Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 3, 265–284
DOI: 10.1214/11-BJPS165

Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas

Christian Genest[a], Aristidis K. Nikoloulopoulos[b], Louis-Paul Rivest[c] and Mathieu Fortin[d]

[a]*McGill University*
[b]*University of East Anglia*
[c]*Université Laval*
[d]*Centre INRA–AgroParisTech de Nancy*

Abstract. The authors consider copula models for vectors of binary response variables having marginal distributions that depend on covariates through logistic regressions. They show how to test for residual pairwise dependence between responses, given the explanatory variables. The procedure they propose is based on the score statistic derived from the assumed copula structure under the alternative. The authors further argue that conditional dependence can be conveniently modelled with meta-elliptical copulas, which offer a wide range of positive and negative degrees of association. They call on a composite likelihood to estimate the copula parameters and they provide standard error estimates of the same via linearization. They illustrate their results with Canadian data on the presence or absence of various log grades in trees.

1 Introduction

Consider a multivariate binary regression setup in which $d \geq 2$ dependent 0–1 random variables $Y_1, \ldots, Y_d$ are observed together with a vector $\mathbf{x} \in \mathbb{R}^p$ of explanatory variables. A frequent objective of statistical analysis is to estimate the joint distribution of $(Y_1, \ldots, Y_d)$ given $\mathbf{x}$. In forestry, for example, the presence or absence of various log grades in a tree is of considerable interest to maximize timber harvest. As the information about log grades is generally unknown until the tree is felled down and sawn, characteristics of the standing tree such as species or diameter at breast height can be used to estimate the probability of each configuration of 0's and 1's. A multivariate binary regression model is thus useful for forest management purposes.

A natural way to build a model for $(Y_1, \ldots, Y_d)|\mathbf{x}$ is to specify its joint distribution by successive conditionings, for example, $Y_1|\mathbf{x}$, $Y_2|(Y_1, \mathbf{x})$, $Y_3|(Y_1, Y_2, \mathbf{x}), \ldots$ This is the approach adopted, for example, by Bonney (1987). It may be convenient when the response variables are naturally ordered, but the lack of permutation invariance of the resulting model is often problematic (Prentice, 1988).

Key words and phrases. Composite likelihood, dependent binary data, Mantel–Haenszel statistic, meta-elliptical copula, score statistic, test of independence, rectangle probability.

Received December 2010; accepted July 2011.

Copula modelling provides a convenient solution to this problem. In this approach, the marginal distribution functions $F_1(\cdot|\mathbf{x}), \dots, F_d(\cdot|\mathbf{x})$ are fitted separately and dependence between them is induced through a copula, that is, a d-variate distribution function with uniform margins on [0, 1]. In other words, a copula model for $(Y_1, \dots, Y_d)$ consists of assuming that the relation

$$\Pr(Y_1 \leq y_1, \dots, Y_d \leq y_d|\mathbf{x}) = C\{F_1(y_1|\mathbf{x}), \dots, F_d(y_d|\mathbf{x})\} \qquad (1.1)$$

holds for a specific copula C and all values of $y_1, \dots, y_d \in \{0, 1\}$ and $\mathbf{x} \in \mathbb{R}^p$.

Multivariate binary regression data modelling through copulas is not new. The approach was originally proposed by Meester and MacKay (1994), who explored the merits of Archimedean copulas, and Frank's family in particular, as exchangeable dependence structures. Nonexchangeable copulas were first used for the analysis of trivariate data: Molenberghs and Lesaffre (1994) considered the Plackett family, which does not extend beyond the case $d = 3$, while Gauvreau and Pagano (1997) called on the class of Farlie–Gumbel–Morgenstern copulas, which may be defined in any dimension but can only accommodate weak degrees of dependence between the variables.

In arbitrary dimension $d \geq 2$, Gaussian copulas were promoted early on by Song (2000, 2007) as a highly flexible class of dependence structures, but his applications were largely limited to situations involving exchangeable pairs, in which case multivariate Normal integrals conveniently reduce to one dimension (Joe, 1995). More recently, Nikoloulopoulos and Karlis (2008) called on mixtures of max-infinitely divisible copulas in this context. Although the latter class covers a wide range of pairwise dependence (Joe and Hu, 1996), it does not allow for negative associations, which is the case in the above-mentioned forestry application, which motivated this work.

This paper considers the use of meta-elliptical copulas for multivariate binary regression data modelling purposes. These copulas, which stem from multivariate elliptical distributions, were first investigated by Fang, Fang and Kotz (2002) as an extension of the Gaussian dependence structure. They were further studied by Abdous, Genest and Rémillard (2005) and are now commonly used, for example, in actuarial science and finance.

