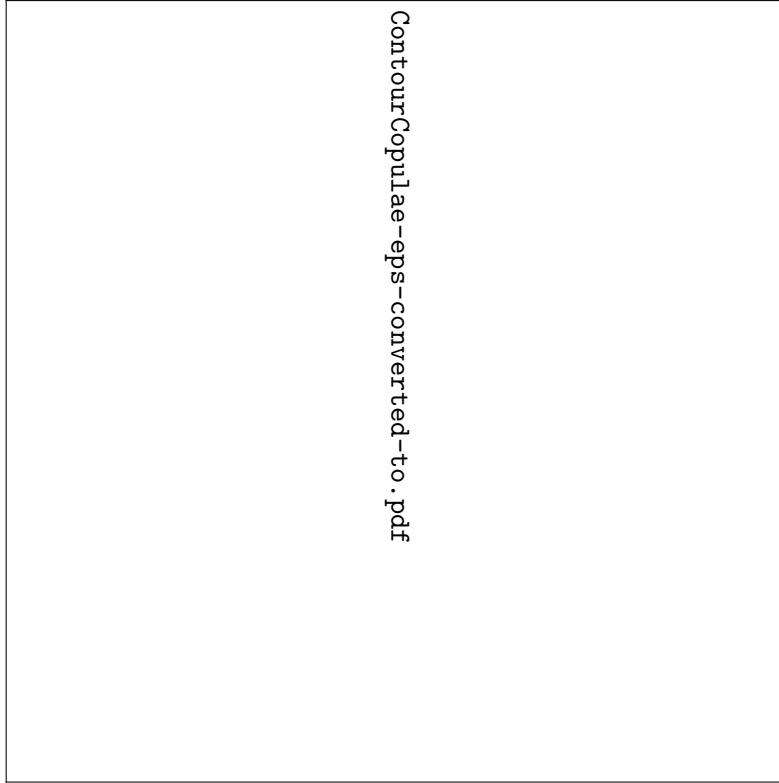
Figure S1: Contour plots of some of the copula functions with Standard Normal margins for data simulated using association parameters γ of 2, 5.74, 2 and 2.86, respectively (these values are consistent with a medium positive correlation). The Frank copula allows for equal degrees of positive and negative dependence, whereas Clayton is asymmetric with a strong lower tail dependence but a weaker upper tail dependence. Vice versa for the Gumbel and Joe copulae.

S.2 Data Generating Processes Employed in Simulations

DGP for Figure 2

$$
\begin{aligned}
y_{1,i}^* &= x_{1,i} + 2x_{2,i} + x_{3,i} + s_{1,1}(v_{1,i}) + s_{1,2}(v_{2,i}) + \epsilon_1 \\
y_{2,i}^* &= 2x_{1,i} - 2x_{2,i} + s_{2,1}(v_{1,i}) + \epsilon_2
\end{aligned} \qquad \epsilon_j \sim \mathcal{N}(0, 1).
$$

The test functions are given by

$$\begin{aligned}
s_{1,1}(v_{1,i}) &= -0.7\{4v_{1,i} + 2.5v_{1,i}^2 + 0.7\sin(5v_{1,i}) + \cos(7.5v_{1,i})\} \\
s_{1,2}(v_{2,i}) &= -0.4\{-0.3 - 1.6v_{2,i} + \sin(5v_{2,i})\} \\
s_{2,1}(v_{1,i}) &= 0.6\{\exp\{v_{1,i}\} + \sin(2, 9v_{1,i})\},
\end{aligned}$$

and the ordered values of $y_{j,i}$ have been computed following the observation rule

$$y_{j,i} = \sum_{k_j \in \mathcal{K}_j} k_j \mathbb{1}_{c_{j,k_j-1} < y_{j,i}^* \le c_{j,k_j}}$$

for every $j \in \{1, 2\}$, and obtained by setting the threshold parameters at $\boldsymbol{c}_{1,k} := (-2, -1, 0, 2)^\top$ and $\boldsymbol{c}_{2,k} := (-1.4, -0.7, -0.2, 0.7, 3)^\top$. Furthermore, we have set the copula association parameters at $\gamma_{\text{clayton}} = 0.2222$, $\gamma_{\text{frank}} = 0.9074$ and $\gamma_{\text{joe}} = 1.1944$, all corresponding to a Kendall's Tau of 0.1.

