Research Article

Statistics in Medicine

Received 11 July 2009, Accepted 23 August 2010 Published online 20 October 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4087

Copula-based regression models for a bivariate mixed discrete and continuous outcome

A. R. de Leon[*†] **and B. Wu**

This paper is concerned with regression models for correlated mixed discrete and continuous outcomes constructed using copulas. Our approach entails specifying marginal regression models for the outcomes, and combining them via a copula to form a joint model. Specifically, we propose marginal regression models (e.g. generalized linear models) to link the outcomes' marginal means to covariates. To account for associations between outcomes, we adopt the Gaussian copula to indirectly specify their joint distributions. Our approach has two advantages over current methods: one, regression parameters in models for both outcomes are marginally meaningful, and two, the association is 'margin-free', in the sense that it is characterized by the copula alone. By assuming a latent variable framework to describe discrete outcomes, the copula used still uniquely determines the joint distribution. In addition, association measures between outcomes can be interpreted in the usual way. We report results of simulations concerning the bias and efficiency of two likelihood-based estimation methods for the model. Finally, we illustrate the model using data on burn injuries. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: inference function for margins; joint analysis; likelihood estimation; marginal models; mixed binary-normal data; t-distribution

1. Introduction

Mixed (or non-commensurate) outcomes are ubiquitous in applications in health and medicine, and joint analysis of such outcomes entails specification of models flexible enough to accommodate them. Such joint models are potentially advantageous in several statistical and practical respects. For example, a multivariate model enables analysts to account for relationships between outcomes and assess at the same time the joint influence of predictors/covariates on them. From a statistical standpoint, joint analysis avoids multiple testing and naturally leads to global tests, thus resulting in increased power and better control of Type I error rates [1]. Significant efficiency gains over separate univariate analyses have also been reported, especially in settings where there are missing data (see, e.g. [2]).

A number of joint modelling strategies for mixed outcomes have been studied in the literature; see Teixeira-Pinto and Normand [3] for a recent survey. The challenge is that models for joint distributions of mixed outcomes are uncommon. One of the earliest proposals directly specifies the joint distribution by factorizing it into a conditional distribution of one set of outcomes and a marginal distribution of the other set. This suggest two formulations of mixed-outcome joint distributions: a marginal distribution for discrete outcomes and a conditional distribution for continuous outcomes, given discrete outcomes, or a marginal distribution for continuous outcomes and a conditional distribution for discrete outcomes, given continuous outcomes. The first formulation has received much attention in mixed-data literature, beginning with Olkin and Tate's [4] so-called general location model, which assumes different multivariate normal distributions for continuous outcomes given discrete outcomes, and a marginal multinomial distribution for the latter (see also [5, 6]). Examples of the second formulation include Cox and Wermuth [7], who suggest a logistic or probit conditional distribution for binary given continuous outcomes and a marginal multivariate normal distribution for the latter. Mixed-data models based on latent variables have likewise been studied [8, 9]; models based on the Plackett–Dale idea [10] can also be viewed within this framework. Refinements and extensions of both formulations have since been investigated by several authors. Recent references include de Leon and Carrière [11], George *et al*. [12], and Yang *et al*. [13], among others.

Indirect approaches to specifying mixed-outcome joint distributions are also possible. One approach introduces shared or correlated random effects to incorporate correlations between mixed outcomes in the resulting joint model. Munkin and

Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4
[*]*Correspondence to: A. R. de Leon, Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada T2N 1N4.*
[†]*E-mail: adeleon@math.ucalgary.ca*

Trivedi [14] adopt the approach in a health-care demand application. Timbie and Normand [15] and Daniels and Normand [16] use this strategy, albeit in a Bayesian framework, to profile health-care units. Applications to high-dimensional mixed data analysis are provided by Faes *et al*. [17].

A recent alternative strategy involves the use of copulas, as discussed in Kolev and Paiva [18], Song *et al*. [19], and Zimmer and Trivedi [20], among others. In the bivariate case, the approach relates an arbitrary joint distribution $F_{X,Y}$ to its corresponding univariate marginal distributions F_X and F_Y via a copula C as

$$F_{X,Y}(x,y) = C(u,v;\rho), \tag{1}$$

where u and v are respective realizations of probability integral transforms $U = F_X(X) \sim \text{uniform}[0,1]$ and $V = F_Y(Y) \sim \text{uniform}[0,1]$, and ρ is the dependence parameter measuring dependence between marginals F_X and F_Y. The distribution $F_{X,Y}$ is thus specified via its marginals and a copula that 'glues' them together. In a parametric context, note that the marginals need not come from the same parametric families, allowing researchers great flexibility in modelling different outcomes. The parameter ρ accounts for 'dependence' between variables in a way that is separate from their marginal specifications.

Sklar's Theorem [21, p. 126] states that copula (1) is unique if and only if X and Y are continuous random variables. However, the same is not true for discrete random variables, in which case the copula is only uniquely identified on the Cartesian product of the ranges of the marginals; moreover, the uniformity of the probability integral transforms does not hold in the discrete case [22, 23]. Genest and Nešlehovà [24] caution against the dangers and limitations of careless applications of copula to discrete settings. One consequence of this is that common association measures like Kendall's tau τ and Spearman's rho ρ_S [21, pp. 129–132] may now depend on the margins; another is that interpretations of parameter ρ in terms of dependence become problematic in some cases.

In this paper, we introduce a new joint regression model constructed from copulas that overcomes many difficulties of earlier approaches [25] to analysis of mixed data. The model adopts a latent variable formulation of the discrete outcome and is related to an earlier regression model studied by Song *et al*. [19]; in fact, it can be viewed as an extension of their model to mixed polychotomous and continuous outcomes. We take a fully parametric specification of the marginal models, and a parametric copula is adopted to model between-outcome associations. This approach is quite flexible since it does not resort to conditioning, resulting in a symmetrical treatment of outcomes while preserving the marginal interpretations of regression parameters, an attractive feature in practice. Likelihood-based methods are used for estimation, affording us a whole battery of available procedures for model inference, model checking, and validation. To illustrate our approach, we re-analyze Song *et al*.'s [19] burn injury data via a version of our model with a normal continuous outcome and a robit model [26] for the binary response. The latter provides a rich class of models, including logistic and probit regression models as special cases, for analysis of binary data.

The paper is organized as follows. We introduce a class of copula-based regression models for bivariate mixed outcomes in Section 2. Section 3 next discusses two likelihood-based methods for model estimation: one based on the full likelihood and another on the method of inference functions for margins [27]. Section 4 reports simulation results concerning two special cases of our model, namely the robit-normal and probit-exponential models. In Section 5, data from burn injuries are used to illustrate our methodology. Section 6 concludes the paper.

2. Regression model for mixed bivariate outcome

Consider correlated mixed outcomes X_i and Y_i obtained from each of N subjects, where Y_i is continuous and X_i is discrete. These outcomes may arise in a study involving patients suffering from burn injuries, for example, where the discrete outcome represents death or survival from burn injury and the continuous outcome corresponds to total burn area [19]. Assume $X_i \sim F_{X_i}$ and $Y_i \sim F_{Y_i}$, $i = 1, \ldots, N$. In addition, let X_i have $C+1$ distinct values, say, $s_0, s_1, \ldots, s_C$, possibly representing ordinal scores or nominal states. To model the joint distribution F_{X_i,Y_i} of X_i and Y_i, let $Y_i^* \sim F_{Y_i^*}$ be the unobserved continuous latent variable underlying X_i, such that

$$X_i = \begin{cases} s_0, & \text{if } Y_i^* \in (-\infty, \gamma_1) \\ \vdots & \quad\vdots \\ s_k, & \text{if } Y_i^* \in [\gamma_k, \gamma_{k+1}) \\ \vdots & \quad\vdots \\ s_C, & \text{if } Y_i^* \in [\gamma_C, \infty) \end{cases}, \tag{2}$$

where $\gamma_1 < \cdots < \gamma_C$ are unknown thresholds, with $\gamma_0 = -\infty$ and $\gamma_{C+1} = \infty$. In what follows, we use the Gaussian copula, an important copula family which has been used in a variety of applications (e.g. [19, 28]) for its flexibility and analytical tractability. While the choice of the copula is arbitrary, adopting a Gaussian copula is convenient because of its nice marginal and conditional properties.

