A Trivariate Additive Regression Model with Arbitrary Link Functions and Varying Correlation Matrix *

Panagiota Filippou† Thomas Kneib‡ Giampiero Marra §

Rosalba Radice¶

2017-09-26

Abstract

In many empirical situations, modelling simultaneously three or more outcomes as well as their dependence structure can be of considerable relevance. Copulae provide a powerful framework to build multivariate distributions and allow one to view the specification of the marginal responses and their dependence as separate but related issues. We propose a generalisation of the trivariate additive probit model where the link functions can in principle be derived from any parametric distribution and the parameters describing the association between the responses can be made dependent on several types of covariate effects (such as linear, nonlinear, random, and spatial effects). All the coefficients of the model are estimated simultaneously within a penalized likelihood framework that uses a trust region algorithm with integrated automatic multiple smoothing parameter selection. The effectiveness of the model is assessed in simulation as well as empirically by modelling jointly three adverse birth binary outcomes in North Carolina. The approach can be easily employed via the `gjrm()` function in the `R` package `GJRM`.

*E-mail for correspondence: `panagiota.filippou.12@ucl.ac.uk`.

†Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK.

‡Chairs of Statistics and Econometrics, Georg-August-Universität Göttingen, Humboldtallee 3, 37073 Göttingen, Germany.

§Department of Statistical Science, University College London, Gower Street, London WC1E 6BT, UK.

¶Department of Economics, Mathematics and Statistics, Birkbeck, University of London, Malet Street, London WC1E 7HX, UK.

Keywords Additive predictor; Binary response; Cholesky decomposition; Penalized regression spline; Simultaneous parameter estimation; Trivariate distribution.

1 Introduction

When the researcher is interested in modelling more than one response, univariate regression will not yield valid inferences if there is residual dependence between the outcomes conditional on covariates. The case of trivariate models has been discussed in literature in various contexts. For example, Loureiro et al. (2010) assessed the effect of parental smoking habits on their children's smoking habits by estimating a three-equation probit regression model, whereas Zhong et al. (2012) evaluated the safety of a treatment and identified an optimal dose by jointly modelling the probabilities of toxicity, efficacy, and surrogate efficacy given a specific dose. Król et al. (2016) examined the response to a treatment on patients with metastatic colorectal cancer by analysing simultaneously three outcomes: a longitudinal marker, a set of recurrent events, and a terminal event. A mixture of powers copula-based approach to model jointly three binary and discrete outcomes was employed by Zimmer & Trivedi (2006), whereas Zhang et al. (2015) developed a Bayesian algorithm to estimate trivariate probit-ordered models affected by double sample selection.

This paper contributes to the literature by introducing a generalization of the trivariate additive probit model. Specifically, we extend and therefore enhance the model proposed by Filippou et al. (2017) by allowing (i) the link functions to be virtually derived from any parametric distribution and (ii) the model's association parameters to depend on several types of covariate effects (such as linear, nonlinear, random, and spatial effects). The first extension allows for the use of link functions other than probit. In particular, the additional link functions implemented for this work are the logit and complementary log-log which are used extensively in numerous disciplines, including the medical and social sciences. In clinical research logit models are widely used as they provide direct information about which treatment has the best odds of benefiting a patient, for instance. Complementary log-log models have important applications in survival analysis where they can, for example, provide a clear insight into the relative reduction of risk for death or progression. Extension (ii) is of some relevance since it can help to gain insights into the way the residual association between the responses is modified by the presence of covariates. To the best of our knowledge, the two proposed developments have not yet been considered in the context of trivariate (or more generally, multivariate) binary response regression models.

It is worth noting that our proposal can also be regarded as an extension of the bivariate regression approaches introduced by Marra & Radice (2017a), Klein & Kneib (2016) and Radice et al. (2016) as well as of the popular generalized additive models (GAMs) and GAMs for location, scale and shape of Wood (2017) and Rigby & Stasinopoulos (2005). Despite we have focused on trivariate binary models, the theoretical results in the paper can be straightforwardly extended to the case of more dimensions. Function `gjrm()` in the `R` package `GJRM` (Marra & Radice, 2017b) implements various types of joint models and includes the developments in this article.

The next section introduces the proposed model, Section 3 describes the log-likelihood and Section 4 provides the key details on estimation. The proposal is empirically evaluated in a simulation study, presented in Section 5, and then applied to a case study in Section 6, where the interest is in modelling jointly three adverse birth binary outcomes in North Carolina. Section 7 concludes the paper.

2 Model specification

This section introduces an extension of the trivariate probit that is based on copulae, additive predictors and a modified Cholesky decomposition of the model's correlation matrix.

In general, a multivariate distribution can be constructed using a copula function that joins together marginal distributions which may come from different families (Joe, 1997). Suppose that $\mathcal{C}$ denotes a joint cumulative distribution function (cdf) with support in $[0,1]^3$ and whose one-dimensional margins are uniform. Let also $\mathcal{U}_m^{-1} : (0,1) \rightarrow \mathbb{R}$ be a quantile function, $\forall m = 1,2,3$, $F_m(\eta_{mi}) : \mathbb{R} \rightarrow [0,1]$ a univariate cdf, $F\left(\mathcal{U}_1^{-1}\left\{F_1(\eta_{1i})\right\}, \mathcal{U}_2^{-1}\left\{F_2(\eta_{2i})\right\}, \mathcal{U}_3^{-1}\left\{F_3(\eta_{3i})\right\}\right)$ a joint cdf, and η_{mi} an additive predictor (made up of regression coefficients and covariates as described in Section 2.2) for $i = 1, \ldots, n$, where n denotes the sample size. Then there exists a three-dimensional copula function $\mathcal{C} : [0, 1]^3 \to [0, 1]$ defined as

$$\mathcal{C}(F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i})) = F\left(\mathcal{U}_1^{-1}\left\{F_1(\eta_{1i})\right\}, \mathcal{U}_2^{-1}\left\{F_2(\eta_{2i})\right\}, \mathcal{U}_3^{-1}\left\{F_3(\eta_{3i})\right\}\right), \tag{1}$$

which satisfies: (C.1) $\mathcal{C}\left(F_1(\eta_{1i}), 1, 1\right) = F_1(\eta_{1i})$, $\mathcal{C}\left(1, F_2(\eta_{2i}), 1\right) = F_2(\eta_{2i})$, $\mathcal{C}\left(1, 1, F_3(\eta_{3i})\right) = F_3(\eta_{3i})$, $\forall F_m(\eta_{mi}) \in [0, 1]$ and $m \leq 3$; (C.2) $\mathcal{C}\left(F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i})\right) = 0$ if $F_m(\eta_{mi}) = 0$ for any $m \leq 3$; and (C.3) $\mathcal{C}$ is 3-increasing (Sklar, 1959). Condition (C.1) states that if the realizations of two variables are known each with marginal probability of one, then the joint probability of the three outcomes is the same as the probability of the remaining uncertain outcome. Condition (C.2) is sometimes referred to as the grounded property of a copula and states that the joint probability of all outcomes is zero if the marginal probability of any outcome is zero. Condition (C.3) means that the copula volume of any 3-dimensional interval is non-negative. A copula $\mathcal{C}$ is unique on the cartesian product of the ranges of the marginal cdfs $\mathbf{Ran}(F_1(\eta_{1i}))\times\mathbf{Ran}(F_2(\eta_{2i}))\times\mathbf{Ran}(F_3(\eta_{3i}))$. The copula is unique if the margins are continuous. Any copula lies always in the interval

$$\max\left\{\sum_{m=1}^{3} F_m(\eta_{mi}) - 2, 0\right\} \leq \mathcal{C}\left(F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i})\right) \leq \min\left\{F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i})\right\},$$

the so-called *Fréchet–Hoeffding* bounds. A desirable feature of a copula is that it should cover the sample space between the lower and upper bounds, and that as the association parameters approach the lower (upper) bound of their permissible ranges, the copula approaches the Fréchet–Hoeffding lower (upper) bound. Knowledge of the Fréchet–Hoeffding bounds is therefore important in selecting an appropriate copula. For more details see, for instance, Trivedi & Zimmer (2007) and references therein.