DGP for Figure 3 The same as of the previous paragraph but with smooth curves

$$\begin{aligned}
s_{1,1}(v_{1,i}) &= 1 - v_{1,i} + 1.6v_{1,i}^4 - \sin(5v_{1,i}) \\
s_{1,2}(v_{2,i}) &= 4v_{2,i} \\
s_{2,1}(v_{1,i}) &= 0.08\{v_{1,i}^{11}[10(1 - v_{1,i})]^6\} + 10(10v_{1,i})(1 - v_{1,i})^{10},
\end{aligned}$$

and cut points $\boldsymbol{c}_{1,k} := (-0.8, -0.3, 0.6, 4)^\top$ and $\boldsymbol{c}_{2,k} := (-1.4, -0.7, -0.2, 0.7, 3)^\top$.

S.3 Data Analysis: Further Details

	Joe$_0$ model				Independent model			
	Driver		Other occupant		Driver		Other occupant	
Variables	estimates	(se)	estimates	(se)	estimates	(se)	estimates	(se)
Occupant Characteristics								
Gender (male)								
female	0.1697	(0.0694)	0.0625	(0.0630)	0.1293	(0.0714)	0.0556	(0.0645)
Seat (other/missing)								
front, passenger	–	–	0.2048	(0.1664)	–	–	0.0665	(0.1687)
rear, driver's side	–	–	0.0364	(0.2344)	–	–	−0.0802	(0.2402)
rear, opposite driver	–	–	−0.0036	(0.2171)	–	–	0.0187	(0.2214)
Motorway Characteristics								
Intersection (off intersection)								
X	0.0395	(0.2358)	−0.2162	(0.2428)	−0.0219	(0.2435)	−0.2347	(0.2467)
T	−0.0924	(0.2535)	−0.3344	(0.2590)	−0.1526	(0.2636)	−0.3319	(0.2611)
Y	−0.2814	(0.3907)	−0.3344	(0.3682)	−0.3370	(0.4052)	−0.3527	(0.3764)
> 4 branches	−0.8819	(0.4609)	−0.4818	(0.4177)	−0.9907	(0.4903)	−0.4882	(0.4221)
roundabout	0.3647	(0.2109)	0.3174	(0.2188)	0.3367	(0.2166)	0.3223	(0.2196)
circus/square	0.1333	(1.0256)	−0.4455	(1.0725)	0.1483	(1.0507)	−0.4455	(1.1038)
other	0.3505	(0.2982)	0.2448	(0.3088)	0.3213	(0.3094)	0.2516	(0.3082)
Type (motorway)								
Route Nationale	−0.1418	(0.1333)	−0.0205	(0.1336)	−0.1639	(0.1363)	−0.0415	(0.1359)
Route Départementale	0.0070	(0.0981)	0.1478	(0.0993)	0.0066	(0.1001)	0.1435	(0.1009)
Voie Communale	−0.3588	(0.1056)	−0.0741	(0.1051)	−0.3595	(0.1076)	−0.0712	(0.1073)
other	−0.4516	(0.3301)	−0.3340	(0.3357)	−0.4898	(0.3437)	−0.3417	(0.3387)
Circulation regime (missing)								
one-way	0.0751	(0.1530)	−0.3744	(0.1574)	0.0943	(0.1570)	−0.4131	(0.1584)
two-way	0.3640	(0.1291)	0.1020	(0.1338)	0.3908	(0.1323)	0.0813	(0.1344)
presence of median	0.1197	(0.1430)	−0.1610	(0.1485)	0.1503	(0.1469)	−0.1744	(0.1494)
other	1.0587	(0.5738)	0.1987	(0.5722)	1.0814	(0.5777)	0.0835	(0.5878)
Horizontal alignment (straight)								
left curve	−0.1137	(0.0800)	0.2048	(0.0812)	−0.1311	(0.0816)	0.2102	(0.0820)
right curve	0.0004	(0.0892)	0.0364	(0.0890)	−0.0163	(0.0905)	0.0190	(0.0907)