Assuming the joint distribution $F_{Y_i^*, Y_i}$ of Y_i^* and Y_i is determined by a Gaussian copula, we get

$$F_{Y_i^*, Y_i}(y^*, y) = \Phi_2(\Phi^{-1}\{F_{Y_i^*}(y^*)\}, \Phi^{-1}\{F_{Y_i}(y)\}; \rho), \tag{3}$$

where Φ is the standard normal distribution, Φ_2 is the standard bivariate normal distribution (i.e. zero means and unit variances) with correlation ρ, and

$$E(Y_i^*) = \mu_{1i}(\mathbf{z}_{1i}, \boldsymbol{\alpha}), \quad E(Y_i) = \mu_{2i}(\mathbf{z}_{2i}, \boldsymbol{\beta}). \tag{4}$$

Here, the margins $F_{Y_i^*}$ and F_{Y_i} are absolutely continuous distributions, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of regression coefficients, $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$ are outcome-specific covariate vectors, and μ_{1i} and μ_{2i} are link functions specifying how the covariates are incorporated in the marginal means, as for example, in the linear models $\mu_{1i} = \mathbf{z}_{1i}^\top \boldsymbol{\alpha}$ and $\mu_{2i} = \mathbf{z}_{2i}^\top \boldsymbol{\beta}$. For identifiability reasons, we assume that Y_i^* has unit variance (or scale parameter) and $\boldsymbol{\alpha}$ has no intercept term; in the binary case, we may arbitrarily assume the single cutpoint in (2) to be zero, and $\boldsymbol{\alpha}$ will need to include an intercept term (e.g. [9]).

The joint distribution of X_i and Y_i is then given by

$$P(X_i = x, Y_i \leqslant y) = \begin{cases} F_{Y_i^*, Y_i}(\gamma_1, y) & \text{if } x = s_0 \\ \vdots & \vdots \\ F_{Y_i^*, Y_i}(\gamma_{k+1}, y) - F_{Y_i^*, Y_i}(\gamma_k, y) & \text{if } x = s_k \\ \vdots & \vdots \\ F_{Y_i}(y) - F_{Y_i^*, Y_i}(\gamma_C, y) & \text{if } x = s_C \end{cases}, \tag{5}$$

where $F_{Y_i^*, Y_i}$ is defined in (3). The joint density $f_{X_i, Y_i}(x, y) = \partial P(X_i = x, Y \leqslant y)/\partial y$ of X_i and Y_i follows in a straightforward way:

$$f_{X_i, Y_i}(x, y) = \begin{cases} \Phi\left(\dfrac{q_1^1 - \rho q_2}{\sqrt{1-\rho^2}}\right) f_{Y_i}(y) & \text{if } x = s_0 \\ \vdots & \vdots \\ \left\{\Phi\left(\dfrac{q_1^{k+1} - \rho q_2}{\sqrt{1-\rho^2}}\right) - \Phi\left(\dfrac{q_1^k - \rho q_2}{\sqrt{1-\rho^2}}\right)\right\} f_{Y_i}(y) & \text{if } x = s_k \\ \vdots & \vdots \\ \Phi\left(-\dfrac{q_1^C - \rho q_2}{\sqrt{1-\rho^2}}\right) f_{Y_i}(y) & \text{if } x = s_C \end{cases}, \tag{6}$$

where f_{Y_i} is the marginal density of Y_i, $q_1^k = \Phi^{-1}(u_1^k) = \Phi^{-1}\{F_{Y_i^*}(\gamma_k)\}$, $q_2 = \Phi^{-1}(u_2) = \Phi^{-1}\{F_{Y_i}(y)\}$, and ϕ is the standard normal density. Observe that (6) is a proper joint density.

Our model, being completely determined by the marginal distributions and the copula, yields marginally interpretable regression parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. It is also possible to obtain the conditional distribution of outcomes, an important requirement in situations where conditional behaviour of outcomes is of interest.

Song *et al*. [19] (see also [21], Chapter 6) recently adapted copula-based regression analysis to mixed-outcome settings using the family of Gaussian copula-generated exponential dispersion models. While our model can be viewed as an extension of their model to mixed polychotomous discrete and continuous settings, our use of a latent variable formulation of discrete outcome X_i leads to a completely unique copula model [29]. In addition, the dependence parameter represents the Pearson correlation coefficient between the normal scores based on Y_i and Y_i^*, and can thus be interpreted as a proxy for the polyserial correlation between Y_i and X_i, a common association measure between continuous and discrete outcomes [30].

Taking both F_{Y_i} and $F_{Y_i^*}$ as normal results in a model akin to those studied by Gueorguieva and Agresti [31] and Catalano and Ryan [9]. Alternatively, a logistic latent distribution for the latent variable results in a binary-outcome model similar to Nikoloulopoulos and Karlis' [32] logit copula model. It is also possible to model $F_{Y_i^*}$ using a (generalized) t-distribution [33] to yield a robit regression model [26] for binary outcomes, a robust and general alternative to probit and logit regression. However, our approach may not be suitable to mixed continuous and count data; in addition, the notion of continuous latent variables underlying nominal outcomes may not be appropriate.

3. Estimation

Let $\{x_i, y_i, \mathbf{z}_{1i}, \mathbf{z}_{2i}\}$, $i=1,\ldots,N$, denote the observed data, with outcome-specific covariate vectors $\mathbf{z}_{1i}$ and $\mathbf{z}_{2i}$. From (6) with $\boldsymbol{\theta}$ as the parameter vector containing the regression parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, the association parameter ρ, and respective marginal parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of $F_{Y_i^*}$ and F_{Y_i}, the log-likelihood function is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log f_{Y_i}(y_i;\mathbf{z}_{2i},\boldsymbol{\beta},\boldsymbol{\theta}_2) + \sum_{\substack{k=0,\ldots,C\\ \forall x_i=s_k}} \log\left\{\Phi\left(\frac{q_{1i}^{k+1}(\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)-\rho q_{2i}(\mathbf{z}_{2i},\boldsymbol{\beta},\boldsymbol{\theta}_2)}{\sqrt{1-\rho^2}}\right)\right.$$
$$\left. -\Phi\left(\frac{q_{1i}^{k}(\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)-\rho q_{2i}(\mathbf{z}_{2i},\boldsymbol{\beta},\boldsymbol{\theta}_2)}{\sqrt{1-\rho^2}}\right)\right\}. \tag{7}$$

with $q_{1i}^{0}(\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)=-\infty$ and $q_{1i}^{C+1}(\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)=\infty$, and where we include $\boldsymbol{\alpha},\boldsymbol{\beta},\mathbf{z}_{1i},\mathbf{z}_{2i},\boldsymbol{\theta}_1$, and $\boldsymbol{\theta}_2$ in notations to emphasize that the margins depend on them through (4). Note that $q_{1i}^{k}(\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)=\Phi^{-1}\{F_{Y_i^*}(\gamma_k;\mathbf{z}_{1i},\boldsymbol{\alpha},\boldsymbol{\theta}_1)\}$, $k=1,\ldots,C$, and $q_{2i}(\mathbf{z}_{2i},\boldsymbol{\beta},\boldsymbol{\theta}_2)=\Phi^{-1}\{F_{Y_i}(y_i;\mathbf{z}_{2i},\boldsymbol{\beta},\boldsymbol{\theta}_2)\}$, for $i=1,\ldots,N$.