Meta-elliptical copulas share with Gaussian copulas the ability to accommodate any feasible pattern of association in a set of random variables. However, the meta-elliptical class offers greater flexibility than the Gaussian in modelling the nature of dependence between the variables. For example, Student t copulas can account for tail dependence in multivariate continuous data (Nikoloulopoulos, Joe and Li, 2009), whereas Gaussian copulas cannot. It is of interest, therefore, to investigate this class of dependence structures for binary data. As they cannot be expressed in closed form, however, their use in a truly multivariate context poses non-negligible numerical challenges (Nikoloulopoulos and Karlis, 2009).

Background material on copula-based multivariate logistic regression modelling is provided in Section 2. As a preliminary step to model construction, tests of residual pairwise dependence are introduced in Section 3. Given a family of bivariate copula alternatives, the optimal procedure is identified; it reduces in many cases to a generalized Mantel–Haenszel statistic. Meta-elliptical copulas are reviewed in Section 4, along with the interpretation of their parameters when the data are binary. In Section 5, the composite likelihood estimation method for multivariate binary regression is adapted to this context. To avoid the computational burden of the jackknife procedure of Zhao and Joe (2005), an explicit variance estimator is constructed via linearization. A real-life application is then presented in Section 6; it features data on the presence or absence of various log grades in trees. Concluding remarks are made in Section 7, followed by a technical Appendix.

2 Multivariate logistic regression modelling through copulas

Let $(Y_1, \dots, Y_d)$ be a vector of Bernoulli random variables and let $\mathbf{x} \in \mathbb{R}^p$ be a vector of covariates. Suppose that the conditional distribution function of $(Y_1, \dots, Y_d)|\mathbf{x}$ is of the form (1.1) for a specific choice of copula C (independent of $\mathbf{x}$) and marginal distributions $F_1(\cdot|\mathbf{x}), \dots, F_d(\cdot|\mathbf{x})$.

As the response variables are dichotomous, $F_j(\cdot|\mathbf{x})$ is completely specified by $\pi_j(\mathbf{x}) = \Pr(Y_j = 1|\mathbf{x})$ for each $j \in \{1, \dots, d\}$. In particular, the copula C induces a multivariate logistic regression model for $(Y_1, \dots, Y_d)$ if, for all j,

$$\pi_j(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \beta_j)}{1 + \exp(\mathbf{x}^\top \beta_j)}, \tag{2.1}$$

where β_j is a $p \times 1$ vector of parameters.

The role of the copula C is to account for possible dependence between the residuals of the marginal models. As explained, for example, in Genest and Nešlehová (2007), this copula is uniquely defined only on

$$\mathcal{C}(\mathbf{x}) = \mathrm{Ran}\{F_1(\cdot|\mathbf{x})\} \times \cdots \times \mathrm{Ran}\{F_d(\cdot|\mathbf{x})\},$$

where, in general, $\mathrm{Ran}(F)$ denotes the range of F. In the present case,

$$\mathrm{Ran}\{F_j(\cdot|\mathbf{x})\} = \{0, \bar\pi_j(\mathbf{x}), 1\}, \qquad \bar\pi_j(\mathbf{x}) = 1 - \pi_j(\mathbf{x})$$

for all $j \in \{1, \dots, d\}$. Thus, $\Pr(Y_1 = y_1, \dots, Y_d = y_d|\mathbf{x})$ is given at arbitrary $y_1, \dots, y_d \in \{0, 1\}$ by

$$h_C(y_1, \dots, y_d|\mathbf{x}) = \sum \mathrm{sign}(v) C\{F_1(v_1|\mathbf{x}), \dots, F_d(v_d|\mathbf{x})\}, \tag{2.2}$$

where the sum is over all $v = (v_1, \dots, v_d) \in \{y_1, y_1 - 1\} \times \cdots \times \{y_d, y_d - 1\}$ and $\mathrm{sign}(v) \in \{-1, 1\}$ equals 1 if and only if $\#\{j : v_j = y_j - 1\}$ is even. This formula uses the facts that $F_j(y|\mathbf{x}) = 0$ if $y < 0$ and $C(u_1, \dots, u_d) = 0$ if $u_j = 0$ for at

Table 1 *Conditional distribution of* (Y_1, Y_2) *as a function of* $\pi_j = \pi_j(\mathbf{x})$, $j = 1, 2$

	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	$C(\bar{\pi}_1, \bar{\pi}_2)$	$\bar{\pi}_1 - C(\bar{\pi}_1, \bar{\pi}_2)$
$Y_1 = 1$	$\bar{\pi}_2 - C(\bar{\pi}_1, \bar{\pi}_2)$	$1 - \bar{\pi}_1 - \bar{\pi}_2 + C(\bar{\pi}_1, \bar{\pi}_2)$

Table 2 *Conditional distribution of* (Y_1, Y_2, Y_3) *as a function of* $\pi_j = \pi_j(\mathbf{x})$, $j = 1, 2, 3$