S	−0.1379	(0.1665)	−0.0036	(0.1657)	−0.0956	(0.1664)	0.0154	(0.1681)
Location (other/missing)								
roadway	−0.4146	(0.1080)	−0.4878	(0.1103)	−0.4539	(0.1095)	−0.5009	(0.1119)
emergency lane	−0.3953	(0.2359)	0.0222	(0.2395)	0.4926	(0.2420)	0.0128	(0.2407)
shoulder	−0.0446	(0.1203)	−0.0932	(0.1221)	−0.0833	(0.1215)	−0.1110	(0.1239)
sidewalk	−0.2552	(0.1864)	−0.6954	(0.1892)	−0.3016	(0.1895)	−0.7241	(0.1924)
Obstacle (other/missing)								
fixed object	0.1938	(0.0887)	0.0984	(0.0904)	0.2044	(0.0907)	0.1057	(0.0915)
pedestrian	−7.2444	(5.9 10^5)	−0.9181	(1.1237)	−7.8729	(3.8175)	−0.8593	(1.1256)
vehicle	−0.1082	(0.1832)	−0.0178	(0.1818)	−0.1191	(0.1883)	−0.0177	(0.1847)
animal	0.1777	(0.3076)	0.0467	(0.3031)	0.2189	(0.3092)	0.0365	(0.3114)
Accident Characteristics								
Lighting (daylight)								
sunrise/sunset	0.0164	(0.1188)	0.1235	(0.1210)	−0.0280	(0.1219)	0.1255	(0.1218)
night without street lights	0.1814	(0.0801)	−0.0430	(0.0788)	0.1591	(0.0824)	−0.0575	(0.0801)
night, street lights in force	−0.0173	(0.0910)	0.1334	(0.0905)	−0.0444	(0.0941)	0.1340	(0.0920)
Atmospheric conditions (normal)								
light rain	−0.2105	(0.0889)	−0.1271	(0.0880)	−0.2097	(0.0904)	−0.1453	(0.0897)
heavy rain	0.4468	(0.1658)	−0.1548	(0.1678)	0.4517	(0.1681)	−0.1743	(0.1714)
snow	0.6575	(0.4330)	−0.6317	(0.4231)	0.8153	(0.4157)	−0.5592	(0.4271)
fog	−0.2960	(0.2864)	0.1544	(0.2890)	−0.2323	(0.2874)	0.1490	(0.2935)
heavy wind/storm	0.0588	(0.4496)	1.1834	(0.4899)	0.0722	(0.4512)	1.1485	(0.4979)
clear	1.0461	(0.4596)	0.6979	(0.4539)	1.0211	(0.4686)	0.6613	(0.4646)
clouds	0.1347	(0.1432)	−0.1661	(0.1458)	0.1352	(0.1461)	−0.1528	(0.1468)
Manner of collision (missing/other)								
heads-on	−0.1864	(0.0761)	0.0176	(0.0773)	−0.1728	(0.0773)	0.0388	(0.0781)
rear-end	−0.3630	(0.1981)	−0.1348	(0.1971)	−0.3834	(0.2054)	−0.1048	(0.1997)
sideswipe, right	−0.1723	(0.1739)	0.7225	(0.1774)	−0.1673	(0.1758)	0.7382	(0.1796)
sideswipe, left	0.6712	(0.1677)	−0.0826	(0.1648)	0.6937	(0.1684)	−0.0902	(0.1682)
Security device (not put on)								
put on	−0.4585	(0.0785)	−0.2646	(0.0771)	−0.4282	(0.0805)	−0.2304	(0.0789)
$c_{j,1}$	−0.9966	(0.2097)	−1.6341	(0.2814)	−0.9943	(0.2142)	−1.6197	(0.2852)
$c_{j,2}$	−0.0786	(0.0196)	−0.2296	(0.0198)	−0.0788	(0.0198)	−0.2111	(0.0199)
$c_{j,3}$	1.3514	(0.0271)	1.3742	(0.0243)	1.3500	(0.0281)	1.3936	(0.0248)
No. observations	1,232		1,232		1,232		1,232	