3.1. Joint estimation

Putting $s(\boldsymbol{\theta})=\partial\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ as the score function and $h(\boldsymbol{\theta})=\partial^2\ell(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ as the Hessian matrix, the maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ is obtained by solving the likelihood equations $s(\boldsymbol{\theta})=\mathbf{0}$ iteratively via a Newton–Raphson updating scheme. It can be easily verified that $\widehat{\boldsymbol{\theta}}$ is consistent and asymptotically multivariate normal with mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})=E\{-h(\boldsymbol{\theta})\}=E\{s(\boldsymbol{\theta})s^\top(\boldsymbol{\theta})\}$. Standard errors (SEs) for $\widehat{\boldsymbol{\theta}}$ are calculated from diagonals of $\{s(\widehat{\boldsymbol{\theta}})s^\top(\widehat{\boldsymbol{\theta}})\}^{-1}$; we may also use $-h^{-1}(\widehat{\boldsymbol{\theta}})$, provided the Hessian matrix is invertible.

3.2. Marginal estimation

Direct maximization of (7) may be computationally prohibitive in practice. To circumvent this, (7) may be replaced by a more computationally tractable function, a so-called pseudo-likelihood function. One such alternative which achieves important computational economies is Joe and Xu's [27] method of inference functions for margins (IFM), which estimates marginal parameters solely from margins and only uses the copula as basis for estimating association parameters. The method is particularly suited to our setting, where the marginal models and the dependence between outcomes are specified separately. Using the theory of inference functions [27], it can be shown that the IFM estimate $\widetilde{\boldsymbol{\theta}}$ is asymptotically multivariate normal with mean $\boldsymbol{\theta}$ and covariance matrix $\mathbf{V}=\mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1}$ [21, Chapter 3], where $\mathbf{J}$ is a block-diagonal matrix with symmetric diagonal blocks and $\mathbf{K}$ is a symmetric block matrix; explicit forms of $\mathbf{J}$ and $\mathbf{K}$ for the robit-normal model below are detailed in the online supplementary material. Standard errors of $\widetilde{\boldsymbol{\theta}}$ are obtained from the diagonals of $\widetilde{\mathbf{V}}=\widetilde{\mathbf{J}}^{-1}\widetilde{\mathbf{K}}\widetilde{\mathbf{J}}^{-1}$, where $\widetilde{\mathbf{J}}$ and $\widetilde{\mathbf{K}}$ are the respective estimates of $\mathbf{J}$ and $\mathbf{K}$ obtained from $\widetilde{\boldsymbol{\theta}}$. Computational and statistical performance (i.e. bias and efficiency) of this method has been shown to range from acceptably good to excellent.

4. Simulation study

In this section, the performance of joint and marginal estimates are evaluated in two settings: the first considers a joint model assembled from a marginal normal specification for the continuous outcome and a marginal t-distribution for the latent variable underlying the binary outcome; the second is based on a marginally exponentially distributed outcome and a marginal probit model for a trichotomous discrete outcome. In both cases, we adopt the Gaussian copula to construct the joint model. Results indicate that joint estimation should be preferred to the marginal approach in settings comprising mixtures of continuous and so-called 'high-information' non-binary discrete outcomes; however, the two methods perform generally similarly for mixed binary and continuous data. It also appears that both sets of estimates of scale/variance and correlation parameters exhibit slight bias, which decreases with increasing sample size.

4.1. Robit-normal model

Let continuous $Y_i \sim \text{normal}(\mu_{2i} = \beta_1 + \beta_2 z_{2i}, \sigma^2)$, $i = 1, \ldots, N$, and assume, underlying binary X_i, a continuous latent variable $Y_i^* \sim t_1(\mu_{1i} = \alpha_1 + \alpha_2 z_{1i}, 1, v)$ (i.e. the univariate t-distribution with mean μ_{1i}, unit scale, and degrees of freedom v; see [33]) with density

$$f_{Y_i^*}(y^*) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\pi v}\,\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{(y^* - \mu_{1i})^2}{v}\right)^{-\frac{v+1}{2}}, \tag{8}$$

such that $X_i = I\{Y_i^* > 0\}$ (i.e. threshold model (2) with $C = 1, s_0 = 0$, and $s_C = 1$). The t-latent distribution is assumed to have unit scale for identifiability reasons; in addition, we arbitrarily take the threshold as zero, which then requires that $\boldsymbol{\alpha}$ contains an intercept term. The t-distribution has attracted the attention of researchers as a robust alternative to a normal latent distribution [see, e.g. [26, 34]]. Liu [26] introduced the term 'robit regression' based on a t-latent distribution as a robust alternative to and extension of logit and probit regression models.

A total of $R = 3000$ repeated samples of sizes $N = 100$ and 200 were generated with $\alpha_1 = 3, \alpha_2 = 4, \beta_1 = 1, \beta_2 = 2$, $v = 5$, and $\sigma = 1$, and with varying correlation $\rho = 0, 0.25, 0.5, 0.75, 0.9$. To avoid problems associated with estimating v [33, pp. 235–236], we employ the method of profile likelihood [35]. The method entails maximizing the log-likelihood at fixed grid points $v \in (2, B]$, for some suitably large constant B; in the simulations, we fixed $B = 8$. Our estimates $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\sigma}$, and $\widehat{\rho}$ correspond to $\widehat{v}$ at which the profile likelihood is maximum on $(2, B]$. This method is easy to implement using dense grid points on $(2, B]$, and is more computationally efficient than full maximum likelihood estimation. Asymptotic properties analogous to those for MLEs can be similarly established.

Alternatively, we adopt the IFM method to estimate marginal parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma$, and v from the respective marginal log-likelihoods, and use the joint log-likelihood only to estimate the correlation ρ. For the binary part, we use the method of (marginal) profile likelihood by maximizing the (marginal) profile log-likelihood based on (8) at fixed grid points $v \in (2, B]$; estimates of the continuous data parameters are easily obtained by least squares. To estimate ρ, we maximize the (joint) log-likelihood evaluated at the marginal estimates. Asymptotic results similar to those for joint estimates apply as well to these estimates.

Both methods were implemented in R using the optimization function `optim`, with $\rho \in (-1, 1)$ re-parameterized as Fisher's z-transformation $\eta = \log\{(1+\rho)/(1-\rho)\}/2$.

Figure 1 presents results on the relative bias of joint and marginal estimates for $\rho = 0, 0.25, 0.5, 0.75, 0.9$. Note that relative bias is undefined for $\widehat{\rho}$ at $\rho = 0$, hence, we only considered $\rho = 0.25, 0.5, 0.75, 0.9$. Using the approximate margin of error $1.96 \times (\text{SD/parameter})$ based on $R = 3000$ simulation repeats, relative biases clearly suggest that both full and marginal likelihood approaches yield reasonably unbiased estimates of the regression parameters; however, there is clear bias in both marginal and joint estimates of σ and ρ which improves from moderate to slight with increasing sample size.

Figure 2 displays relative efficiencies of both sets of estimates. From the plots we can see that the efficiencies of estimates of regression parameters are all relatively close to unity, indicating that our methods of precision estimation yield SEs which reflect true sampling variability; almost all their efficiency values, except for a few corresponding to $N = 100$, fall within the approximate margins of error based on the normal approximation. Marginal and joint estimates of σ and ρ show slight inefficiency, which slowly deteriorated with increasing ρ.

Literature on mixed-data analysis suggests that joint estimation can result in significant efficiency gains for discrete outcomes but little to no improvement for continuous outcomes, especially in missing data contexts [2]. However, joint and marginal methods showed similar results in our simulations, notwithstanding slight differences. This lack of efficiency gain from joint estimation appears to be due to the binary (or so-called 'low-information') nature of the discrete outcome. We investigate this further in the next simulation.