In this paper, we employ the trivariate Gaussian copula with dependence structure characterized by coefficients $\vartheta_{12,i}$, $\vartheta_{13,i}$ and $\vartheta_{23,i}$ which form the model's correlation matrix $\mathbf{\Sigma}_i$. Based on (1), we express the trivariate Gaussian copula as $\Phi_3\left(\Phi^{-1}\left\{F_1(\eta_{1i})\right\}, \Phi^{-1}\left\{F_2(\eta_{2i})\right\}, \Phi^{-1}\left\{F_3(\eta_{3i})\right\}; \mathbf{0}, \mathbf{\Sigma}_i\right)$, where Φ^{-1} is the quantile function of a standard normal, $F_m(\eta_{mi})$ is derived in this case from the standardised normal, logistic or Gumbel univariate cdf which are defined as

$$F_m(\eta_{mi}) = \Phi(\eta_{mi}), \quad F_m(\eta_{mi}) = \frac{\exp(\eta_{mi})}{1+\exp(\eta_{mi})} \quad \text{and} \quad F_m(\eta_{mi}) = 1 - \exp\left\{-\exp(\eta_{mi})\right\},$$

and matrix $\boldsymbol{\Sigma}_i$ is equal to

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} 1 & \vartheta_{12,i} & \vartheta_{13,i} \\ \vartheta_{12,i} & 1 & \vartheta_{23,i} \\ \vartheta_{13,i} & \vartheta_{23,i} & 1 \end{pmatrix}, \tag{2}$$

where $\vartheta_{k_1 k_2,i}$ is the correlation coefficient between the k_1^{th} and k_2^{th} responses for subject i, for $k_1 = 1, 2,\ k_2 = 2, 3$, with $k_1 \neq k_2$. The case of non-normal dependence is tricky. In this work, we have considered several ways of modelling non-Gaussian structures by reviewing the growing literature on multivariate models. Supplementary Material A discusses five different ways for potentially achieving this aim in our case: Archimedean copulae, mixtures of powers, pair-copulae constructions, the trivariate Student-t distribution, and the composite likelihood approach. Although these approaches allow for non-Gaussian dependencies, the majority of them make certain strong assumptions which may be regarded as acceptable only in specific applied contexts. In fact, such methods would limit the generality as well as applicability of the modelling approach presented here. The only suitable alternative would appear to be the trivariate Student-t distribution, however, as shown in the Supplementary Material A, there is not much to be gained by using such distribution in our context. In conclusion, the Gaussian copula seems to be a sensible and tractable modelling choice for the case of trivariate binary data.

Each coefficient in matrix (2) is allowed to be expressed as a function of an additive predictor. The challenge to address here is that the range of each correlation's additive predictor has to be unbounded to avoid constrained optimization and that the correlation matrix $\boldsymbol{\Sigma}_i$ must be positive definite with each of its coefficients taking values in $[-1, 1]$. This makes the parameter space of

$\mathbf{\Sigma}_i$ somewhat complex with restrictions for each parameter depending on the values of the others. To this end, we propose using a modified Cholesky decomposition approach which is described in the next section.

2.1 Unconstrained parametrization for the correlation matrix

The standard Cholesky decomposition of a positive-definite correlation matrix $\mathbf{\Sigma}$ is of the form $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^\top$, where $\mathbf{C}$ is a unique lower-triangular matrix with positive diagonal entries. Modifications of the standard Cholesky decomposition can be found in the literature. For example, Pourahmadi (1999, 2000) shows that the modified Cholesky decomposition of $\mathbf{\Sigma}^{-1}$ offers a simple unconstrained reparametrization of the covariance matrix, while Chen & Dunson (2003) propose an alternative modified Cholesky decomposition to factorize the covariance matrix. As shown by Pourahmadi (2007), who provides an overview of the two methods, estimation of the new parameters in the latter decomposition may be more demanding computationally. In this paper, we employ a modification of the work by Pourahmadi (1999, 2000), where we employ the modified Cholesky approach with unit variance constraints to deal with correlation matrices.

Let $\mathbf{\Sigma}_i^*$ denote a symmetric positive-definite correlation matrix, $\forall i$, defined as

$$\mathbf{\Sigma}_i^* = \mathbf{C}_i^*\mathbf{C}_i^{*\top} = \begin{pmatrix} 1 & \eta_{12,i} & \eta_{13,i} \\ \eta_{12,i} & 1+\eta_{12,i}^2 & \eta_{12,i}\eta_{13,i}+\eta_{23,i} \\ \eta_{13,i} & \eta_{12,i}\eta_{13,i}+\eta_{23,i} & 1+\eta_{13,i}^2\eta_{23,i}^2 \end{pmatrix},$$

where $\eta_{k_1k_2,i} \in \mathbb{R}$, $\forall k_1, k_2$, and $\mathbf{C}_i^*$ is equal to

$$\mathbf{C}_i^* = \begin{pmatrix} 1 & 0 & 0 \\ \eta_{12,i} & 1 & 0 \\ \eta_{13,i} & \eta_{23,i} & 1 \end{pmatrix}.$$

By using the variance-correlation decomposition $\mathbf{\Sigma}_i = \mathbf{T}_i\mathbf{\Sigma}_i^*\mathbf{T}_i$ with $\mathbf{T}_i = \mathrm{diag}\left(1, \left(1+\eta_{12,i}^2\right)^{-1/2},\right.$

$\left(1+\eta_{13,i}^2+\eta_{23,i}^2\right)^{-1/2}\Big)$, we have that the correlation matrix $\mathbf{\Sigma}_i$ can be expressed as

$$\mathbf{\Sigma}_i = \begin{pmatrix} 1 & \frac{\eta_{12,i}}{\sqrt{1+\eta_{12,i}^2}} & \frac{\eta_{13,i}}{\sqrt{1+\eta_{13,i}^2+\eta_{23,i}^2}} \\ \frac{\eta_{12,i}}{\sqrt{1+\eta_{12,i}^2}} & 1 & \frac{\eta_{12,i}\eta_{13,i}+\eta_{23,i}}{\sqrt{(1+\eta_{12,i}^2)(1+\eta_{12,i}^2+\eta_{23,i}^2)}} \\ \frac{\eta_{13,i}}{\sqrt{1+\eta_{13,i}^2+\eta_{23,i}^2}} & \frac{\eta_{12,i}\eta_{13,i}+\eta_{23,i}}{\sqrt{(1+\eta_{12,i}^2)(1+\eta_{12,i}^2+\eta_{23,i}^2)}} & 1 \end{pmatrix}.$$

The correlation parameters can therefore be defined as $\vartheta_{12,i}=\eta_{12,i}/\sqrt{1+\eta_{12,i}^2}$, $\vartheta_{13,i}=\eta_{13,i}/\sqrt{1+\eta_{13,i}^2+\eta_{23,i}^2}$ and $\vartheta_{23,i}=(\eta_{12,i}\eta_{13,i}+\eta_{23,i})/\sqrt{(1+\eta_{12,i}^2)(1+\eta_{12,i}^2+\eta_{23,i}^2)}$. It follows that

$$\eta_{12,i}=\mathscr{F}_{12}(\vartheta_{12,i})=\sqrt{\frac{\vartheta_{12,i}^2}{1-\vartheta_{12,i}^2}},\quad \eta_{13,i}=\mathscr{F}_{13}(\vartheta_{13,i})=\sqrt{\frac{\vartheta_{13,i}^2\left(1+\frac{\mathsf{A}}{1-\mathsf{A}}\right)}{1-\vartheta_{13,i}^2}},\quad \eta_{23,i}=\mathscr{F}_{23}(\vartheta_{23,i})=\sqrt{\frac{\mathsf{A}}{1-\mathsf{A}}},$$

where $\mathsf{A}=\left(\frac{\vartheta_{23,i}\sqrt{1+\eta_{12,i}^2}-\eta_{12,i}\vartheta_{13,i}}{\sqrt{1-\vartheta_{13,i}^2}}\right)^2$. Therefore, by construction we have that $\vartheta_{k_1k_2,i}\in[-1,1]$, $\eta_{k_1k_2,i}\in\mathbb{R}$, $\forall k_1,k_2,i$ and the resulting correlation matrix is positive definite, as required.

2.2 Additive predictor

All the model's parameters are related to covariates and regression coefficients via additive predictors. Let us define a generic predictor η_i as a function of parametric components and smooth functions. That is,

$$\eta_i=\mathbf{v}_i^{\top}\boldsymbol{\gamma}+\sum_{\nu=1}^{\tilde{N}}s_\nu(z_{\nu i}),\ \ i=1,\ldots,n, \tag{3}$$

where $\mathbf{v}_i$ contains binary and/or categorical predictors, vector $\boldsymbol{\gamma}$ represents the effects of the variables in $\mathbf{v}_i$, and $s_\nu(z_{\nu i})$ is a smooth function of covariate $z_{\nu i}$, $\forall\nu=1,\ldots,\tilde{N}$ with $\tilde{N}$ being the number of smooth terms in (3). The smooth functions are represented using the regression spline approach popularized by Eilers & Marx (1996) because of its computational efficiency, theoretical properties and flexibility in representing several types of covariate effects (e.g., Wood, 2017). Using this approach, $s_\nu(z_{\nu i})$ is approximated by a linear combination of known basis functions