Table S1: Estimates and associated standard errors (in brackets) obtained by applying `CopulaCLM` to the BAAC 2014 data. Results are for the parametric model components of Scenario I, when using the Joe$_0$ copula function is used. The last columns report the results corresponding to the independent model. The reference categories are given in round brackets next to the variable names to which they refer.
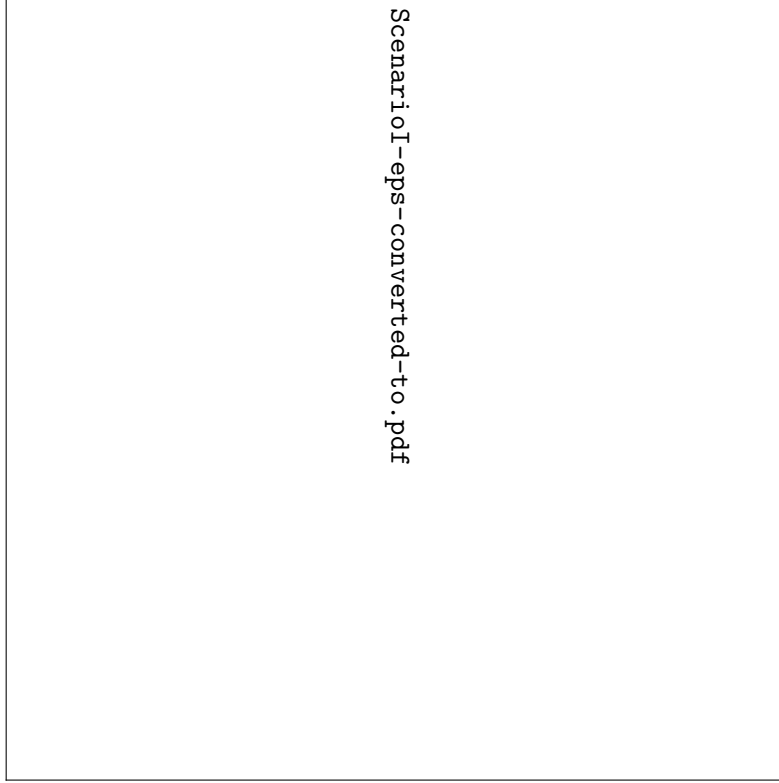
Figure S2: Smooth functions estimates and associated 95% point-wise confidence intervals corresponding to the two equations (first and second row, respectively) of the bivariate model applied to the BAAC 2014 data under Scenario I, when the Joe_0 copula function is used. The maps quantify the estimates obtained for the regional variable in each of the 96 Departments of continental France. We refer to the caption of Figure 4 for further details.

S.4 Asymptotic Considerations

Definition of B-splines of order q On the closed interval $[0,1]$ we define the sequence of knot points $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_H$ and further $2q$ knots: $\kappa_H = \kappa_{H+1} = \cdots = \kappa_{H+q}$ and $\kappa_{-p+1} = \kappa_{-p+2} = \cdots = \kappa_{-1} = \kappa_0$. The basis functions are then derived recursively as:

$$B_{h,q}(\mathrm{v}_{j,l_j,i}) = \frac{\mathrm{v}_{j,l_j,i} - \kappa_{h-1}}{\kappa_{h+q-1} - \kappa_{h-1}} B_{h,q-1}(\mathrm{v}_{j,l_j,i}) + \frac{\kappa_{h+q} - \mathrm{v}_{j,l_j,i}}{\kappa_{h+q} - \kappa_h} B_{h+1,q-1}(\mathrm{v}_{j,l_j,i}),$$

for $h = -q+1, \ldots, H$, with

$$B_{h,0}(\mathrm{v}_{j,l_j,i}) = \left\{ \begin{array}{ll} 1 & \kappa_{h-1} \leq \mathrm{v}_{j,l_j,i} < \kappa_h \\ 0 & o/w \end{array} \right. .$$

The above construction defines precisely $H+q$ basis functions: $B_{-q+1,q}(\mathrm{v}_{j,l_j,i}), \ldots, B_{H,q}(\mathrm{v}_{j,l_j,i})$. Each evaluation $s_{j,l_j}(\mathrm{v}_{j,l_j,i})$ is then approximated by

$$s_{j,l_j}(\mathrm{v}_{j,l_j,i}) \approx \sum_{h=-q+1}^{H} \beta_{j,l_j} B_{h,q}(\mathrm{v}_{j,l_j,i}),$$

and we write the penalized log-likelihood as

$$\ell_{\mathrm{p}}(\boldsymbol{\vartheta}|\cdot) := \ell(\boldsymbol{\vartheta}|\cdot) - \frac{1}{2} \boldsymbol{\vartheta}^{\top} \mathbf{S}_{\boldsymbol{\lambda}} \boldsymbol{\vartheta},$$