4.2. Probit-exponential model

We next consider the case of an exponential-distributed continuous outcome and a probit model for a trichotomous discrete response. Let continuous outcome $Y_i \sim \text{exponential}(\mu_{2i} = \exp\{\beta_1 + \beta_2 z_{2i}\})$ and take the continuous latent variable $Y_i^* \sim \text{normal}(\mu_{1i} = \alpha z_{1i}, 1)$, such that discrete outcome X_i is linked to Y_i^* by threshold model (2) with $C = 2, s_0 = 0, s_1 = 1$, and $s_C = 2$. Note that no intercept is included in latent mean μ_{1i} as it is not identifiable from the cutpoints γ_1 and γ_2; $R = 5000$ simulated data of size $N = 200$ were generated from the joint model constructed from a Gaussian copula with $\rho = 0.8$ and the specified probit and exponential margins. Estimation for the model is done jointly and marginally as outlined in Section 3.

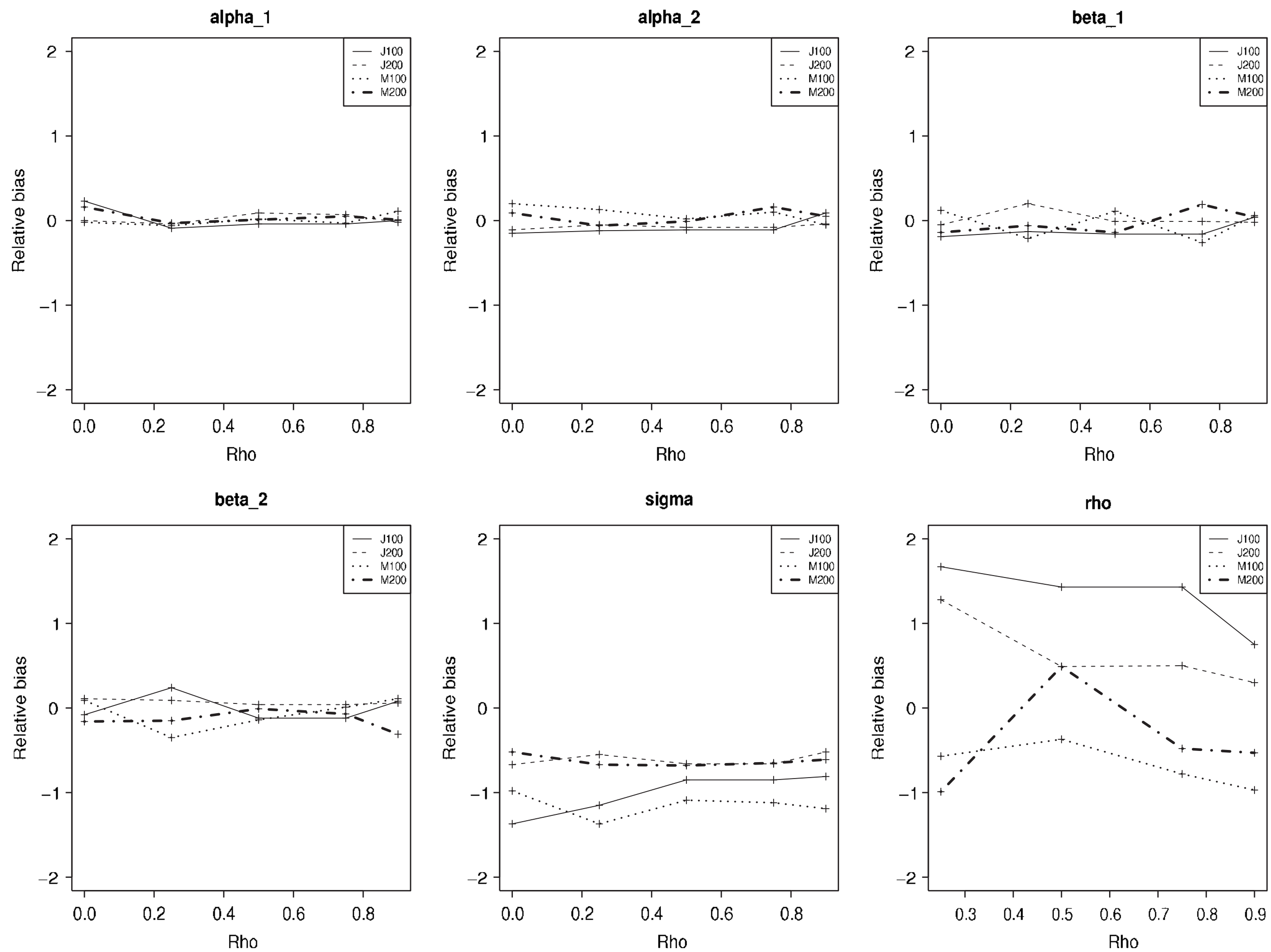

Figure 1. Relative bias = (estimate–parameter)/parameter × 100% of joint and marginal estimates of $\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma$, and ρ, using outcome-specific covariates. Solid and dashed plots correspond to relative biases of joint estimates for $N=100$ and 200, respectively; dotted and dashed-dotted plots correspond to those of marginal estimates for $N=100$ and 200, respectively.

Results displayed in Table I suggest that joint estimates perform better than marginal estimates. Relative bias of the latter are noticeably larger than that of the former, especially for correlation ρ. The marginal method tended to severely underestimate ρ and also yielded the lowest relative efficiency of 49 per cent. This contrasts with uniformly high relative efficiencies (i.e. very close to unity) for joint estimates, indicating that SEs are able to capture true sampling variability of joint estimates. We note that although relative efficiencies of marginal estimates (except for ρ) were also generally relatively close to unity, their empirical SDs were consistently higher than corresponding SDs for joint estimates, unlike in the robit-normal case. This indicates efficiency gains from joint estimation which were absent in the robit-normal case, a possible explanation for which is our use of a non-binary 'high-information' discrete outcome in this case.

5. Application to burn injury data

In this section, we revisit the burn injury data analyzed by Song *et al.* [19]. The data involve $N=981$ patients of different ages suffering from burn injuries. The interest lies in linking two outcomes, namely, severity of burn injury Y_i (continuous, measured by total burn area) and disposition X_i (binary, with 1 for death and 0 for survival), to the common covariate z_i, the patient's age. Following [19], we let $Y_i = \log(\text{burn area}+1) \sim \text{normal}(\mu_{2i}, \sigma^2)$ as our continuous outcome. We assume a latent variable $Y_i^* \sim t_1(\mu_{1i}, 1, v)$ such that $X_i = I\{Y_i^* > 0\}$, and consider the following marginal linear models relating means μ_{1i} and μ_{2i} to patient's age z_i:

$$\mu_{1i} = \alpha_1 + \alpha_2 z_i, \quad \mu_{2i} = \beta_1 + \beta_2 z_i. \tag{9}$$