$b_{\nu j}(z_{\nu i})$ and regression parameters $\alpha_{\nu j}$. That is,

$$s_\nu(z_{\nu i}) \approx \sum_{j=1}^{J_\nu} \alpha_{\nu,j} b_{\nu,j}(z_{\nu i}) = \mathbf{L}_\nu(z_{\nu i})\boldsymbol{\alpha}_\nu, \tag{4}$$

where $\mathbf{L}_\nu(z_{\nu i})$ is a vector containing the J_ν basis functions evaluated at $z_{\nu i}$, that is $\mathbf{L}_\nu(z_{\nu i}) = \{b_{\nu,1}(z_{\nu i}), b_{\nu,2}(z_{\nu i}), \ldots, b_{\nu,J_\nu}(z_{\nu i})\}$, and $\boldsymbol{\alpha}_\nu$ is the corresponding parameter vector defined as $\boldsymbol{\alpha}_\nu = (\alpha_{\nu,1}, \alpha_{\nu,2}, \ldots, \alpha_{\nu,J_\nu})^\top$, $\forall \nu$. Each term has an associated quadratic penalty $\lambda_\nu \boldsymbol{\alpha}_\nu^\top \mathbf{S}_\nu \boldsymbol{\alpha}_\nu$ which enforces specific properties on the ν^{th} function (such as smoothness) and that is therefore used during model fitting. Smoothing parameter $\lambda_\nu \in [0, \infty)$ controls the trade-off between fit and smoothness. The overall penalty can be written as $\boldsymbol{\alpha}^\top \mathbf{S} \boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = \left(\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_{\tilde{N}}^\top\right)^\top$, $\mathbf{S} = \mathrm{diag}\left(\mathbf{0}_{\tilde{P}}^\top, \lambda_1 \mathbf{S}_1, \ldots, \lambda_{\tilde{N}} \mathbf{S}_{\tilde{N}}\right)$, $\tilde{P}$ denotes the number of parametric components in the additive predictor and the $\mathbf{S}_\nu$ are positive definite or semi-definite symmetric known square matrices. Centering constraint $\sum_i s_\nu(z_{\nu i}) = 0$ is imposed on all smooth terms in the model for identification purposes. The above formulation allows us to represent many types of covariate effects depending on the nature of the covariate(s) considered. These include random, spatial and non-linear effects. We refer the reader to Filippou et al. (2017), and references therein, for an overview of some common examples.

3 Log-likelihood

To avoid over-fitting, simultaneous estimation of all parameters of the trivariate additive binary model is achieved by solving

$$\hat{\boldsymbol{\delta}} \;:=\; \arg\min_{\boldsymbol{\delta}} -\ell_p(\boldsymbol{\delta}) = \arg\min_{\boldsymbol{\delta}} -\{\log \mathcal{L}(\mathbf{Y}; \boldsymbol{\delta}) - \frac{1}{2}\boldsymbol{\delta}^\top \mathbf{S}_{\boldsymbol{\lambda}} \boldsymbol{\delta}\}, \tag{5}$$

where $\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n)^\top$ with $\boldsymbol{y}_i = (y_{1i}, y_{2i}, y_{3i})^\top$ which denotes the three correlated binary responses, $\boldsymbol{\delta} = (\boldsymbol{\beta}^\top, \boldsymbol{\beta}_{\boldsymbol{\vartheta}}^\top)^\top$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)^\top$, $\boldsymbol{\beta}_{\boldsymbol{\vartheta}} = (\boldsymbol{\beta}_{12}, \boldsymbol{\beta}_{13}, \boldsymbol{\beta}_{23})^\top$, $\boldsymbol{\beta}_m$ includes the regression coefficients in m^{th} equation, $\boldsymbol{\beta}_{k_1 k_2}$ denotes the coefficients in additive predictor $\eta_{k_1 k_2, i}$, $\mathbf{S}_{\boldsymbol{\lambda}} =$

$\mathrm{diag}\Big(\mathbf{0}^{\top}_{\tilde{P}_1}, \lambda_{1\nu_1}\mathbf{S}_{1\nu_1}, \ldots, \lambda_{1\tilde{N}_1}\mathbf{S}_{1\tilde{N}_1}, \mathbf{0}^{\top}_{\tilde{P}_2}, \lambda_{2\nu_2}\mathbf{S}_{2\nu_2}, \ldots, \lambda_{2\tilde{N}_2}\mathbf{S}_{2\tilde{N}_2}, \mathbf{0}^{\top}_{\tilde{P}_3}, \lambda_{3\nu_3}\mathbf{S}_{3\nu_3}, \ldots, \lambda_{3\tilde{N}_3}\mathbf{S}_{3\tilde{N}_3}, \mathbf{0}^{\top}_{\tilde{P}_{12}},$ $\lambda_{12\nu_{12}}\mathbf{S}_{12\nu_{12}}, \ldots, \lambda_{12\tilde{N}_{12}}\mathbf{S}_{12\tilde{N}_{12}}, \mathbf{0}^{\top}_{\tilde{P}_{13}}, \lambda_{13\nu_{13}}\mathbf{S}_{13\nu_{13}}, \ldots, \lambda_{13\tilde{N}_{13}}\mathbf{S}_{13\tilde{N}_{13}}, \mathbf{0}^{\top}_{\tilde{P}_{23}}, \lambda_{23\nu_{23}}\mathbf{S}_{23\nu_{23}}, \ldots, \lambda_{23\tilde{N}_{23}}\mathbf{S}_{23\tilde{N}_{23}}\Big),$ $\mathbf{S}_{m\nu_m}$ and $\mathbf{S}_{k_1k_2\nu_{k_1k_2}}$ are defined following a similar construction as $\mathbf{S}_{\nu}$, $\lambda_{m\nu_m}$ and $\lambda_{k_1k_2\nu_{k_1k_2}}$ are defined similarly as λ_{ν}, $\boldsymbol{\lambda}$ is a vector containing all smoothing parameters, $\tilde{P}_{k_1k_2}$ denotes the number of parametric components in $\eta_{k_1k_2,i}$ and $\tilde{P}_m$ that in the m^{th} equation. For a 3-D binary response vector we have 2^3 trivariate probabilities expressed via the trivariate Gaussian copula function. The likelihood is given by the joint density of observed outcomes

$$\mathcal{L}(\mathbf{Y};\boldsymbol{\delta}) = \prod_{i=1}^{n}\prod_{\tilde{k}=1}^{2^3}\mathcal{L}_{i\tilde{k}}(\boldsymbol{y}_i;\boldsymbol{\delta}) = \prod_{i=1}^{n}\prod_{\tilde{k}=1}^{2^3}\boldsymbol{\Psi}_{i\tilde{k}}^{\mathcal{Y}_{i\tilde{k}}},$$

where $\mathcal{L}_{i\tilde{k}}$ is derived from Lemma 1 for $M=3$. Term $\mathcal{Y}_{i\tilde{k}}$ denotes an indicator variable for the $\tilde{k}^{th}$ combination of the three possible events $y_{1i}=\bar{e}_1, y_{2i}=\bar{e}_2, y_{3i}=\bar{e}_3$ with $\bar{e}_m\in\{0,1\}\ \forall m$ and $\boldsymbol{\Psi}_{i\tilde{k}}$ is the corresponding trivariate Gaussian copula function. Note that for each $\tilde{k}$ the form of $\boldsymbol{\Psi}_{i\tilde{k}}$ and $\mathcal{Y}_{i\tilde{k}}$ is different as their structure depends on the $\tilde{k}^{th}$ combination of the three possible events. The calculation of the multivariate normal probabilities is described in detail in Filippou et al. (2017).

Lemma 1. *Quantity $\mathcal{L}_{i\tilde{k}}$, evaluated at the vector $(\boldsymbol{\mathcal{B}}_i\mathbf{H}_i)_{\tilde{k}}$ is equal to the cdf of a multivariate standardized normal vector with correlation matrix $(\boldsymbol{\mathcal{B}}_i\boldsymbol{\Sigma}_i\boldsymbol{\mathcal{B}}_i)_{\tilde{k}}$, that is*

$$\mathcal{L}_{i\tilde{k}}(\boldsymbol{y}_i;\boldsymbol{\delta}) = \boldsymbol{\Psi}_{i\tilde{k}}^{\mathcal{Y}_{i\tilde{k}}} = \left\{\Phi_M((\boldsymbol{\mathcal{B}}_i\mathbf{H}_i)_{\tilde{k}};\mathbf{0},(\boldsymbol{\mathcal{B}}_i\boldsymbol{\Sigma}_i\boldsymbol{\mathcal{B}}_i)_{\tilde{k}})\right\}^{\mathcal{Y}_{i\tilde{k}}} = \left\{\Phi_M((\boldsymbol{\mathcal{W}}_i)_{\tilde{k}};\mathbf{0},(\boldsymbol{\Upsilon}_i)_{\tilde{k}})\right\}^{\mathcal{Y}_{i\tilde{k}}},$$

where $\boldsymbol{\mathcal{W}}_i = \boldsymbol{\mathcal{B}}_i\mathbf{H}_i = (\mathcal{W}_{1,i},\ldots,\mathcal{W}_{M,i})^{\top}$, $\mathbf{H}_i = (\Phi^{-1}(F_1(\eta_{1i})),\ldots,\Phi^{-1}(F_M(\eta_{Mi})))^{\top}$, $\boldsymbol{\Upsilon}_i = \boldsymbol{\mathcal{B}}_i\boldsymbol{\Sigma}_i\boldsymbol{\mathcal{B}}_i$, $\mathcal{W}_{m,i}=\tilde{y}_{mi}\Phi^{-1}(F_m(\eta_{mi}))$, for $\tilde{y}_{mi}=(2y_{mi}-1)$, y_{mi} denotes the m^{th} binary response, $F_m(\eta_{mi})$ denotes the univariate cdf, η_{mi} is an additive predictor and $\boldsymbol{\mathcal{B}}_i$ denotes a diagonal $M\times M$ matrix with main diagonal elements $\tilde{y}_{mi}=(2y_{mi}-1)$, that is $\boldsymbol{\mathcal{B}}_i=\mathrm{diag}(2y_{1i}-1,2y_{2i}-1,\ldots,2y_{Mi}-1)$.