where $\mathbf{S}_{\boldsymbol{\lambda}} := \mathrm{diag}(\mathbf{0}_{K_1+K_2-1}, \mathbf{Q}_{1,m}, \mathbf{Q}_{2,m}, \mathbf{Q}_{\gamma,m})$, $\mathbf{Q}_{j,m} := \mathrm{diag}(\lambda_{j,1} \Delta_m^{\top} \Delta_m, \ldots, \lambda_{j,L_j} \Delta_m^{\top} \Delta_m)$ and Δ_m is the $(H+q-m) \times (H+q)$ m-th difference matrix.

Lemma 1

Proof. The results follow from tedious computation after having noticed that the elements of the gradient vector can be written as

$$\frac{\partial \ell_i}{\partial \tilde{\boldsymbol{c}}_{k_j}} = \boldsymbol{d}_{k_j} \mathrm{u}_{k_j, i}, \quad \frac{\partial \ell_i}{\partial \boldsymbol{\beta}_j} = -\mathbf{x}_{j,i} \mathrm{u}_{j,i}, \quad \frac{\partial \ell_i}{\partial \boldsymbol{\beta}_\gamma} = \mathbf{x}_{\gamma,i} \mathrm{u}_{\gamma,i}$$

for $j \in \{1, 2\}$. ∎

Lemma 2

Proof. Since we work under the same maintained assumptions, Lemma 1 in **?** applies. Therefore we have

$$\begin{array}{rcl} \mathbf{X}_{j,l_j}^\top \mathbf{X}_{j,l_j} & = & O(n/H_n) \quad \text{for } j \in \{1, 2, \gamma\} \\ \mathbf{X}_{j,l_j}^\top \mathbf{X}_{j,\bar{l}_j} & = & O(n/H_n^2) \quad \text{for } j \in \{1, 2, \gamma\} \text{ and } l_j \neq \bar{l}_j \end{array}$$

and $\mathbf{X}_j^\top \mathbf{X}_{\bar{j}}$ is globally of asymptotic order $O(n/H_n)$ for all $j, \bar{j} \in \{1, 2, \gamma\}$. Moreover, we note that the matrices $\tilde{\mathbf{D}}_{k_1}$ and $\tilde{\mathbf{D}}_{k_2}$ are made up of finite constant terms only, so each is of order $O(1)$. As a consequence, $\tilde{\mathbf{D}}_k^\top \tilde{\mathbf{D}}_{\bar{k}}$ and $\tilde{\mathbf{D}}_k^\top \mathbf{X}_j$ are also bounded for $k, \bar{k} \in \{k_1, k_2\}$ and every j.

We turn now to the elements of the sub-matrices $\mathbb{E}_{\boldsymbol{\vartheta}_0}[\tilde{\mathbf{W}}_{\bar{h}}]$. Since the derivatives of the copula functions are assumed to be bounded, it only remains to consider the terms $\partial \Phi(\mathrm{y}_j^*)/\partial \eta_j$. Notice that the subscript i has been omitted for no good reason. In order to bound these quantities, we use the fact that the Normal distribution is a member of the exponential family; thus we write

$$f_{1,j}(\mathrm{y}_j^*) = \exp\left\{\frac{\mathrm{y}_j^* \eta_j - b(\eta_j)}{a(\phi)} + c(\mathrm{y}_j^*, \phi)\right\} = \exp\left\{\mathrm{y}_j^* \eta_j - \eta_j^2/2 + \left(-\frac{1}{2}\log(2\phi) - \frac{(\mathrm{y}_j^*)^2}{2}\right)\right\}.$$