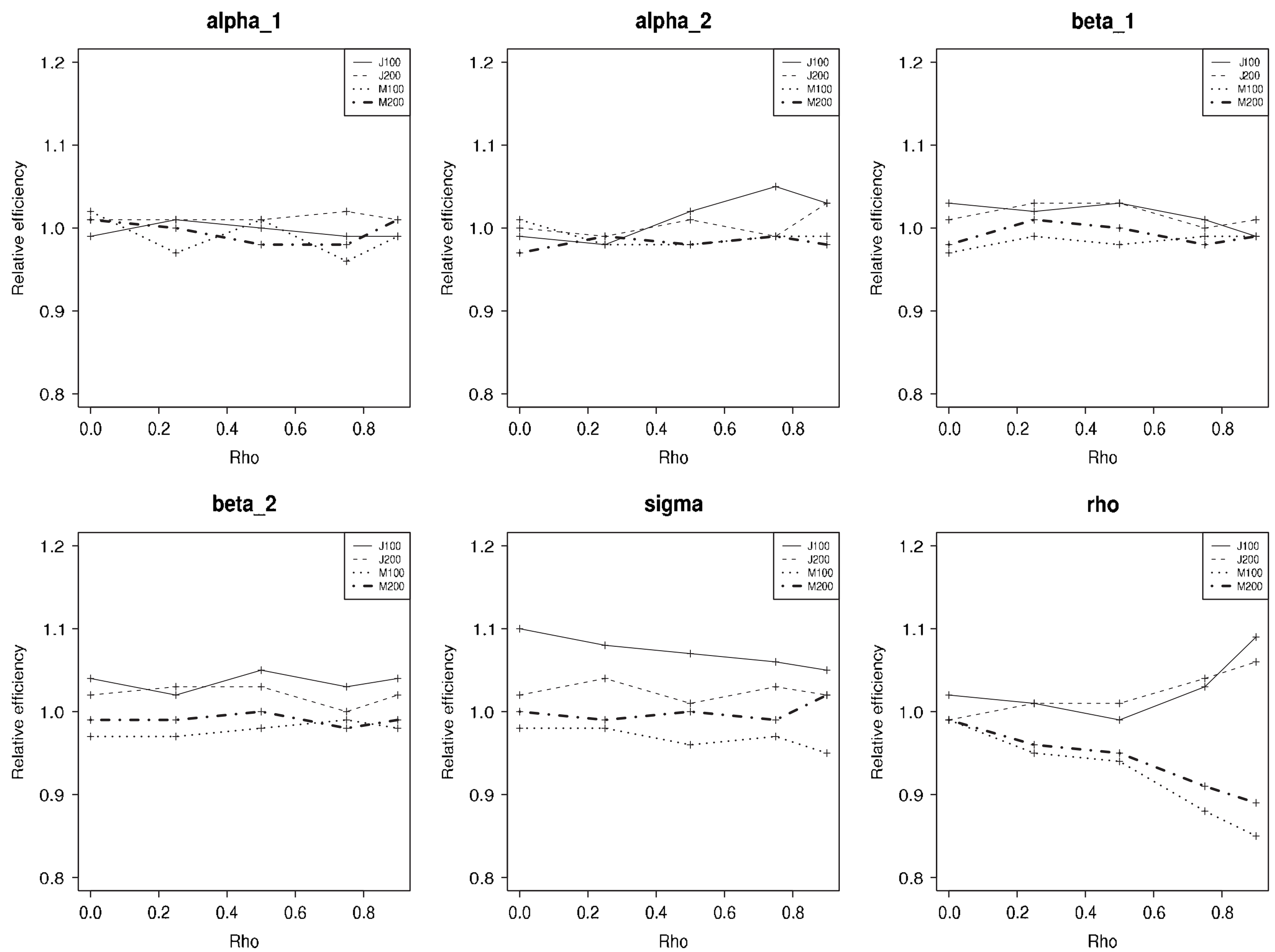

Figure 2. Relative efficiency=mean SE÷standard deviation of joint and marginal estimates of $\alpha_1, \alpha_2, \beta_1, \beta_2, \sigma$, and ρ, using outcome-specific covariates and score function to calculate SEs. Solid and dashed plots correspond to relative efficiencies of joint estimates for $N=100$ and 200, respectively; dotted and dashed-dotted plots correspond to those of marginal estimates for $N=100$ and 200, respectively.

Table I. Relative bias (RB, in per cent), empirical standard deviation (SD), and relative efficiency (RE=mean SE÷SD) of joint and marginal estimates for probit-exponential model with $N=200$ and $R=5000$.

	Marginal				Joint			
Parameter	Est	RB	SD	RE	Est	RB	SD	RE
$\alpha=2$	1.998	−0.082	0.279	1.359	2.001	0.051	0.185	1.001
$\gamma_1=-0.43$	−0.435	0.942	0.169	1.089	−0.429	−0.48	0.125	0.996
$\gamma_2=0.43$	0.432	0.37	0.168	1.109	0.434	0.852	0.132	0.976
$\beta_1=3$	3.048	1.605	0.551	0.979	3.025	0.846	0.414	0.994
$\beta_2=4$	4.001	0.034	1.196	0.983	4.002	0.062	0.869	1.006
$\rho=0.8$	0.544	−31.977	0.098	0.489	0.804	0.459	0.029	0.999

A joint model to glue the marginal models in (9) is specified by a bivariate Gaussian copula with correlation ρ. The choice of a marginal robit regression of X_i on z_i—using a t-latent distribution for Y_i^*—results in inferences that are robust to the presence of outliers. Liu [26] also shows that robit models approximate both logit (with $v \approx 7$) and probit (with large v) regressions, and thus provide a general approach to binary regression modelling.

Both full and marginal estimation methods outlined in Section 4 are implemented for the data. Resulting estimates are reported in Table II. In both cases, degrees of freedom v was estimated using profile likelihood over the range $v \in (2, 8]$ with respective grid lengths 0.5 and 0.1 for joint and marginal approaches; profile plots

Table II. Joint and marginal estimates (Est), standard errors (SE), and z-values for burn injury data.

	Marginal			Joint		
Parameter	Est	SE	z-value	Est	SE	z-value
Death disposition						
α_1	−6.5548	0.586	−11.19	−3.9095	0.3829	−10.22
α_2	0.0849	0.0097	8.75	0.0554	0.0063	8.79
v	2.8			2.5		
Burn severity						
β_1	6.7413	0.635	106.16	6.6721	0.067	99.58
β_2	0.0029	0.0017	1.71	0.004	0.0017	2.35
σ	1.2683	0.0382	33.2	1.2048	0.0269	44.79
Association						
ρ	0.8654	0.0176	49.17	0.8197	0.0261	31.41
τ	0.6659	0.0224	29.73	0.6117	0.029	21.09
ρ_S	0.8546	0.0186	45.95	0.8065	0.0273	29.54

indicated the profile likelihood to be concave over this interval. From Table II, we can see that the joint and marginal estimates are comparable, especially for outcome burn severity. Joint and marginal estimates of the regression coefficient α_2 for age in the robit regression of death disposition on age indicate that the probability of death increases with age. Similarly, estimates of regression coefficient β_2 for age indicate that mean log(burn area+1) slowly increases with age. Note that SEs for both sets of estimates are relatively close; z-values suggest both age coefficients for burn severity and death disposition are significantly positive. Estimates of correlation ρ suggest a high positive association between burn severity and the latent variable underlying death disposition, indicating that burn severity and death disposition are strongly positively associated; estimates of non-parametric rank correlations $\tau = 2\sin^{-1}(\rho)/\pi$ and $\rho_S = 6\sin^{-1}(\rho/2)/\pi$ are also displayed along with their SEs obtained by the delta method.

We also looked at the conditional behaviours of the outcomes. From (6), we get

$$P(X_i = 1|Y_i = y) = 1 - \Phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y-\beta_1-\beta_2 z_i}{\sigma}\right)}{\sqrt{1-\rho^2}}\right); \tag{10}$$

similarly, we get

$$E(Y_i|X_i = x) = E(Y_i) + \begin{cases} \dfrac{\rho\sigma}{1-u_i}\phi\{\Phi^{-1}(u_i)\}, & x = 1 \\ -\dfrac{\rho\sigma}{u_i}\phi\{\Phi^{-1}(u_i)\}, & x = 0 \end{cases}, \tag{11}$$

where $u_i = F_{Y_i^*}(0)$. Figure 3 plots the following: conditional probability of death in (10), as a function of age with log(burn area+1) fixed at its quartiles, estimated (a) marginally and (d) jointly; conditional probability of death in (10), as a function of log(burn area+1) with age fixed at its quartiles, estimated (b) marginally and (e) jointly; and marginal and conditional (on death or survival) regression functions in (11), of log(burn area+1) on age, estimated (c) marginally and (f) jointly.

We see that our estimates, especially those jointly obtained, are generally comparable in magnitude to those reported in [19]. Recall that [19] used a logistic-normal model to analyze the data; as a check, we applied our model with $v = 7$ (at which t approximates logistic, [26]), and obtained estimates very close to those in [19]. Note that our approach generalizes the logistic-normal and probit-normal models, since the t-latent distribution provides excellent approximations to both logistic and probit models for binary data at certain values of degrees of freedom v. Our approach thus provides a means to select among these models.