	$Y_3 = 0$		$Y_3 = 1$	
	$Y_2 = 0$	$Y_2 = 1$	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	$C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$C(\bar{\pi}_1, 1, \bar{\pi}_3) - C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$C(\bar{\pi}_1, \bar{\pi}_2, 1) - C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$\bar{\pi}_1 - C(\bar{\pi}_1, 1, \bar{\pi}_3) - C(\bar{\pi}_1, \bar{\pi}_2, 1) + C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$
$Y_1 = 1$	$C(1, \bar{\pi}_2, \bar{\pi}_3) - C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$\bar{\pi}_3 - C(\bar{\pi}_1, 1, \bar{\pi}_3) - C(1, \bar{\pi}_2, \bar{\pi}_3) + C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$\bar{\pi}_2 - C(\bar{\pi}_1, \bar{\pi}_2, 1) - C(1, \bar{\pi}_2, \bar{\pi}_3) + C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$	$1 - \bar{\pi}_1 - \bar{\pi}_2 - \bar{\pi}_3 + C(1, \bar{\pi}_2, \bar{\pi}_3) + C(\bar{\pi}_1, 1, \bar{\pi}_3) + C(\bar{\pi}_1, \bar{\pi}_2, 1) - C(\bar{\pi}_1, \bar{\pi}_2, \bar{\pi}_3)$

least one $j \in \{1, \ldots, d\}$. Tables 1 and 2 show what the general formula reduces to when $d = 2$ and 3, respectively.

Although C is not uniquely defined outside $\mathcal{C}(\mathbf{x})$, there is no harm in assuming that it arises from a parametric class of copulas. This approach was taken, for example, by Meester and MacKay (1994), Molenberghs and Lesaffre (1994), Gauvreau and Pagano (1997), Song (2000, 2007), and Nikoloulopoulos and Karlis (2008). Further, note that if the probit transform is preferred to (2.1), a Gaussian copula in (1.1) leads to the model considered, for example, by Joe (1997), Chib and Greenberg (1998), Gueorguieva and Agresti (2001), and Papathomas and O'Hagan (2005).

3 Score tests of independence

An advantage of model (1.1) is that the selection of an appropriate structure C for dependence can be performed independently from the inference on the marginal distributions $F_1(\cdot|\mathbf{x}), \ldots, F_d(\cdot|\mathbf{x})$. The first step of the analysis consists in fitting logistic regressions to each of the response variables $Y_1, \dots, Y_d$. This results in estimates $\hat{\beta}_1, \dots, \hat{\beta}_d$ for the regression parameters $\beta_1, \dots, \beta_d$, respectively. The second step involves selecting a copula family and fitting it to the residuals. Before proceeding, however, tests of independence should be carried out. One such procedure is proposed here, which is geared to detect the presence of pairwise dependence among residuals.

Suppose that it is desired to test for independence in the pair (Y_k, Y_ℓ) for given $k, \ell \in \{1, \dots, d\}$ with $k \neq \ell$. Further assume that the alternative is represented by a family (C_θ) of copulas in which θ_0 corresponds to independence, that is, $C_{\theta_0}(u_k, u_\ell) = u_k u_\ell$ for all u_k, $u_\ell \in [0, 1]$. Denote the observed pairs by $(Y_{1k}, Y_{1\ell}), \dots, (Y_{nk}, Y_{n\ell})$ and to get compact expressions, set $\pi_{ij} = \pi_j(\mathbf{x}_i) = 1 - \bar{\pi}_{ij}$ for $j \in \{k, \ell\}$ and all $i \in \{1, \dots, n\}$.

The log-likelihood for the pair (Y_k, Y_ℓ) is derived using the expressions given in Table 1. Upon differentiation of this expression with respect to θ, the score function for θ is seen to be

$$\begin{aligned} s_\theta(\theta, \beta_k, \beta_\ell) = \sum_{i=1}^{n} \dot{C}_\theta(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) \biggl\{ & \frac{(1 - Y_{ik})(1 - Y_{i\ell})}{C_\theta(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} - \frac{(1 - Y_{ik}) Y_{i\ell}}{\bar{\pi}_{i\ell} - C_\theta(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\ & - \frac{Y_{ik}(1 - Y_{i\ell})}{\bar{\pi}_{ik} - C_\theta(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\ & + \frac{Y_{ik} Y_{i\ell}}{1 - \bar{\pi}_{ik} - \bar{\pi}_{i\ell} + C_\theta(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \biggr\}, \end{aligned} \tag{3.1}$$

where $\dot{C}_\theta(u_k, u_\ell) = \partial C_\theta(u_k, u_\ell)/\partial\theta$ is assumed to exist for arbitrary u_k, $u_\ell \in (0, 1)$. At independence, this reduces to

$$s_\theta(\theta_0, \beta_k, \beta_\ell) = \sum_{i=1}^{n} \dot{C}_{\theta_0}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) \frac{(Y_{ik} - \pi_{ik})(Y_{i\ell} - \pi_{i\ell})}{\pi_{ik}\pi_{i\ell}\bar{\pi}_{ik}\bar{\pi}_{i\ell}}.$$