In particular, notice that the scale parameter reduces to $a(\phi) = 1$ in our context, and the moments of the random variable Y_j^* correspond to $\mathbb{E}[Y_j^*] = b'(\eta_j) = \eta_j < \infty$ and $\mathbb{V}[Y_j^*] = b''(\eta_j)$. A sufficient condition for $\mathbb{E}[Y_j^*] < \infty$ is that Y_j^* has not zero counts on its last level K_j. In this case, in fact, the inequality $c_{j,K_j-1} < c_{j,K_j} = \infty$ holds strictly. Under usual smoothness assumption, we find that

$$\left|\frac{\partial \Phi(\mathrm{y}_j^*)}{\partial \eta_j}\right| = \left|\int_{-\infty}^{\mathrm{y}_j^*} \frac{\partial}{\partial \eta_j} f_{1,j}(t_j) \mathrm{d}t_j\right| \leq \left|\int_{\mathbb{R}} f_{1,j}(t_j)\left(t_j - b'(\eta_j)\right) \mathrm{d}t_j\right| \leq 2\mathbb{E}_{\boldsymbol{\vartheta}_0}|Y_j^*|$$

and, with similar arguments,

$$\left|\frac{\partial \Phi(\mathrm{y}_j^*)^2}{\partial^2 \eta_j}\right| \leq 2\mathbb{V}_{\boldsymbol{\vartheta}_0}|Y_j^*|.$$

Hence we have $\mathbb{E}_{\boldsymbol{\vartheta}_0}[w_{\bar{h},i}] = O(1)$ for all $i = 1, \ldots, n$ and $\bar{h} \in \{k_1, (k_1, k_2), \ldots, \gamma\}$ and $\mathbf{F}_n(\boldsymbol{\vartheta}_0) := \tilde{\mathbf{D}}^\top \mathbb{E}_{\boldsymbol{\vartheta}_0}[\tilde{\mathbf{W}}_j]\tilde{\mathbf{D}} = O(n/H_n)$.

To prove the second part of the statement, we stress that, by the properties of the B-spline basis, the (i, l_j)-th component of the matrices $\mathbf{X}_{j,l_j}^\top \mathbf{X}_{j,\bar{l}_j}$, $l_j, \bar{l}_j = 1, \ldots, L_j$ and all j, is 0 if $|i - l_j| > q$. Hence $\mathbf{X}_j^\top \mathbb{E}_{\boldsymbol{\vartheta}_0}[\tilde{\mathbf{W}}_j]\mathbf{X}_j$ has a band structure and the assertion follows by exploiting the order of the penalty terms (e.g. **?**). ∎

Variables	Driver		Other occupant		Independent model	
	estimates	(se)	estimates	(se)	estimates	(se)

Occupant Characteristics

Gender (male)						
female	0.1697	(0.0694)	0.0625	(0.0630)	0.1094	(0.0466)
Seat (driver)						
other/missing	−	−	ref.	−	0.2970	(0.2347)
front, passenger	−	−	0.2048	(0.1664)	0.3754	(0.0453)
rear, driver's side	−	−	0.0364	(0.2344)	0.1932	(0.1565)
rear, opposite driver	−	−	−0.0036	(0.2171)	0.3175	(0.1333)

Motorway Characteristics

Intersection (off intersection)						
X	0.0395	(0.2358)	−0.2162	(0.2428)	−0.0927	(0.1714)
T	−0.0924	(0.2535)	−0.3344	(0.2590)	−0.2378	(0.1832)
Y	−0.2814	(0.3907)	−0.3344	(0.3682)	−0.3054	(0.2694)
> 4 branches	−0.8819	(0.4609)	−0.4818	(0.4177)	−0.7115	(0.3076)
roundabout	0.3647	(0.2109)	0.3174	(0.2188)	0.3176	(0.1532)
circus/square	0.1333	(1.0256)	−0.4455	(1.0725)	−0.1332	(0.7553)
other	0.3505	(0.2982)	0.2448	(0.3088)	0.2856	(0.2173)
Type (motorway)						
Route Nationale	−0.1418	(0.1333)	−0.0205	(0.1336)	−0.1005	(0.0952)
Route Départementale	0.0070	(0.0981)	0.1478	(0.0993)	0.0718	(0.0703)
Voie Communale	−0.3588	(0.1056)	−0.0741	(0.1051)	−0.2200	(0.0751)
other	−0.4516	(0.3301)	−0.3340	(0.3357)	−0.4088	(0.2383)
Circulation regime (missing)						
one-way	0.0751	(0.1530)	−0.3744	(0.1574)	−0.1300	(0.1102)
two-way	0.3640	(0.1291)	0.1020	(0.1338)	0.2490	(0.0931)
presence of median	0.1197	(0.1430)	−0.1610	(0.1485)	0.0113	(0.1033)
other	1.0587	(0.5738)	0.1987	(0.5722)	0.6379	(0.4112)
Horizontal alignment (straight)						
left curve	−0.1137	(0.0800)	0.2048	(0.0812)	0.0246	(0.0572)
right curve	0.0004	(0.0892)	0.0364	(0.0890)	−0.0027	(0.0636)
S	−0.1379	(0.1665)	−0.0036	(0.1657)	−0.0397	(0.1175)
Location (other/missing)						
roadway	−0.4146	(0.1080)	−0.4878	(0.1103)	−0.4533	(0.0775)
emergency lane	−0.3953	(0.2359)	0.0222	(0.2395)	−0.2426	(0.1688)
shoulder	−0.0446	(0.1203)	−0.0932	(0.1221)	−0.0861	(0.0860)
sidewalk	−0.2552	(0.1864)	−0.6954	(0.1892)	−0.4705	(0.1336)
Obstacle (other/missing)						
fixed object	0.1938	(0.0887)	0.0984	(0.0904)	0.1594	(0.0636)
pedestrian	−7.2444	$(5.9\,10^5)$	−0.9181	(1.1237)	−1.0730	(0.8588)
vehicle	−0.1082	(0.1832)	−0.0178	(0.1818)	−0.0639	(0.1301)
animal	0.1777	(0.3076)	0.0467	(0.3031)	0.1418	(0.2180)