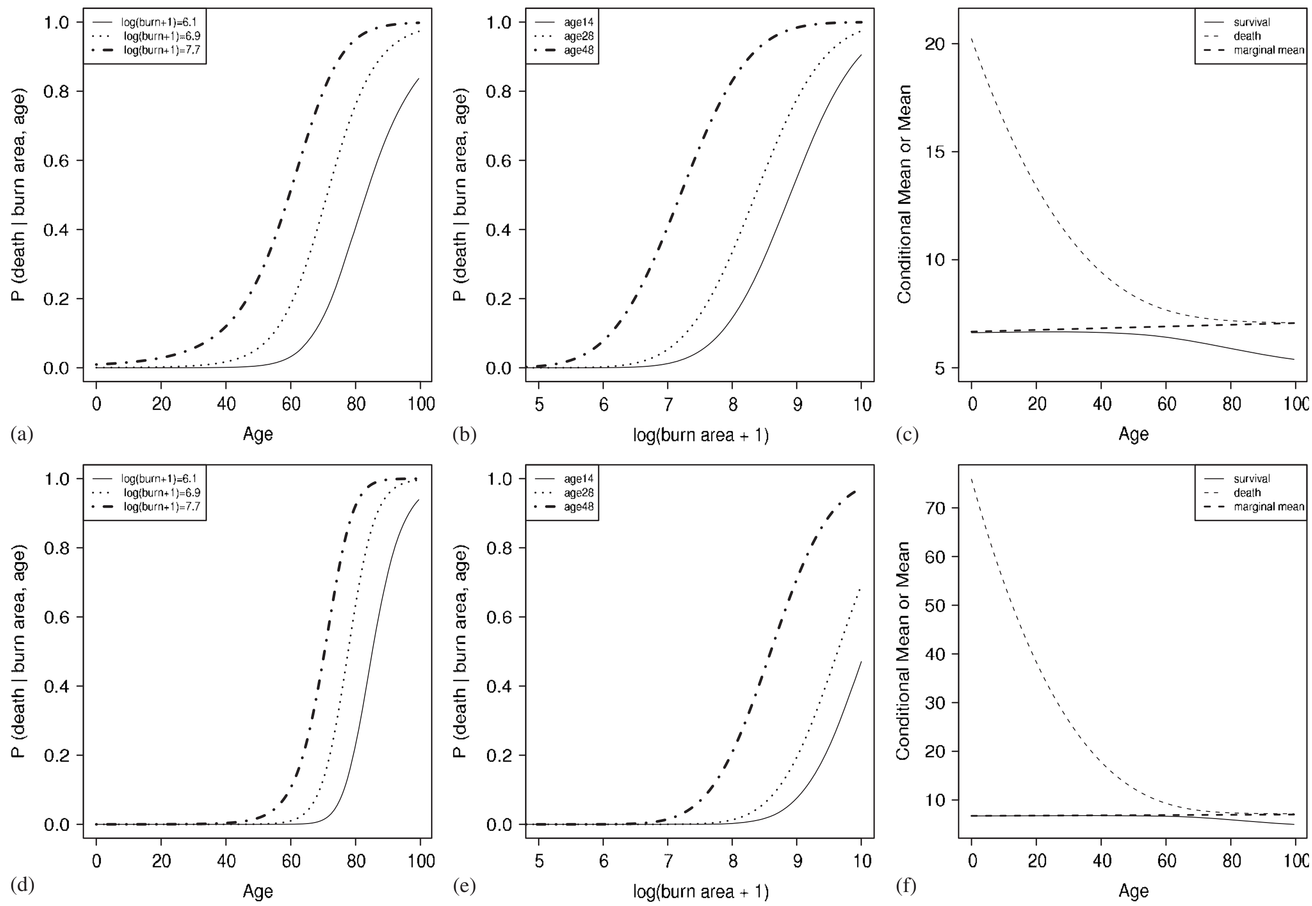

Figure 3. Top panel plots (a)–(c) correspond respectively to joint estimates of conditional probability of death as a function of age at fixed $\log(\text{burn area}+1) = 6.1$, 6.9, 7.7; conditional probability of death as a function of $\log(\text{burn area}+1)$ at fixed age $=14$, 28, and 48 weeks; and regression functions of $\log(\text{burn area}+1)$ (marginal and conditional on death/survival); bottom panel plots (d)–(f) correspond to marginal estimates of the same quantities.

6. Discussion

In this paper, we develop copula-based regression models for mixed outcomes by adopting a latent-variable formulation of the discrete outcomes and using Gaussian copulas to 'glue' mixed-outcome marginal regression models. Two likelihood estimation strategies are proposed, one method uses the full likelihood function to estimate parameters simultaneously, the other applies the IFM method to estimate marginal parameters marginally and shared parameters (e.g. association parameters) jointly.

We study in detail two specific versions of our model: one for binary and normal continuous outcomes, where a t-latent distribution is used to define the marginal distribution of the binary outcome, and another for trichotomous discrete and exponential continuous outcomes, where a normal model is adopted for the latent distribution. The choice of a t-latent distribution corresponds to so-called robit regression, a robust alternative to and extension of logit and probit models. Simulation results indicate both estimation methods are able to recover parameters reasonably well and their standard errors reflect true sampling variability. Both methods performed relatively similarly in the binary-continuous case; however, joint estimation clearly outperformed marginal estimation in the case of trichotomous and continuous outcomes. We attribute this efficiency gain to the non-binary 'high-information' nature of the discrete variable, in which case joint estimation should be preferred to marginal estimation. To illustrate our approach, we apply our model to data on patients with burn injuries. Results obtained from our analysis are generally comparable to previous results by Song *et al*. [19].

We assumed throughout the paper that observations are independent; however, many applications in practice have correlated observations in the sense that the basic sampling unit is a cluster, as is the case in studies on developmental toxicology, where the sampling unit is a dam or a litter, and hence, observations belonging to the same dam must be correlated. Work on extending our methodology to allow for clustering in the data is on-going. We are also exploring generalizations of the model to the multivariate case of more than one discrete and more than one continuous outcome, including the incorporation of random effects. The challenge here lies in defining models that allow for different levels of association among outcomes, as in longitudinal studies, and always guarantee proper joint distributions.

Acknowledgements

This research is supported by Grant 76-1060 from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to three anonymous reviewers for insightful comments leading to an improved article. Part of this work was carried out while A. R. de Leon was visiting the School of Statistics, University of the Philippines.