Given that logistic regressions were fitted to the marginals $j \in \{k, \ell\}$, the score function for β_j equals

$$s_j(\theta_0, \beta_j) = \sum_{i=1}^{n} (Y_{ij} - \pi_{ij}) \mathbf{x}_i.$$

Under the null hypothesis of independence, one gets for $j \in \{k, \ell\}$,

$$\begin{aligned} \operatorname{cov}(s_j, s_\theta) &= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \frac{\dot{C}_{\theta_0}(\bar{\pi}_{i_2k}, \bar{\pi}_{i_2\ell})}{\pi_{i_2k}\pi_{i_2\ell}\bar{\pi}_{i_2k}\bar{\pi}_{i_2\ell}} \\ &\qquad \times \operatorname{cov}\{(Y_{i_2k} - \pi_{i_2\ell})(Y_{i_2\ell} - \pi_{i_2\ell}), (Y_{i_1j} - \pi_{i_1j})\} \mathbf{x}_{i_1} \\ &= 0. \end{aligned}$$

This shows that the Fisher Information Matrix for $(\theta, \beta_k, \beta_\ell)$ is block diagonal at independence. Accordingly, score tests of $H_0 : \theta = \theta_0$ are obtained by dividing the

Table 3 *Functional form of $\dot{C}_{\theta_0}$ for common families of bivariate copulas*

Parametric families of copulas	Functional form of $\dot{C}(u_k, u_\ell)$ at independence
Ali–Mikhail–Haq, Dąbrowska, Farlie–Gumbel–Morgenstern, Frank, Plackett	$u_k u_\ell(1-u_k)(1-u_\ell)$
Archimedean copula $\phi_\theta^{-1}\{\phi_\theta(u_k)+\phi_\theta(u_\ell)\}$	$u_k u_\ell\{\dot{\phi}_{\theta_0}(u_k u_\ell)-\dot{\phi}_{\theta_0}(u_k)-\dot{\phi}_{\theta_0}(u_\ell)\}$
Clayton, Gumbel–Barnett	$u_k u_\ell \ln(u_k)\ln(u_\ell)$
Gaussian	$\Phi'\{\Phi^{-1}(u_k)\}\Phi'\{\Phi^{-1}(u_\ell)\}$

statistic $s_\theta(\theta_0, \hat{\beta}_k, \hat{\beta}_\ell)$ by the square root of an estimate of the Fisher information for θ. The resulting procedure rejects H_0 if

$$z_{\text{obs}} = \sum_{i=1}^{n} \frac{\dot{C}_{\theta_0}(\hat{\bar{\pi}}_{ik}, \hat{\bar{\pi}}_{i\ell})(Y_{ik}-\hat{\pi}_{ik})(Y_{i\ell}-\hat{\pi}_{i\ell})}{\hat{\pi}_{ik}\hat{\pi}_{i\ell}\hat{\bar{\pi}}_{ik}\hat{\bar{\pi}}_{i\ell}} \Bigg/ \sqrt{\sum_{i=1}^{n} \frac{\dot{C}^2_{\theta_0}(\hat{\bar{\pi}}_{ik}, \hat{\bar{\pi}}_{i\ell})}{\hat{\pi}_{ik}\hat{\pi}_{i\ell}\hat{\bar{\pi}}_{ik}\hat{\bar{\pi}}_{i\ell}}}$$

is larger in absolute value than a critical value derived from the standard Normal distribution, denoted $\mathcal{N}(0, 1)$.

As observed by various authors in other contexts, the most common copula models listed in the books of Joe (1997) or Nelsen (2006) can be clustered into broad classes, according to the functional form of $\dot{C}_{\theta_0}$. Table 3 summarizes the calculations presented in Section 5 of Genest, Quessy and Rémillard (2006) and Proposition 4 therein. It is worth noting that when $\dot{C}_{\theta_0}(u_k, u_\ell) \propto u_k u_\ell(1-u_k)(1-u_\ell)$, the score test is a generalized Mantel–Haenszel statistic, viz.

$$z_{\text{obs}} = \sum_{i=1}^{n}(Y_{ik}-\hat{\pi}_{ik})(Y_{i\ell}-\hat{\pi}_{i\ell}) \Bigg/ \sqrt{\sum_{i=1}^{n} \hat{\pi}_{ik}\hat{\pi}_{i\ell}\hat{\bar{\pi}}_{ik}\hat{\bar{\pi}}_{i\ell}}. \tag{3.2}$$

In fact, the latter is exactly equal to a standard Mantel–Haenszel statistic when the explanatory variables in the marginal logistic regressions define mutually exclusive strata. When the intercept is the only explanatory variable in the logistic regressions, z^2_{obs} is Pearson's classical chi-squared statistic for testing independence in the 2 × 2 marginal contingency table.