Accident Characteristics

Lighting (daylight)						
sunrise/sunset	0.0164	(0.1188)	0.1235	(0.1210)	0.0424	(0.0853)
night without street lights	0.1814	(0.0801)	−0.0430	(0.0788)	0.0709	(0.0567)
night, street lights in force	−0.0173	(0.0910)	0.1334	(0.0905)	0.0420	(0.0648)
Atmospheric conditions (normal)						
light rain	−0.2105	(0.0889)	−0.1271	(0.0880)	−0.1761	(0.0630)
heavy rain	0.4468	(0.1658)	−0.1548	(0.1678)	0.1702	(0.1189)
snow	0.6575	(0.4330)	−0.6317	(0.4231)	0.1938	(0.2940)
fog	−0.2960	(0.2864)	0.1544	(0.2890)	−0.0662	(0.2038)
heavy wind/storm	0.0588	(0.4496)	1.1834	(0.4899)	0.4839	(0.3227)
clear	1.0461	(0.4596)	0.6979	(0.4539)	0.8283	(0.3260)
clouds	0.1347	(0.1432)	−0.1661	(0.1458)	−0.0031	(0.1027)
Manner of collision (missing/other)						
head-on	−0.1864	(0.0761)	0.0176	(0.0773)	−0.0831	(0.0544)
rear-end	−0.3630	(0.1981)	−0.1348	(0.1971)	−0.2355	(0.1411)
sideswipe, right	−0.1723	(0.1739)	0.7225	(0.1774)	0.2286	(0.1237)
sideswipe, left	0.6712	(0.1677)	−0.0826	(0.1648)	0.3176	(0.1178)
Security device (not put on)						
put on	−0.4585	(0.0785)	−0.2646	(0.0771)	−0.3452	(0.0554)
$c_{j,1}$	−0.9966	(0.2097)	−1.6341	(0.2814)	−1.0693	(0.1515)
$c_{j,2}$	−0.0786	(0.0196)	−0.2296	(0.0198)	0.0378	(0.0138)
$c_{j,3}$	1.3514	(0.0271)	1.3742	(0.0243)	1.5235	(0.0180)
No. observations	1,232		1,232		2,464	

Table S2: Estimates for Scenario I: the independent model is obtained under a univariate model where all the observations are pooled together.