References

1. de Leon AR, Zhu Y. ANOVA extensions for mixed discrete and continuous data. *Computational Statistics and Data Analysis* 2008; **52**:2218–2227.
2. McCulloch C. Joint modeling of mixed outcome types using latent variables. *Statistical Methods in Medical Research* 2007; **17**:1–21.
3. Teixeira-Pinto A, Normand SLT. Correlated bivariate continuous and binary outcomes: issues and applications. *Statistics in Medicine* 2009; **28**:1753–1773.
4. Olkin I, Tate, RF. Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics* 1961; **32**:448–465. (Correction in **36**: 343–344).
5. Fitzmaurice GM, Laird NM. Regression models for mixed discrete and continuous with potentially missing values. *Biometrics* 1997; **53**:110–122.
6. Fitzmaurice GM, Laird NM. Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American Statistical Association* 1995; **90**:845–852.
7. Cox DR, Wermuth N. Response models for mixed binary and quantitative variables. *Biometrika* 1992; **79**:441–461.
8. Regan MM, Catalano PJ. Combined continuous and discrete outcomes. In *Topics in Modelling of Clustered Data*, Aerts M, Geys H, Molenberghs G, Ryan L (eds). Chapman & Hall: London, 2002; 233–261.
9. Catalano PJ, Ryan LM. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* 1992; **87**:651–658.
10. Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in Medicine* 2001; **20**:3023–3038.
11. de Leon AR, Carrière KC. General mixed-data model: extension of general location and grouped continuous models. *Canadian Journal of Statistics* 2007; **35**:533–548.
12. George EO, Armstrong A, Catalano PJ, Srivastava DK. Regression models for analyzing clustered binary and continuous outcomes under an assumption of exchangeability. *Journal of Statistical Planning and Inference* 2007; **137**:3462–3474.
13. Yang Y, Kang J, Mao K, Zhang J. Regression models for mixed Poisson and continuous longitudinal data. *Statistics in Medicine* 2007; **26**:3782–3800.
14. Munkin M, Trivedi PK. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with applications. *Econometrics Journal* 1999; **1**:29–48.
15. Timbie JW, Normand SLT. A comparison of methods for combining quality and efficiency performance measures: profiling the value of hospital care following acute myocardial infarction. *Statistics in Medicine* 2008; **27**:1351–1370.
16. Daniels MJ, Normand SLT. Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics* 2006; **7**:1–15.
17. Faes C, Aerts M, Molenberghs G, Geys H, Teuns G, Bijnens L. A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine* 2009; **27**:4408–4427.
18. Kolev N, Paiva D. Copula-based regression models: a survey. *Journal of Statistical Planning and Inference* 2009; **139**:3847–3856.
19. Song PXK, Li M, Yuan Y. Joint regression analysis of correlated data using Gaussian copula. *Biometrics* 2009; **65**:60–68.
20. Zimmer DM, Trivedi PK. Using trivariate copulas to model sample selection and treatment effects: application to family health care demand. *Journal of Business and Economic Statistics* 2006; **24**:63–76.
21. Song PXK. *Correlated Data Analysis: Modeling, Analytics, and Applications*. Springer: Berlin, 2007.
22. Heinen A, Rengifo E. Multivariate reduced rank regression in non-Gaussian contexts, using copulas. *Computational Statistics and Data Analysis* 2008; **52**:2931–2944.
23. Heinen A, Rengifo E. Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance* 2007; **14**:564–583.
24. Genest C, Nešlehovà J. A primer on copulas for count data. *ASTIN Bulletin* 2007; **37**:475–515.
25. de Leon AR, Carrière Chough K. Mixed-outcome data. In *Encyclopedia of Biopharmaceutical Statistics* (3rd edn), Chow S-C (ed.). Taylor & Francis: London, 2010.
26. Liu C. Robit regression: a simple robust alternative to logistic and probit regression. In *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives*, Gelman A, Meng XL (eds). Wiley: New York, 2004; 227–238.
27. Joe H, Xu JJ. The estimation method of inference functions for margins for multivariate models. *Technical Report*, Department of Statistics, University of British Columbia, 1996.
28. Pitt M, Chan D, Kohn R. Efficient Bayesian interface for Gaussian copula regression models. *Biometrika* 2006; **93**:537–554.
29. Braeken J, Tuerlinckx F, De Boeck P. Copula functions for residual dependency. *Psychometrika* 2007; **72**:393–411.

30. Drasgow F. Polychoric and polyserial correlations. In *Encyclopedia of Statistical Sciences*, Kotz S, Johnson NL (eds). Wiley: New York, 1986; 68–74.
31. Gueorguieva RV, Agresti A. A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* 2001; **96**:1102–1112.
32. Nikoloulopoulos AK, Karlis D. Multivariate logit copula model with an application to dental data. *Statistics in Medicine* 2008; **27**: 6393–6406.
33. Kotz S, Nadarajah S. *Multivariate t Distributions and their Applications*. Cambridge University Press: New York, 2004.
34. Tan M, Qu Y, Rao JS. Robustness of the latent variable model for correlated binary data. *Biometrics* 1999; **55**:258–263.
35. Song PXK, Zhang P, Qu A. Maximum likelihood inference in robust linear mixed-effects models using multivariate *t* distributions. *Statistica Sinica* 2007; **17**:929–943.

Copula-based regression models for a bivariate discrete and continuous outcome

Online Supplementary Materials

A. R. de LEON & B. WU

Score function for joint estimation.

Given data $\{x_i, y_i, \mathbf{z}_{1i}, \mathbf{z}_{2i}\}$, $i = 1, \cdots, N$, the log-likelihood function is given by (7) as

$$\begin{aligned}\ell(\boldsymbol{\theta}, \nu) &= -N\log\sigma + \sum_{i=1}^{N}\log\phi\left(\frac{y_i - \mu_{2i}}{\sigma}\right) + \sum_{\forall x_i=0}\log\Phi\left(-\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i-\mu_{2i}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) \\ &\quad + \sum_{\forall x_i=1}\log\Phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i-\mu_{2i}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right), \qquad (1)\end{aligned}$$

where $\boldsymbol{\theta}^\top = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \sigma, \rho)$ and $u_i = P(Y_i^* \leq 0) = T_\nu(-\mu_{1i})$, the distribution function evaluated at $-\mu_{1i}$, of the Student's t-distribution t_ν (i.e., $t_1(0,1,\nu)$) with ν degrees of freedom. Writing $\ell(\boldsymbol{\theta}, \nu)$ with ν fixed as $\ell_\nu(\boldsymbol{\theta})$, we get $\ell_\nu(\boldsymbol{\theta}) = \sum_{i=1}^N \ell_{\nu i}(\boldsymbol{\theta}) = \sum_{i=1}^N \ell_{\nu i}$. Elements of $s_\nu(\boldsymbol{\theta}) = \partial\ell_\nu(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ are easily derived as follows:

$$\frac{\partial\ell_{\nu i}}{\partial\alpha_1} = \begin{cases} -\left(\frac{\left(\frac{\nu+1}{\nu}\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu(1-\rho^2)}\Gamma\left(\frac{\nu}{2}\right)}\right)\frac{\int_{-\infty}^{0}(y^*-\mathbf{z}_i^\top\boldsymbol{\alpha})\left(1+\frac{(y^*-\mathbf{z}_i^\top\boldsymbol{\alpha})^2}{\nu}\right)^{-\frac{\nu+3}{2}}\mathrm{d}y^*}{\Phi\left(-\frac{\Phi^{-1}(u_i)-\rho\left(\frac{y_i-\mathbf{z}_i^\top\boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)} & \text{, if } x_i = 0 \\ \times\exp\left(-\frac{\left(\Phi^{-1}(u_i)-\rho\left\{\frac{y_i-\mathbf{z}_i^\top\boldsymbol{\beta}}{\sigma}\right\}\right)^2}{2(1-\rho^2)} + \frac{1}{2}\left(\Phi^{-1}\{u_i\}\right)^2\right) & \\ \left(\frac{\left(\frac{\nu+1}{\nu}\right)\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu(1-\rho^2)}\Gamma\left(\frac{\nu}{2}\right)}\right)\frac{\int_{-\infty}^{0}(y^*-\mathbf{z}_i^\top\boldsymbol{\alpha})\left(1+\frac{(y^*+\mathbf{z}_i^\top\boldsymbol{\alpha})^2}{\nu}\right)^{-\frac{\nu+3}{2}}\mathrm{d}y^*}{\Phi\left(\frac{\Phi^{-1}(u_i)-\rho\left(\frac{y_i-\mathbf{z}_i^\top\boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)} & \text{, if } x_i = 1 \\ \times\exp\left(-\frac{\left(\Phi^{-1}(u_i)-\rho\left\{\frac{y_i-\mathbf{z}_i^\top\boldsymbol{\beta}}{\sigma}\right\}\right)^2}{2(1-\rho^2)} + \frac{1}{2}\left(\Phi^{-1}\{u_i\}\right)^2\right) & \end{cases} \qquad (2)$$

$$
\frac{\partial \ell_{\nu i}}{\partial \beta_1} = \begin{cases} \left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma^2}\right) - \dfrac{\frac{\rho}{\sigma\sqrt{1-\rho^2}}\phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)}{\Phi\left(-\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)} & \text{, if } x_i = 0 \\ \left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma^2}\right) + \dfrac{\frac{\rho}{\sigma\sqrt{1-\rho^2}}\phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)}{\Phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)} & \text{, if } x_i = 1 \end{cases} \tag{3}
$$