To check whether the $\mathcal{N}(0, 1)$ is a good approximation for the distribution of the Mantel–Haenszel score test statistic under the hypothesis of independence, samples of sizes 100, 300, 1000, 3000 and 10,000 were generated from two independent Bernoulli distributions with parameters

$$\pi_j(\mathbf{x}) = \frac{\exp(j + jx_1 - jx_2 - jx_3)}{1+\exp(j + jx_1 - jx_2 - jx_3)}, \qquad j \in \{1, 2\}. \tag{3.3}$$

The explanatory variables were assumed to be mutually independent; x_1 and x_2 were drawn randomly (once and for all) from a $\mathcal{N}(0, 1)$, while x_3 was taken to be Bernoulli$(1/2)$.

Table 4 *Empirical level of the Mantel–Haenszel score test*

	Sample size				
Nominal level	100	300	1000	3000	10,000
1%	1.2	1.0	1.1	1.2	1.0
5%	5.6	5.4	5.4	5.1	4.9
10%	11.1	10.5	10.4	9.9	9.8

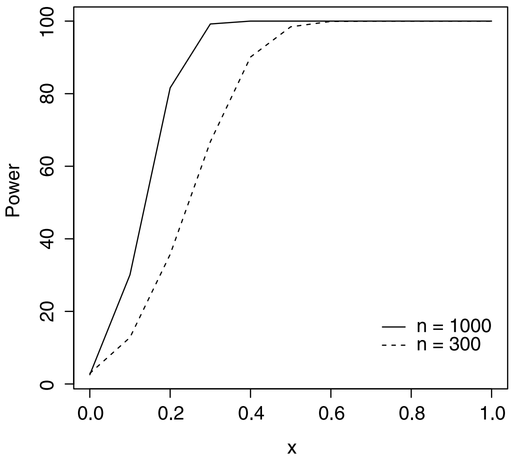

Figure 1 *Power of the Mantel–Haenszel score test for independence, based on random samples of size* $n = 300$ (*dotted*) *and* $n = 1000$ (*solid*) *from bivariate Gaussian copula alternatives with correlation varying in* $(0, 1)$.

Table 4 reports the observed level of the bilateral test for three common nominal levels. As the results are based on 10,000 repetitions, the standard error never exceeds 0.5%. While H_0 tends to be rejected somewhat too often in small samples, this is not a concern when n is sufficiently large.

Figure 1 shows the percentage of rejection of H_0 based on 10,000 random samples of sizes $n = 300$ and 1000 from the bivariate Gaussian copula with latent correlation $\rho = 0$ to 1 in 0.1 increments and with Bernoulli margins with parameters specified in (3.3). The graph shows that the power increases as a function of n and ρ, as expected.

4 Meta-elliptical copulas

Stated succinctly, meta-elliptical copulas are the dependence structures associated with random vectors having elliptically contoured distributions. They were introduced in the statistical literature by Fang, Fang and Kotz (2002) and further studied by Abdous, Genest and Rémillard (2005). They can be viewed as a broad extension of the class of Gaussian copulas which includes, for example, the multivariate Student t copulas.

To be specific, a *continuous* random vector $\mathbf{Z} = (Z_1, \dots, Z_d)$ is said to have an elliptically contoured distribution H if it can be expressed as $\mathbf{Z} = R\Sigma^{1/2}\mathcal{U}$ in terms of a strictly positive random variable R, a $d \times d$ positive semi-definite correlation matrix Σ, and a vector $\mathcal{U}$ which is uniformly distributed on the unit sphere $\mathcal{S}_d = \{(s_1, \dots, s_d) \in \mathbb{R}^d : s_1^2 + \cdots + s_d^2 = 1\}$. It is easily seen that the margins of H are identical, that is, $F_1 = \cdots = F_d = F$. The copula associated with $\mathbf{Z}$ is thus the cumulative distribution function of the vector $(F(Z_1), \dots, F(Z_d))$, that is, it is given by $C(u_1, \dots, u_d) = H(F^{-1}(u_1), \dots, F^{-1}(u_d))$ for all $u_1, \dots, u_d \in [0, 1]$.