Proof. See Supplementary Material B. □

4 Estimation details

To minimize (5), we have extended the efficient and stable trust region algorithm with integrated automatic multiple smoothing parameter selection described by Filippou et al. (2017) to allow for the specification of virtually any parametric link function, and for the correlation matrix to depend on covariate effects as described earlier. The practical success of these extensions depends on the availability of the analytical score and Hessian matrix of the model which are fundamental for a reliable, stable and efficient implementation of the above mentioned algorithm. This requires to amend and generalise the results presented in the work by Filippou et al. (2017). Specifically, we compute the analytical score function $\boldsymbol{g}_i(\boldsymbol{\delta}^{[\varkappa]}) = \nabla_{\boldsymbol{\delta}}\ell_i\left(\boldsymbol{\delta}^{[\varkappa]}\right)$, and Hessian matrix $\boldsymbol{\mathcal{H}}_i(\boldsymbol{\delta}^{[\varkappa]}) = \nabla_{\boldsymbol{\delta}}\nabla_{\boldsymbol{\delta}}^{\top}\ell_i\left(\boldsymbol{\delta}^{[\varkappa]}\right)$ as

$$\nabla_{\boldsymbol{\delta}}\ell_i(\boldsymbol{\delta}) = \left(\frac{\partial\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}}\right)^{\top}\frac{\partial\ell_i(\boldsymbol{\delta})}{\partial\boldsymbol{\eta}_i} = \left(\frac{\partial\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}}\right)^{\top}\left\{\frac{\partial\ell_i(\boldsymbol{\delta})}{\partial\boldsymbol{F}_i}\frac{\boldsymbol{F}_i}{\partial\boldsymbol{\eta}_i}\right\} = \left(\frac{\partial\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}}\right)^{\top}\left\{\frac{1}{\boldsymbol{\Psi}_{i\tilde{k}}}\frac{\partial\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{F}_i}\frac{\partial\boldsymbol{F}_i}{\partial\boldsymbol{\eta}_i}\right\}, \quad (6)$$

$$\begin{aligned}\nabla\nabla_{\boldsymbol{\delta}\boldsymbol{\delta}^{\top}}\ell_i(\boldsymbol{\delta}) &= \left\{\frac{1}{\boldsymbol{\Psi}_{i\tilde{k}}}\frac{\partial\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{F}_i}\frac{\partial\boldsymbol{F}_i}{\partial\boldsymbol{\eta}_i}\right\}\frac{\partial^2\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}^{\top}} + \left(\frac{\partial\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}}\right)^{\top}\left\{-\frac{1}{\boldsymbol{\Psi}_{i\tilde{k}}\boldsymbol{\Psi}_{i\tilde{k}}^{\top}}\frac{\partial\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{F}_i}\frac{\partial\boldsymbol{F}_i}{\partial\boldsymbol{\eta}_i}\left(\frac{\partial\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{\eta}_i}\right)^{\top} + \right.\\ &\left.\frac{1}{\boldsymbol{\Psi}_{i\tilde{k}}}\left[\frac{\partial^2\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{F}_i\boldsymbol{F}_i^{\top}}\left(\frac{\partial\boldsymbol{F}_i}{\partial\boldsymbol{\eta}_i}\right)^2 + \frac{\partial\boldsymbol{\Psi}_{i\tilde{k}}}{\partial\boldsymbol{F}_i}\frac{\partial^2\boldsymbol{F}_i}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^{\top}}\right]\right\}\left(\frac{\partial\boldsymbol{\eta}_i}{\partial\boldsymbol{\delta}}\right), \end{aligned} \quad (7)$$

where, $\boldsymbol{\eta}_i = (\eta_{1i}, \eta_{2i}, \eta_{3i}, \eta_{12,i}, \eta_{13,i}, \eta_{23,i})^{\top}$, $\boldsymbol{F}_i = (F_1(\eta_{1i}), F_2(\eta_{2i}), F_3(\eta_{3i}), F_4(\eta_{4i}), F_5(\eta_{5i}), F_6(\eta_{6i}))^{\top}$ with $(F_4(\eta_{4i}), F_5(\eta_{5i}), F_6(\eta_{6i})) = (\vartheta_{12,i}, \vartheta_{13,i}, \vartheta_{23,i})$, $\partial\boldsymbol{\eta}_i/\partial\boldsymbol{\delta} = \mathrm{diag}\,(\partial\eta_{1i}/\partial\boldsymbol{\beta}_1, \partial\eta_{2i}/\partial\boldsymbol{\beta}_2, \partial\eta_{3i}/\partial\boldsymbol{\beta}_3,$ $\partial\eta_{12,i}/\partial\boldsymbol{\beta}_{12}, \partial\eta_{13,i}/\partial\boldsymbol{\beta}_{13}, \partial\eta_{23,i}/\partial\boldsymbol{\beta}_{23})$ and $\partial\ell(\boldsymbol{\delta})/\partial\boldsymbol{\eta}_i$=$(\partial\ell(\boldsymbol{\delta})/\partial\eta_{1i}, \partial\ell(\boldsymbol{\delta})/\partial\eta_{2i}, \partial\ell(\boldsymbol{\delta})/\partial\eta_{3i},$ $\partial\ell(\boldsymbol{\delta})/\partial\eta_{12,i}, \partial\ell(\boldsymbol{\delta})/\partial\eta_{13,i}, \partial\ell(\boldsymbol{\delta})/\partial\eta_{23,i})^{\top}$. Predictor $\boldsymbol{\eta}_i$ is functionally dependent on $\boldsymbol{\delta}$, that is $\boldsymbol{\eta}_i = \boldsymbol{\eta}_i(\boldsymbol{\delta})$. Implementation of (6) and (7) has been a tedious and non-trivial task, especially because of the presence of a varying correlation matrix. This extension required, for instance, the use of the multivariate chain rule which was employed as follows. As shown in Section 2.1, $\vartheta_{k_1k_2,i}$ may depend on $\eta_{k_1k_2,i}$ and $\boldsymbol{\eta}_{-k_1k_2,i}$, where $\boldsymbol{\eta}_{-k_1k_2,i} \in \bar{\boldsymbol{\eta}}_i \setminus \eta_{k_1k_2,i}$, for $\bar{\boldsymbol{\eta}}_i = (\eta_{12,i}, \eta_{13,i}, \eta_{23,i})^{\top}$. Hence, term $\partial\bar{\boldsymbol{F}}_i/\partial\bar{\boldsymbol{\eta}}_i$, for $\bar{\boldsymbol{F}}_i = (\vartheta_{12,i}, \vartheta_{13,i}, \vartheta_{23,i})^{\top}$, is a 3×3 Jacobian matrix containing all the derivatives of $\bar{\boldsymbol{F}}_i$ with respect to $\bar{\boldsymbol{\eta}}_i$. That is,

$$\frac{\partial \bar{\boldsymbol{F}}_i}{\partial \bar{\boldsymbol{\eta}}_i} = \begin{pmatrix} \frac{\partial \vartheta_{12,i}}{\eta_{12,i}} & \frac{\partial \vartheta_{12,i}}{\eta_{13,i}} & \frac{\partial \vartheta_{12,i}}{\eta_{23,i}} \\ \frac{\partial \vartheta_{13,i}}{\eta_{12,i}} & \frac{\partial \vartheta_{13,i}}{\eta_{13,i}} & \frac{\partial \vartheta_{13,i}}{\eta_{23,i}} \\ \frac{\partial \vartheta_{23,i}}{\eta_{12,i}} & \frac{\partial \vartheta_{23,i}}{\eta_{13,i}} & \frac{\partial \vartheta_{23,i}}{\eta_{23,i}} \end{pmatrix}.$$

The above accounts for the dependencies between $\vartheta_{k_1k_2,i}$ and $\eta_{k_1k_2,i}$ as well as $\boldsymbol{\eta}_{-k_1k_2,i}$. Second-order derivatives were derived in a similar way. More generically, implementation of (6) and (7) was achieved via Propositions 2 and 3 by setting $M = 3$.