$$
\frac{\partial \ell_{\nu i}}{\partial \sigma} = \begin{cases} -\frac{\rho(y_i - \mathbf{z}_i^\top \boldsymbol{\beta})}{\sigma^2\sqrt{1-\rho^2}} \left\{\Phi\left(-\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)\right\}^{-1} & \text{, if } x_i = 0 \\ \quad \times \phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) - \frac{1}{\sigma} + \frac{(y_i - \mathbf{z}_i^\top \boldsymbol{\beta})^2}{\sigma^3} & \\ \frac{\rho(y_i - \mathbf{z}_i^\top \boldsymbol{\beta})}{\sigma^2\sqrt{1-\rho^2}} \left\{\Phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right)\right\}^{-1} & \text{, if } x_i = 1 \\ \quad \times \phi\left(\frac{\Phi^{-1}(u_i) - \rho\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{\sqrt{1-\rho^2}}\right) - \frac{1}{\sigma} + \frac{(y_i - \mathbf{z}_i^\top \boldsymbol{\beta})^2}{\sigma^3} & \end{cases} \tag{4}
$$

$$
\frac{\partial \ell_{\nu i}}{\partial \eta} = \begin{cases} -\frac{1}{2e^\eta}\left\{(e^{2\eta} - 1)\Phi^{-1}(u_i) - (e^{2\eta} + 1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)\right\} & \\ \times \left\{1 - \Phi\left(\frac{(e^{2\eta}+1)\Phi^{-1}(u_i) - (e^{2\eta}-1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{2e^\eta}\right)\right\}^{-1} & \text{, if } x_i = 0 \\ \quad \times \phi\left(\frac{(e^{2\eta}+1)\Phi^{-1}(u_i) - (e^{2\eta}-1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{2e^\eta}\right) & \\ \frac{1}{2e^\eta}\left\{(e^{2\eta} - 1)\Phi^{-1}(u_i) - (e^{2\eta} + 1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)\right\} & \\ \times \left\{\Phi\left(\frac{(e^{2\eta}+1)\Phi^{-1}(u_i) - (e^{2\eta}-1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{2e^\eta}\right)\right\}^{-1} & \text{, if } x_i = 1 \\ \quad \times \phi\left(\frac{(e^{2\eta}+1)\Phi^{-1}(u_i) - (e^{2\eta}-1)\left(\frac{y_i - \mathbf{z}_i^\top \boldsymbol{\beta}}{\sigma}\right)}{2e^\eta}\right) & \end{cases}, \tag{5}
$$

and $\partial \ell_{\nu i}/\partial \alpha_2 = z_i \partial \ell_{\nu i}/\partial \alpha_1$ and $\partial \ell_{\nu i}/\partial \beta_2 = z_i \partial \ell_{\nu i}/\partial \beta_1$.

Score functions and asymptotic variances for marginal estimation.

The respective marginal log-likelihood functions for the binary and continuous data are given by

$$\ell_1(\boldsymbol{\alpha}, \nu) = \sum_{\forall x_i=1} \log T_\nu(-\mu_{1i}) + \sum_{\forall x_i=0} \log T_\nu(\mu_{1i}) \tag{6}$$

$$\ell_2(\boldsymbol{\beta}, \sigma) = -\frac{N}{2}\log(2\pi) - N\log\sigma - \frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\mu}_2)^\top(\mathbf{y}-\boldsymbol{\mu}_2), \tag{7}$$

where $\boldsymbol{\mu}_2^\top = (\mu_{21}, \cdots, \mu_{2N})$, and $\mathbf{y} = (y_1, \cdots, y_N)^\top$. With ν fixed, write marginal log-likelihood functions (6)–(7) as $\ell_1(\boldsymbol{\alpha}, \nu) = \ell_{1\nu}(\boldsymbol{\alpha}) = \sum_{i=1}^N \ell_{1\nu i}$ and $\ell_2(\boldsymbol{\beta}, \sigma) = \sum_{i=1}^N \ell_{2i}$. We have $\mathbf{J}_\nu$ a block-diagonal matrix with symmetric diagonal blocks

$$\mathbf{J}_1 = -\sum_{i=1}^N E\left(\frac{\partial^2 \ell_{1\nu i}}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^\top}\right) \tag{8}$$

$$\mathbf{J}_2 = -\sum_{i=1}^N E\left(\frac{\partial^2 \ell_{2i}}{\partial(\boldsymbol{\beta}^\top, \sigma)^\top \partial(\boldsymbol{\beta}, \sigma)^\top}\right) \tag{9}$$

$$J_3 = -\sum_{i=1}^N E\left(\frac{\partial^2 \ell_{\nu i}}{\partial\rho}\right), \tag{10}$$

and $\mathbf{K}$ a symmetric block matrix with blocks

$$\mathbf{K}_{11} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{1\nu i}}{\partial\boldsymbol{\alpha}}\right)\left(\frac{\partial \ell_{1\nu i}}{\partial\boldsymbol{\alpha}}\right)^\top\right\} \tag{11}$$

$$\mathbf{K}_{12} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{1\nu i}}{\partial\boldsymbol{\alpha}}\right)\left(\frac{\partial^2 \ell_{2i}}{\partial(\boldsymbol{\beta}^\top, \sigma)^\top}\right)^\top\right\} \tag{12}$$

$$\mathbf{K}_{13} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{1\nu i}}{\partial\boldsymbol{\alpha}}\right)\left(\frac{\partial \ell_{\nu i}}{\partial\rho}\right)\right\} \tag{13}$$

$$\mathbf{K}_{22} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{2i}}{\partial(\boldsymbol{\beta}^\top, \sigma)^\top}\right)\left(\frac{\partial \ell_{2i}}{\partial(\boldsymbol{\beta}^\top, \sigma)^\top}\right)^\top\right\} \tag{14}$$

$$\mathbf{K}_{23} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{2i}}{\partial(\boldsymbol{\beta}^\top, \sigma)^\top}\right)\left(\frac{\partial \ell_{\nu i}}{\partial\rho}\right)\right\} \tag{15}$$

$$K_{33} = \sum_{i=1}^N E\left\{\left(\frac{\partial \ell_{\nu i}}{\partial\rho}\right)^2\right\}. \tag{16}$$

The score function from (6) is easily obtained as follows:

$$\frac{\partial \ell_{1\nu i}}{\partial \alpha_1} = \begin{cases} \frac{(\nu+1)\Gamma\left(\frac{\nu+1}{2}\right)}{\nu\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left\{T_\nu(\mathbf{z}_i^\top \boldsymbol{\alpha})\right\}^{-1} \\ \times \int_{-\infty}^{0} (y^* - \mathbf{z}_i^\top \boldsymbol{\alpha}) \left(1 + \frac{(y^* - \mathbf{z}_i^\top \boldsymbol{\alpha})^2}{\nu}\right)^{-\frac{\nu+3}{2}} \mathrm{d}y^* & \text{, if } x_i = 1 \\ -\frac{(\nu+1)\Gamma\left(\frac{\nu+1}{2}\right)}{\nu\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left\{T_\nu(-\mathbf{z}_i^\top \boldsymbol{\alpha})\right\}^{-1} \\ \times \int_{-\infty}^{0} (y^* - \mathbf{z}_i^\top \boldsymbol{\alpha}) \left(1 + \frac{(y^* - \mathbf{z}_i^\top \boldsymbol{\alpha})^2}{\nu}\right)^{-\frac{\nu+3}{2}} \mathrm{d}y^* & \text{, if } x_i = 0 \end{cases} \tag{17}$$

$$\frac{\partial \ell_{1\nu i}}{\partial \alpha_2} = z_i \frac{\partial \ell_{1\nu i}}{\partial \alpha_1}. \tag{18}$$

The score function $\partial \ell_{\nu i}/\partial \eta$ is given in expression (5). The score function from (7) is the usual score function in normal theory linear regression models.