Meta-elliptical copulas offer a wide range of dependence properties that are governed both by the distribution of R and the choice of correlation matrix $\Sigma = (\rho_{k\ell})$. For example, Hult and Lindskog (2002) show that for arbitrary $i, j \in \{1, \dots, d\}$, the population value of Kendall's tau between Z_k and Z_ℓ is linked to $\rho_{k\ell}$ through the relation

$$\tau_{k\ell} = 2 \arcsin(\rho_{k\ell})/\pi, \tag{4.1}$$

but the case $\tau_{k\ell} = \rho_{k\ell} = 0$ corresponds to independence only when R has a chi-square distribution, that is, when $\mathbf{Z}$ is Gaussian. However, when a meta-elliptical copula is used in (1.1) to induce dependence in a pair (Y_k, Y_ℓ) of *binary* random variables, relation (4.1) fails. In fact, it can be seen (Nikoloulopoulos and Karlis, 2008) that

$$\tau_{k\ell} = 2\{C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) - \bar{\pi}_{ik}\bar{\pi}_{i\ell}\},$$

where $\pi_{ij} = \pi_j(\mathbf{x}_i) = 1 - \bar{\pi}_{ij}$ for $j \in \{k, \ell\}$ and all $i \in \{1, \dots, n\}$. Thus, Kendall's tau is no longer a function of the copula alone but also depends on the marginal probabilities. See, for example, Denuit and Lambert (2005), Mesfioui and Tajar (2005) or Nešlehová (2007) for further discussion.

An important practical consequence of this observation is that in multivariate logistic regression copula modelling, the range of $\tau_{k\ell}$ is substantially smaller than $[-1, 1]$. In fact, consideration of the Fréchet–Hoeffding bounds leads to the conclusion that $\tau_{k\ell} \in [-1/2, 1/2]$. A simple way to compensate for this shorter span is to work with a scale version of Kendall's tau, viz.

$$\gamma_{k\ell} = \frac{\tau_{k\ell}}{2\xi(\rho, k, \ell)},$$

where $\xi(\rho, k, \ell) = 2C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})^2 + \bar{\pi}_{ik}\bar{\pi}_{i\ell} + C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})(-3 + 2\pi_{ik} + 2\pi_{i\ell})$. This coefficient is known as Goodman's gamma (Goodman and Kruskal, 1954). Equivalently, one could consider the odds ratio

$$\lambda_{k\ell} = \frac{C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\{1 - \bar{\pi}_{ik} - \bar{\pi}_{i\ell} + C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\}}{\{\bar{\pi}_{ik} - C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\}\{\bar{\pi}_{i\ell} - C_{\rho_{k\ell}}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\}},$$

as done, for example, by Trégouët et al. (1999). The latter is a strictly increasing function of γ through the relation $\lambda = (1 + \gamma)/(1 - \gamma)$; see Agresti (1980).

5 Composite likelihood estimation

Consider a d-variate family of copulas indexed by a parameter θ. If C is meta-elliptical with a fixed distribution for the radial part R (i.e., Gaussian or Student t with fixed degrees of freedom ν, say), then θ is a parameter of dimension $d(d-1)/2$ whose components are the upper diagonal entries of the symmetric correlation matrix Σ. The global log-likelihood for θ and the logistic regression parameters $\beta_1, \dots, \beta_d$ is given by

$$\sum_{i=1}^{n} \log\{h_{C_\theta}(Y_{i1}, \dots, Y_{id} | \mathbf{x})\},$$

where h_C is defined in (2.2). Each term in this sum requires d-dimensional integration. When $d > 3$, the numerical problems associated with the evaluation of this log-likelihood call for an alternative estimation strategy.

A two-step composite likelihood approach such as the "CL1" technique of Zhao and Joe (2005) is adequate in this context. First, logistic regression models are fitted to the margins, leading to estimates $\hat{\beta}_1, \dots, \hat{\beta}_d$. Next, estimates of the off-diagonal entries $\rho_{k\ell} = \rho_{\ell k}$ of the correlation matrix Σ are obtained by solving a system of $d(d-1)/2$ estimating equations of the form

$$s_{\rho_{k\ell}}(\rho_{k\ell}, \hat{\beta}_k, \hat{\beta}_\ell) = 0,$$

where $k, \ell \in \{1, \dots, d\}$ with $k < \ell$ and $s_{\rho_{k\ell}}$ is given in (3.1).

5.1 An explicit variance estimator for $\hat{\rho}_{k\ell}$

In their paper, Zhao and Joe (2005) suggest the jackknife as a general strategy for estimating the variance of $\hat{\rho}_{k\ell}$. In large samples, however, it may be more convenient to use an explicit variance estimator. Such an estimator can be derived through linearization (Gong and Samaniego, 1981).