Proposition 2. *Assume that $\boldsymbol{\mathcal{W}}_i$ is a multivariate standardized normal vector with correlation matrix equal to $\boldsymbol{\Upsilon}_i$. Then the first-order derivative of the M-variate normal cdf $\Phi_M(\boldsymbol{\mathcal{W}}_i; \mathbf{0}, \boldsymbol{\Upsilon}_i)$ with respect to $\boldsymbol{\beta}_m$, $\forall m = 1, \ldots, M$, can be expressed as*

$$\begin{aligned} \frac{\partial \Phi_M(\boldsymbol{\mathcal{W}}_i; \mathbf{0}, \boldsymbol{\Upsilon}_i)}{\partial \boldsymbol{\beta}_m} &= \phi(\mathcal{W}_{m,i}; 0, 1)\Phi_{M-1}(\boldsymbol{\mathcal{W}}_{-m,i} | \mathcal{W}_{m,i}; \boldsymbol{M}_i^m, \boldsymbol{\Theta}_i^m) \frac{f_m(\eta_{mi})}{\phi\left(\Phi^{-1}(F_m(\eta_{mi}))\right)} \\ &\quad (2y_{mi} - 1)\boldsymbol{x}_{mi}^\top \end{aligned}$$

where M denotes the total number of equations under a multivariate binary framework, $\mathcal{W}_{m,i}$ denotes the linear predictor of the m^{th} equation and is equal to $(2y_{mi} - 1)\Phi^{-1}(F_m(\eta_{mi}))$, $\boldsymbol{\beta}_m$ denotes the parameter vector of covariate vector $\mathbf{x}_{mi}$, the vector of linear predictors $\boldsymbol{\mathcal{W}}_{-m,i}$ is defined as $(\mathcal{W}_{1,i}, \ldots, \mathcal{W}_{m-1,i}, \mathcal{W}_{m+1,i}, \ldots, \mathcal{W}_{M,i})^\top$ and $f_m(\eta_{mi})$ and $F_m(\eta_{mi})$ denote the univariate pdf and cdf respectively which can be specified via the normal, logistic and Gumbel distributions. The mean $\boldsymbol{M}_i^m$ and variance-covariance matrix $\boldsymbol{\Theta}_i^m$ is equal to $\boldsymbol{\Theta}_{21,i}^m \mathcal{W}_{m,i}$ and $\boldsymbol{\Theta}_{22,i}^m - \boldsymbol{\Theta}_{21,i}^m \boldsymbol{\Theta}_{12,i}^m$, respectively, with $\boldsymbol{\Theta}_{12,i}^m$, $\boldsymbol{\Theta}_{21,i}^m$ and $\boldsymbol{\Theta}_{22,i}^m$ defined by re-ordering $\boldsymbol{\Upsilon}_i$ as follows

$$\boldsymbol{\Upsilon}_i^m = \left(\begin{array}{c:c} \overbrace{\Theta_{11,i}^m}^{1\times 1} & \overbrace{\boldsymbol{\Theta}_{12,i}^m}^{1\times(M-1)} \\ \hdashline \underbrace{\boldsymbol{\Theta}_{21,i}^m}_{(M-1)\times 1} & \underbrace{\boldsymbol{\Theta}_{22,i}^m}_{(M-1)\times(M-1)} \end{array}\right).$$

The element $\Theta^m_{11,i}$ is equal to 1, the off-diagonal blocks $\boldsymbol{\Theta}^m_{12,i}$ and $\boldsymbol{\Theta}^m_{21,i}$ consist of the correlations $r_{m\varpi,i} = \mathrm{t}_{mm,i}\mathrm{t}_{\varpi\varpi,i}\sigma^*_{m\varpi,i}(2y_{mi}-1)(2y_{\varpi i}-1)$, where $\mathrm{t}_{mm,i}$ and $\mathrm{t}_{\varpi\varpi,i}$ denote the $(m,m)^{\mathrm{th}}$ and $(\varpi,\varpi)^{\mathrm{th}}$ element of matrix $\mathbf{T}_i$, respectively, $\forall\ \varpi \in \{1:M\} \setminus m, m \neq \varpi$, and $\sigma^*_{m\varpi,i}$ is the $(m,\varpi)^{\mathrm{th}}$ element of matrix $\boldsymbol{\Sigma}^*_i$ (matrices $\mathbf{T}_i$ and $\boldsymbol{\Sigma}^*_i$ are defined in Supplementary Material C). The symmetric sub-matrix $\boldsymbol{\Theta}^m_{22,i}$ has main diagonal elements equal to 1 and off-diagonals equal to $r_{\bar{\varphi}\varpi,i} = \mathrm{t}_{\bar{\varphi}\bar{\varphi},i}\mathrm{t}_{\varpi\varpi,i}\sigma^*_{\bar{\varphi}\varpi,i}(2y_{\bar{\varphi}i}-1)(2y_{\varpi i}-1), \forall\bar{\varphi},\varpi \in \{1:M\}\setminus m$, for $\bar{\varphi} \neq \varpi$.

Proof. See Supplementary Material D.1. □

Proposition 3. *Assume that $\boldsymbol{\mathcal{W}}_i$ is a multivariate standardized normal vector with correlation matrix equal to $\boldsymbol{\Upsilon}_i$. Then the first-order derivative of the M-variate normal cdf $\Phi_M(\boldsymbol{\mathcal{W}}_i;\mathbf{0},\boldsymbol{\Upsilon}_i)$ with respect to $\boldsymbol{\beta}_{k_1k_2}, \forall k_1 = 1,\ldots,M-1, k_2 = k_1+1,\ldots M$, can be expressed as*

$$\begin{aligned}\frac{\partial\Phi_M(\boldsymbol{\mathcal{W}}_i;\mathbf{0},\boldsymbol{\Upsilon}_i)}{\partial\boldsymbol{\beta}_{k_1k_2}} &= \Big(\phi_2(\boldsymbol{\mathcal{W}}_{12,i};\mathbf{0},\boldsymbol{\Theta}^{12}_i)\Phi_{M-2}(\boldsymbol{\mathcal{W}}_{-12,i}|\boldsymbol{\mathcal{W}}_{12,i};\boldsymbol{M}^{-12}_i,\boldsymbol{\Theta}^{-12}_i),\ldots,\\ &\quad \phi_2(\boldsymbol{\mathcal{W}}_{M-1,M,i};\mathbf{0},\boldsymbol{\Theta}^{M-1,M}_i)\Phi_{M-2}(\boldsymbol{\mathcal{W}}_{-M-1,M,i}|\boldsymbol{\mathcal{W}}_{M-1,M,i};\\ &\quad \boldsymbol{M}^{-M-1,M}_i,\boldsymbol{\Theta}^{-M-1,M}_i)\Big)\times\left(\frac{\partial r_{12,i}}{\partial\eta_{k_1k_2,i}},\ldots,\frac{\partial r_{M-1,M,i}}{\partial\eta_{k_1k_2,i}}\right)^{\top}\boldsymbol{x}^{\top}_{k_1k_2,i},\end{aligned}$$

where M denotes the total number of equations under a multivariate binary framework, $\boldsymbol{\beta}_{k_1k_2}$ denotes the parameter vector of covariate vector $\boldsymbol{x}_{k_1k_2,i}$, $\boldsymbol{\mathcal{W}}_{k_1k_2,i} = (\mathcal{W}_{k_1,i},\mathcal{W}_{k_2,i})^{\top}$, $\boldsymbol{\mathcal{W}}_{-k_1k_2,i} = (\mathcal{W}_{1,i},\ldots,\mathcal{W}_{k_1-1,i},\mathcal{W}_{k_1+1,i},\ldots,\mathcal{W}_{k_2-1,i},\mathcal{W}_{k_2+1,i},\ldots,\mathcal{W}_{M,i})^{\top}$, $\forall k_1,k_2$, $\mathcal{W}_{k_1,i}$ and $\mathcal{W}_{k_2,i}$ refer to the linear predictors of the k_1^{th} and k_2^{th} equations respectively and are equal to $(2y_{mi}-1)\Phi^{-1}(F_m(\eta_{mi}))$, $\forall m = k_1,k_2$, and $f_m(\eta_{mi})$ and $F_m(\eta_{mi})$ denote the univariate pdf and cdf respectively which can be specified via the normal, logistic and Gumbel distributions. The variance-covariance matrix $\boldsymbol{\Theta}^{k_1k_2}_i$ is equal to $\boldsymbol{\Theta}^{k_1k_2}_{11,i}$, while the mean $\boldsymbol{M}^{-k_1k_2}_i$ and variance-covariance matrix $\boldsymbol{\Theta}^{-k_1k_2}_i$ is equal to $\boldsymbol{\Theta}^{k_1k_2}_{21,i}\left(\boldsymbol{\Theta}^{k_1k_2}_{11,i}\right)^{-1}\boldsymbol{\mathcal{W}}_{k_1k_2}$ and $\boldsymbol{\Theta}^{k_1k_2}_{22,i} - \boldsymbol{\Theta}^{k_1k_2}_{21,i}\left(\boldsymbol{\Theta}^{k_1k_2}_{11,i}\right)^{-1}\boldsymbol{\Theta}^{k_1k_2}_{12,i}$, respectively, $\forall k_1,k_2$. The sub-matrices $\boldsymbol{\Theta}^{k_1k_2}_{11,i}$, $\boldsymbol{\Theta}^{k_1k_2}_{12,i}$, $\boldsymbol{\Theta}^{k_1k_2}_{21,i}$ and $\boldsymbol{\Theta}^{k_1k_2}_{22,i}$ are defined by re-ordering $\boldsymbol{\Upsilon}_i$ as follows

$$
\boldsymbol{\Upsilon}_i^{k_1k_2} = \left(\begin{array}{c:c} \overbrace{\boldsymbol{\Theta}_{11,i}^{k_1k_2}}^{2\times 2} & \overbrace{\boldsymbol{\Theta}_{12,i}^{k_1k_2}}^{2\times(M-2)} \\ \hdashline \underbrace{\boldsymbol{\Theta}_{21,i}^{k_1k_2}}_{(M-2)\times 2} & \underbrace{\boldsymbol{\Theta}_{22,i}^{k_1k_2}}_{(M-2)\times(M-2)} \end{array} \right).
$$