Variables	Joe$_0$: Semi-parametric		Independent		Joe$_0$: Parametric	
	Driver	Other occupant	Driver	Other occupant	Driver	Other occupant
Occupant Characteristics						
Gender (male)						
female	15.1771	4.4118	11.4417	3.9410	20.0326	6.3791
Seat (other/missing)						
front, passenger	—	3.3171	—	5.0900	—	11.2427
rear, driver's side	—	−0.6495	—	−5.2680	—	−2.2810
rear, opposite driver	—	2.5478	—	1.3643	—	10.9453
Motorway Characteristics						
Intersection (off intersection)						
X	3.5046	−15.5834	−1.9336	−16.9575	20.8335	−8.8219
T	−8.1106	−24.0573	−13.2887	−23.9279	0.2473	−24.1377
Y	−24.0709	−24.0592	−28.5440	−25.4054	−6.8963	−11.5996
> 4 branches	−64.7161	−34.3071	−70.0681	−34.7874	−56.1683	−38.8870
roundabout	32.9388	21.1949	29.4565	21.5968	40.6659	11.2428
circus/square	11.9005	−31.8270	13.1213	−31.8771	21.0943	16.2180
other	31.6143	16.6715	28.16066	17.1848	34.1223	23.2436
Type (motorway)						
Route Nationale	−12.3814	−1.4598	−14.2561	−2.9740	−7.4069	−0.0587
Route Départementale	0.6239	10.2816	0.5852	10.0325	3.5596	12.7703
Voie Communale	−30.2670	−5.3074	−30.3170	−5.1137	−20.9363	−2.6323
other	−37.3652	−24.0312	−40.1773	−24.6226	−31.9717	−30.2226
Circulation regime (missing)						
one-way	6.6856	−26.8814	8.3478	−29.6338	15.9702	−25.6134
two-way	32.8730	7.1588	33.9659	5.7432	36.7234	11.4425
presence of median	10.6766	−11.5934	13.2971	−12.5896	20.2130	−10.2389
other	285.5734	13.6792	73.6875	5.8991	69.6891	13.6003
Horizontal alignment (straight)						
left curve	−8.9555	18.0840	−10.1281	18.7363	−7.1916	15.3640
right curve	0.0402	2.5769	−1.4410	1.3507	−3.3128	−0.5520
S	−10.5973	−0.2560	−7.6513	1.1204	−9.4537	6.7268
Location (other/missing)						
roadway	−34.5838	−34.7099	−37.5420	−35.6480	−33.3844	−35.0592
emergency lane	−33.1021	1.5788	−40.3814	0.9101	−38.7264	−1.8851
shoulder	−3.9319	−6.6893	−7.3046	−7.9924	−9.8526	−9.4138
sidewalk	−21.9244	−48.0965	−25.7039	−49.8746	−23.9135	−51.1506
Obstacle (other/missing)						
fixed object	21.1914	7.7640	22.3652	8.4292	22.3471	8.8375
pedestrian	—	−21.5275	—	−22.3293	—	−22.2966
vehicle	−8.5727	−1.2459	−9.3135	−1.2383	−7.4586	9.0969
animal	19.0619	3.4888	24.3422	2.7093	6.4797	10.0389
Accident Characteristics						
Lighting (daylight)						
sunrise/sunset	1.4600	8.6284	−2.4667	8.7987	−0.4411	6.5159
night without street lights	16.2326	−3.0728	14.0798	−4.1294	20.2537	−5.9967
night, street lights in force	−1.5261	9.3080	3.9102	9.3804	−2.7762	10.9983
Atmospheric conditions (normal)						
light rain	−18.2084	−9.1396	−18.1288	−10.4779	−15.9728	−5.1552
heavy rain	40.7718	−11.1393	38.8988	−12.5850	55.3536	−12.7853
snow	65.8266	−44.1370	63.4047	−39.5186	58.4414	−59.4227
fog	−19.4357	12.9200	−16.3039	12.4447	−17.2725	10.5198
heavy wind/storm	5.2257	45.3357	6.38444	46.4522	−6.1463	53.2024
clear	267.3856	39.6371	72.0389	38.6520	121.7924	52.1828
clouds	12.0293	−11.9607	11.9639	−11.0216	20.5940	−11.6020
Manner of collision (missing/other)						
heads-on	−16.1804	1.2487	−15.0135	2.7577	−22.0029	2.4676
rear-end	−22.3114	−8.3346	−23.1520	−6.7087	−26.7919	−9.9022
sideswipe, right	−12.7931	116.2993	−12.4661	121.8290	−20.3684	129.3245
sideswipe, left	68.0130	−5.9216	56.3359	−6.4898	60.6336	−9.6748
Security device (not put on)						
put on	−25.6391	−14.2845	−24.7310	−12.9336	−27.2415	−14.1851
No. observations	1,232		1,232		1,232	

Table S3: Pseudo-elasticities for Scenario I obtained by applying the preferred Joe$_0$ copula, independent and the fully parametric models. Quantities are computed with respect to the hospitalised injuries.