Fix $k, \ell \in \{1, \dots, d\}$ with $k < \ell$ and to simplify notation, set $\rho = \rho_{k\ell}$ throughout this subsection. Adapted to the present context, Theorem 2.2 in Gong and Samaniego (1981) implies (under appropriate regularity conditions) that an estimator for the asymptotic variance of $\hat{\rho}$ is given by

$$v(\hat{\rho}) = \frac{1}{\hat{I}_{\rho\rho}} + \frac{\hat{I}_{\rho\beta}}{\hat{I}_{\rho\rho}^2}(\hat{\Sigma}_{\beta\beta}\hat{I}_{\rho\beta}^\top - 2\hat{\Sigma}_{\rho\beta}).$$

Here, $\hat{I}_{\rho\rho}$ and $\hat{I}_{\rho\beta}$ are plug-in estimates for the (ρ, ρ) entry and the vectors of the (ρ, β_k) and (ρ, β_ℓ) components of the Fisher information matrix, respectively. Similarly, $\hat{\Sigma}_{\beta\beta}$ is a plug-in estimate of the asymptotic covariance matrix of $(\hat{\beta}_k^\top, \hat{\beta}_\ell^\top)^\top$. As for $\hat{\Sigma}_{\rho\beta}$, it should estimate the asymptotic covariance between

$s_\rho(\rho, \beta_k, \beta_\ell)$ and $(\hat{\beta}_k^\top, \hat{\beta}_\ell^\top)^\top$. As shown in the Appendix,

$$
\begin{aligned}
I_{\rho\rho} = \sum_{i=1}^{n} \dot{C}_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})^2 \biggl\{ & \frac{1}{C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} + \frac{1}{\bar{\pi}_{i\ell} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\
& + \frac{1}{\bar{\pi}_{ik} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\
& + \frac{1}{1 - \bar{\pi}_{ik} - \bar{\pi}_{i\ell} + C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \biggr\},
\end{aligned} \tag{5.1}
$$

and for $j \in \{k, \ell\}$,

$$
\begin{aligned}
I_{\rho\beta_j} = -\sum_{i=1}^{n} \dot{C}_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) \biggl\{ & \frac{\dot{C}_{\rho j}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})}{C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} + \frac{\dot{C}_{\rho j}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})}{\bar{\pi}_{i\ell} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\
& - \frac{1 - \dot{C}_{\rho j}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})}{\bar{\pi}_{ik} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \\
& - \frac{1 - \dot{C}_{\rho j}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})}{1 - \bar{\pi}_{ik} - \bar{\pi}_{i\ell} + C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})} \biggr\} \\
\times\, & \bar{\pi}_{ik}(1 - \bar{\pi}_{ik}) \mathbf{x}_i^\top,
\end{aligned} \tag{5.2}
$$

where $\mathbf{x}_i$ is the vector of explanatory variables for the ith observation in the jth logistic regression and $\dot{C}_{\rho j}(u_k, u_\ell) = \partial C_\rho(u_k, u_\ell)/\partial u_j$. It is further shown in the Appendix that $\Sigma_{\rho\beta} = 0$ and

$$
\Sigma_{\beta\beta} = \begin{pmatrix} I(\beta_k)^{-1} & J(\beta_k, \beta_\ell) \\ J(\beta_k, \beta_\ell)^\top & I(\beta_\ell)^{-1} \end{pmatrix},
$$

where

$$
J(\beta_k, \beta_\ell) = I(\beta_k)^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top \{ C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) - \bar{\pi}_{ik}\bar{\pi}_{i\ell} \} \right] I(\beta_\ell)^{-1}
$$

and for $j \in \{k, \ell\}$, $I(\beta_j)$ is the Fisher information matrix for the marginal logistic regression for Y_j, viz.

$$
I(\beta_j) = \sum_{i=1}^{n} \pi_{ij} \bar{\pi}_{ij} \mathbf{x}_i \mathbf{x}_i^\top .
$$

The final estimate $v(\hat{\rho})$ is then obtained from (5.2) upon replacing ρ, β_k and β_ℓ by their estimates in each of the above expressions.

Remark 5.1. Because $\Sigma_{\rho\beta} = 0$, one always has $v(\hat{\rho}) \geq 1/\hat{I}_{\rho\rho}$. The lower bound is the variance estimate that would be used if the marginal parameters β_k and β_ℓ were known or if Y_k and Y_ℓ were independent given $\mathbf{x}$.

Interestingly, one has $v(\hat{\rho}) = 1/\hat{I}_{\rho\rho}$ when the underlying bivariate copula is from Plackett's family. Typically, the latter is not parametrized by $\rho \in [-1, 1]$ but rather by $\theta \in (0, \infty) \setminus \{1\}$, viz.

$$C_\theta(u, v) = \frac{1 + (u + v)(\theta - 1) - \sqrt{\{1 + (u + v)(\theta - 1)\}^2 - 4uv\theta(\theta - 1)}}{2(\theta - 1)}$$

for all $u, v \in (0, 1)$. As defined by Plackett (1965), copulas in this class are precisely those for which

$$C(u, v)\{1 - u - v + C(u, v)\} = \theta\{u - C(u, v)\}\{v - C(u, v)\} \qquad (5.3)$$

for some $\theta > 0$, the limiting case $\theta = 1$ corresponding to independence.