The sub-matrix $\boldsymbol{\Theta}_{11,i}^{k_1k_2}$ has unit diagonals and off-diagonals defined as $r_{k_1k_2,i} = \mathrm{t}_{k_1k_1,i}\mathrm{t}_{k_2k_2,i}\sigma^*_{k_1k_2,i}(2y_{k_1i} - 1)(2y_{k_2i} - 1)$, where $\mathrm{t}_{mm,i}$ denotes the $(m,m)^{\mathrm{th}}$ element of matrix $\mathbf{T}_i$, $\forall m = k_1, k_2$, and $\sigma^*_{k_1k_2,i}$ is the $(k_1,k_2)^{\mathrm{th}}$ element of matrix $\boldsymbol{\Sigma}_i^*$ (matrices $\mathbf{T}_i$ and $\boldsymbol{\Sigma}_i^*$ are defined in Supplementary Material C). The first row (column) of $\Theta_{12,i}^{k_1k_2}$ ($\Theta_{21,i}^{k_1k_2}$) contains the correlations $r_{k_1\bar{\varrho},i}$, for $\bar{\varrho} \in \{1:M\} \setminus k_1$, while the second row (column) of $\Theta_{12,i}^{k_1k_2}$ ($\Theta_{21,i}^{k_1k_2}$) contains the correlations $r_{\bar{\upsilon}k_2,i}$, for $\bar{\upsilon} \in \{1:M\} \setminus k_2$. The diagonal block $\boldsymbol{\Theta}_{22,i}^{k_1k_2}$ is a symmetric matrix with unit diagonals and off-diagonal elements equal to $r_{\bar{\chi}\bar{\psi},i}$, $\forall\, \bar{\chi}, \bar{\psi} \in \{1:M\} \setminus \{k_1,k_2\}$ for $\bar{\chi} \neq \bar{\psi}$.

Proof. See Supplementary Material D.2. □

The construction of confidence intervals, p-values and information criteria, for instance, are not essentially changed by the extensions introduced in this paper and we refer the reader to the supplementary material of Filippou et al. (2017) for such details.

5 Simulation Study

To gain some insights into the practical performance of the proposed approach, we conducted a simulation study. We considered three binary outcomes, one binary covariate and one continuous regressor. The chosen link functions were logit, cloglog and probit. Exact simulation settings are given in the Supplementary Material E. The syntax to fit the proposed trivariate binary model is

```
out  <- gjrm(formula = f.l, data = dat, Chol = TRUE, Model = "T",
             margins = c("logit", "cloglog", "probit"))
```

where `f.l` consists of a list of six equations

```
eq1 <- y1 ~ v1 + s(z1)
eq2 <- y2 ~ v1 + s(z1)
eq3 <- y3 ~ v1 + s(z1)
```

```
eq12 <-   ~ v1 + s(z1)
eq13 <-   ~ v1 + s(z1)
eq23 <-   ~ v1 + s(z1)
f.l <- list(eqn1, eqn2, eqn3, eq12, eq13, eq23)
```

`v1` and `z1` denote the binary and continuous covariates, respectively, `s()` represents a smooth function that is set up using a penalised thin plate regression spline with 10 bases and penalty based on second order derivatives, the last three equations in `f.l` refer to the additive predictors for the correlation parameters ϑ_{12}, ϑ_{13} and ϑ_{23}, `data` is a data frame containing the variables in the model, `Chol = TRUE` indicates that the modified Cholesky decomposition approach has to be employed, `Model` indicates the type of model (`"T"` for trivariate binary model) and `margins` the three the link functions.

Figures 1 and 2 depict linear and non-linear estimates obtained when applying the proposed approach. Overall, the mean estimates are close to the true values and, as expected, their variability decreases as the sample size grows large. The main exception is perhaps the parametric component of the additive predictor related to ϑ_{23}, where at $n = 1000$ the estimates exhibit some bias and a larger variability as compared to the other parameters. Also note that the uncertainty of the estimates for all the components in the correlations' additive predictors is higher than that of the estimates for the three marginal equations. This is not so surprising given the complexity of the proposed model and the fact that the correlation parameters are usually more difficult to estimate in a flexible regression setting when the outcomes are binary. Overall, the results improve considerably as n increases.

6 Empirical illustration

We illustrate the potential of the proposed model using 2007-2008 birth data from the North Carolina Center for Health Statistics (`http://www.schs.state.nc.us/`). The data contain information on $64,690$ male newborns and builds upon the analysis conducted in Filippou et al. (2017). The choice of variables included in the model was mainly driven by previous work on the subject (e.g., South et al., 2012; Neelon et al., 2014), and the responses are plurality (`mb`, a

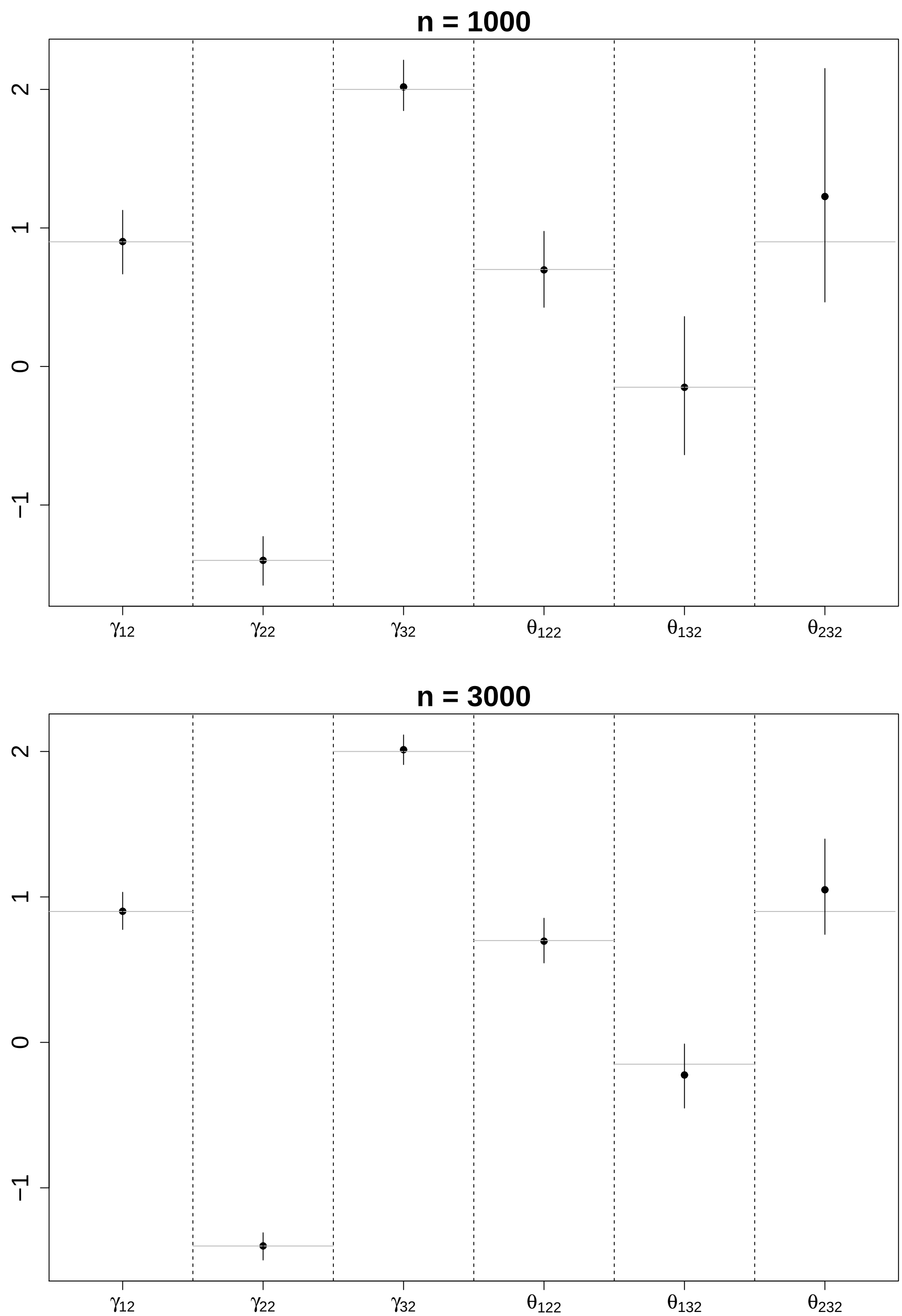

Figure 1: Linear coefficient estimates obtained by applying the proposed model to data simulated from a trivariate Gaussian copula model with logistic, Gumbel and normal margins. Circles indicate mean estimates while bars represent the estimates' ranges resulting from 5% and 95% quantiles. True values are indicated by gray horizontal lines.

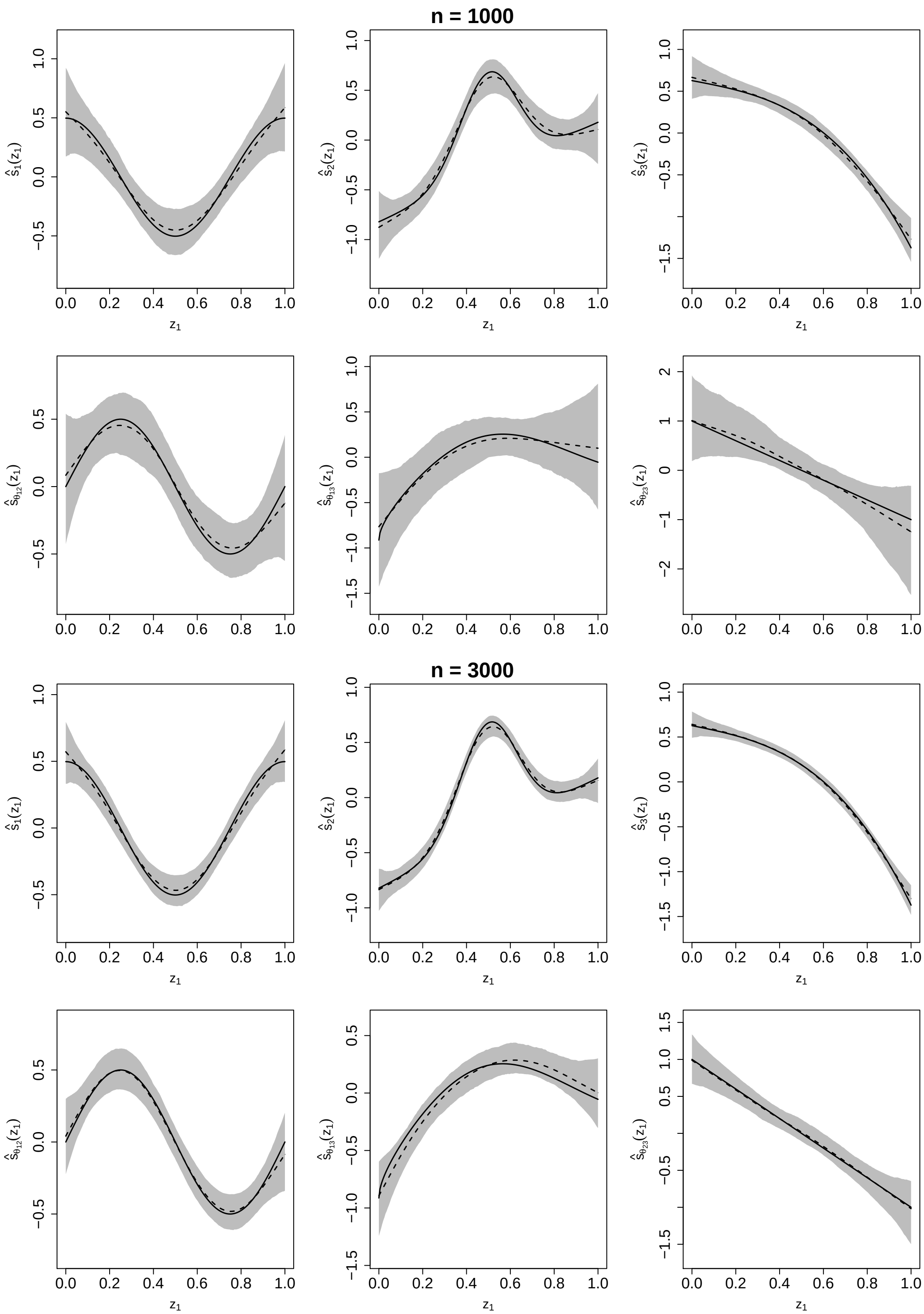

Figure 2: Smooth function estimates obtained by applying the proposed model to data simulated from a trivariate Gaussian copula model with logistic, Gumbel and normal margins. True functions are represented by black solid lines, mean estimates by dashed lines and point-wise ranges resulting from 5% and 95% quantiles by shaded areas.

binary variable that takes value 1 for singleton birth and 0 otherwise), infant's birth weight (`lbw`, which takes value 1 when weight is less than 2500 grams and 0 otherwise) and preterm birth (`ptb` that takes value 1 if the number of gestation weeks is less than 37 and 0 otherwise). The covariates are maternal race categorized as non-white and white (`nwhite`), smoking status with 1 indicating a mother smoking during pregnancy (`smoker`), weight gained by mother during pregnancy in pounds (`gained`), age of mother in years (`mage`) and county in which the birth occurred (`county`).

Filippou et al. (2017) built a model for the joint analysis of `mb`, `lbw` and `ptb`, and showed the impacts that the model's covariates have on the responses as well as some joint probabilities of interest. Here, the focus is on alternative specifications for the link functions and on understanding how the association between the three outcomes is modified by the presence of covariates. We started off with the specification adopted by Filippou et al. (2017) where all model's additive predictors contained all the covariates available in the data. That is, all additive predictors included $\texttt{nwhite}_i$, $\texttt{smoker}_i$, $s(\texttt{gained}_i)$, $s(\texttt{mage}_i)$ and $s_{\texttt{spatial}}(\texttt{county}_i)$, where the smooth functions of $\texttt{gained}_i$ and $\texttt{mage}_i$ were represented using penalized thin plate regression splines, and the spatial smooth for the regional effects was set up using a Markov random field approach (Wood, 2017). To simplify the model building process we used the fact that the specification for the marginal models and their dependence can be addressed separately. For each margin we fitted three univariate GAMs based on the probit, logit and cloglog links. For each margin and link the covariate effects were always all significant. The links chosen were logit, logit and cloglog for `mb`, `lbw` and `ptb`. We then focused on the correlations' additive predictors and viewed all of their covariates effects as being part of a unique equation. We employed the classic backward selection procedure and also looked at the significance of the effects to favor more parsimonious specifications. The additive predictors for the six equations of the final model are:

$$
\begin{aligned}
\eta_{1i} &= \gamma_{11} + \gamma_{12}\texttt{nwhite}_i + \gamma_{13}\texttt{smoker}_i + s_{11}(\texttt{gained}_i) + s_{12}(\texttt{mage}_i) + s_{1\texttt{spatial}}(\texttt{county}_i),\\
\eta_{2i} &= \gamma_{21} + \gamma_{22}\texttt{nwhite}_i + \gamma_{23}\texttt{smoker}_i + s_{21}(\texttt{gained}_i) + s_{22}(\texttt{mage}_i) + s_{2\texttt{spatial}}(\texttt{county}_i),\\
\eta_{3i} &= \gamma_{31} + \gamma_{32}\texttt{nwhite}_i + \gamma_{33}\texttt{smoker}_i + s_{31}(\texttt{gained}_i) + s_{32}(\texttt{mage}_i) + s_{3\texttt{spatial}}(\texttt{county}_i),\\
\eta_{12i} &= \gamma_{12,1} + \gamma_{12,2}\texttt{nwhite}_i + s_{12}(\texttt{gained}_i) + s_{12\texttt{spatial}}(\texttt{county}_i),\\
\eta_{13i} &= \gamma_{13,1} + \gamma_{13,2}\texttt{nwhite}_i + \gamma_{13,3}\texttt{smoker}_i + s_{13,1}(\texttt{gained}_i) + s_{13,2}(\texttt{mage}_i) + s_{13\texttt{spatial}}(\texttt{county}_i),\\
\eta_{23i} &= \gamma_{23,1} + s_{23,1}(\texttt{gained}_i) + s_{23,2}(\texttt{mage}_i),
\end{aligned}
$$

Some results are presented below.

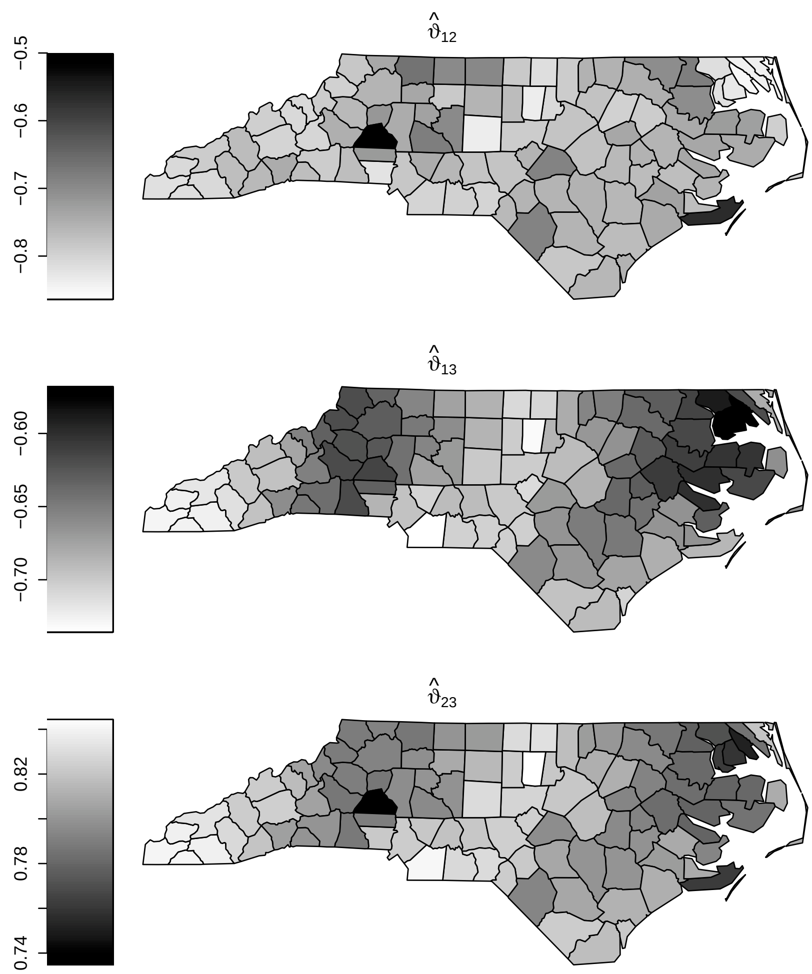

Figure 3: Spatially varying estimates of correlations ϑ_{12} ϑ_{13} and ϑ_{23} obtained by applying the proposed approach to North Carolina data.

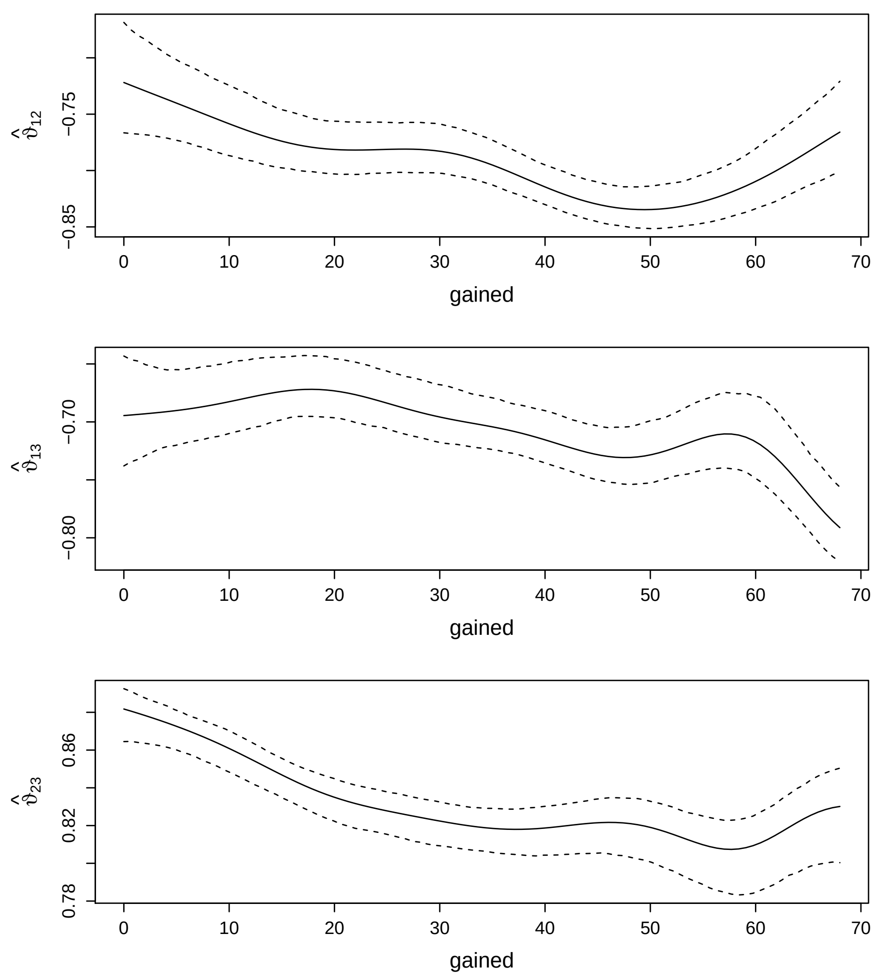

Figure 4: Estimates of correlations ϑ_{12} ϑ_{13} and ϑ_{23} by `gained` obtained by applying the proposed approach to North Carolina data. Point-wise 95% confidence intervals were obtained using the posterior simulation approach described in Filippou et al. (2017).

Figure 3 shows the estimated model's correlations by county in North Carolina. Here, the effects for two binary predictors in the model were set to zero (since the majority of individuals are white and non smokers) while the continuous regressors were set at their average values. Figure 4 displays the estimated correlations by `gained` where the two binary predictors were set at 0, `mage` at its average value and `county` was randomly chosen (although results were very similar across counties).

Generally, the three binary outcomes are strongly correlated with each other even after accounting for covariates at marginal level. Interestingly, as shown in Figure 3, there is a good deal of spatial variation in the strength of the correlations. Specifically, the three responses seem to be more strongly related in the west and central areas of North Carolina than they are otherwise. Figure 4 suggests that the absolute association between `mb` and `lbw` increases for values of `gained` up to 50 and then decreases, the correlation between `mb` and `ptb` overall increases, and the dependence between `lbw` and `ptb` decreases for values of `gained` between 50 and 60 and then increases. These are new findings which open up questions for further research to elucidate the nature of such dependencies in North Carolina.

7 Conclusions

We have proposed a generalisation of the trivariate additive probit model which allows for virtually any parametric link function and for the model's correlation coefficients to depend on flexible additive predictors. The parameters of the model can be estimated simultaneously within a penalized likelihood framework based on a trust region algorithm with automatic smoothing parameter selection, and the model can be easily employed via the `gjrm()` function in the `R` package `GJRM`. The potential of the approach has been demonstrated using simulated and real data.

The proposed extensions are of some applied relevance as link functions other than probit are often used in medical studies and understanding how the residual association between response variables is related to covariates can help to model more general forms of multivariate dependence. We plan to extend the trivariate model to other types of marginal outcomes (e.g., continuous, discrete). This will considerably extend to scope and applicability of the trivariate modelling approach introduced in this article.

Acknowledgement

We would like to thank Nadja Klein for the useful discussions on the modified Cholesky decomposition proposed in this paper.

References

Chen, Z. & Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics*, 59(4), 762–769.

Eilers, P. H. & Marx, B. D. (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 11(2), 89–102.

Filippou, P., Marra, G., & Radice, R. (2017). Penalized likelihood estimation of a trivariate additive probit model. *Biostatistics*, 18(3), 569–585.

Joe, H. (1997). *Multivariate models and multivariate dependence concepts*. CRC Press.

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

Król, A., Ferrer, L., Pignon, J.-P., Proust-Lima, C., Ducreux, M., Bouché, O., Michiels, S., & Rondeau, V. (2016). Joint model for left-censored longitudinal data, recurrent events and terminal event: Predictive abilities of tumor burden for cancer evolution with application to the ffcd 2000–05 trial. *Biometrics*.

Loureiro, M. L., Sanz-de Galdeano, A., & Vuri, D. (2010). Smoking habits: like father, like son, like mother, like daughter? *Oxford Bulletin of Economics and Statistics*, 72(6), 717–743.

Marra, G. & Radice, R. (2017a). Bivariate copula additive models for location, scale and shape. *Computational Statistics & Data Analysis*, 112, 99–113.

Marra, G. & Radice, R. (2017b). *GJRM: Generalised Joint Regression Modelling*. R package version 0.1-1.

Neelon, B., Anthopolos, R., & Miranda, M. L. (2014). A spatial bivariate probit model for correlated binary data with application to adverse birth outcomes. *Statistical Methods in Medical Research*, 23(2), 119–133.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3), 677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika*, 87(2), 425–435.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: orthogonality of variance–correlation parameters. *Biometrika*, 94(4), 1006–1013.

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

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554.

Sklar, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

South, A. P., Jones, D. E., Hall, E. S., Huo, S., Meinzen-Derr, J., Liu, L., & Greenberg, J. M. (2012). Spatial analysis of preterm birth demonstrates opportunities for targeted intervention. *Maternal and Child Health Journal*, 16(2), 470–478.

Trivedi, P. K. & Zimmer, D. M. (2007). *Copula modeling: an introduction for practitioners*. Now Publishers Inc.

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

Zhang, R., Inder, B. A., & Zhang, X. (2015). Bayesian estimation of a discrete response model with double rules of sample selection. *Computational Statistics & Data Analysis*, 86, 81–96.

Zhong, W., Koopmeiners, J. S., & Carlin, B. P. (2012). A trivariate continual reassessment method for phase i/ii trials of toxicity, efficacy, and surrogate efficacy. *Statistics in Medicine*, 31(29), 3885–3895.

Zimmer, D. M. & Trivedi, P. K. (2006). Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business & Economic Statistics*, 24(1), 63–76.