Glonek and McCullach (1995) observed that for Plackett copulas, $I_{\theta\beta_j} = 0$ for $j \in \{1, 2\}$. This result also follows from the above developments. Indeed, upon differentiation with respect to u on both sides of (5.3), one gets

$$\dot{C}(u, v) = \frac{\partial}{\partial u} C(u, v) = \frac{\theta v + (1 - \theta)C(u, v)}{1 - (1 - \theta)\{u + v - 2C(u, v)\}}$$

and hence for all $u, v \in (0, 1)$,

$$\frac{\dot{C}(u, v)}{C(u, v)} + \frac{\dot{C}(u, v)}{v - C(u, v)} - \frac{1 - \dot{C}(u, v)}{u_1 - C(u, v)} - \frac{1 - \dot{C}(u, v)}{1 - u - v + C(u, v)} = 0.$$

Accordingly, (5.2) vanishes and $v(\hat{\rho}) = 1/\hat{I}_{\rho\rho}$ in Plackett's model.

5.2 Performance of the linearized variance estimator

To gauge the small-sample efficiency of the proposed variance estimator, several simulation studies were performed using three- and four-dimensional meta-elliptical copula-based models with marginal logistic regressions. The following experiment is typical of the results that were obtained.

For $n \in \{300, 500, 1000\}$, 10,000 random samples of size n were generated from the trivariate Gaussian copula with correlation matrix

$$\Sigma = \begin{pmatrix} 1.0 & -0.5 & -0.3 \\ -0.5 & 1.0 & 0.3 \\ -0.3 & 0.3 & 1.0 \end{pmatrix}$$

and Bernoulli margins with parameters

$$\pi_j(x) = \frac{\exp(\beta_j + \beta_j x)}{1 + \exp(\beta_j + \beta_j x)}, \qquad j \in \{1, 2, 3\},$$

where $\beta = (-0.5, 0.5, 1)$ and x was drawn from a Uniform distribution on the interval $[-1, 1]$.

Table 5 reports the parameter values, bias, variance, and mean square error (MSE) of the CL1 estimates, along with average values of $v(\hat{\rho})$ and $1/\hat{I}_{\rho\rho}$. From

Table 5 *Observed bias, variance, and MSE of the CL*1 *estimates, along with average values of* $v(\hat{\rho})$ *and* $1/\hat{I}_{\rho\rho}$, *for a trivariate Gaussian copula model with Bernoulli margins*

	n	Bias	Variance	MSE	$v(\hat{\rho})$	$1/\hat{I}_{\rho\rho}$
$\rho_{12} = -0.5$	300	−0.017	0.006	0.006	0.006	0.006
	500	−0.018	0.003	0.004	0.003	0.003
	1000	−0.017	0.002	0.002	0.002	0.002
$\rho_{13} = -0.3$	300	−0.011	0.009	0.009	0.008	0.008
	500	−0.014	0.005	0.005	0.005	0.005
	1000	−0.013	0.003	0.003	0.002	0.002
	300	0.011	0.009	0.009	0.008	0.008
$\rho_{23} = 0.3$	500	0.013	0.005	0.005	0.005	0.005
	1000	0.013	0.003	0.003	0.002	0.002

this simulation and similar ones performed for other choices of parameters and meta-elliptical copulas, it appears that $\hat{\rho}_{k\ell}$ and its variance estimator are nearly unbiased. In addition, it is found that $\mathrm{E}\{v(\hat{\rho}_{k\ell})\} \simeq \mathrm{E}(1/\hat{I}_{\rho_{k\ell}\rho_{k\ell}})$. Accordingly, one can conclude that, in practice, $I_{\rho_{k\ell}\beta} \approx 0$ and, thus, that the composite likelihood estimator $\hat{\rho}_{k\ell}$ is nearly efficient.

6 Illustration

In forest management, industrial needs are often expressed as timber volumes of a specific quality or grade. In order to maximize returns from timber harvest while ensuring a constant supply, information on log grade is required. For practical reasons, information about log grade in a given tree is generally unknown until it is felled down and sawn. Consequently, models for estimating the occurrence of particular log grades in trees are required.

Since 2002, the DAFPP (*Direction de l'aménagement des forêts publiques et privées*) at the *Ministère des Ressources naturelles et de la Faune du Québec* has undertaken a sampling program aimed at providing a representative sample of log grade occurrence in trees for three major hardwood species: paper birch (*Betula papyrifera Marsh.*), yellow birch (*Betula alleghaniensis Britton*), and sugar maple (*Acer saccharum Marsh.*). Data were collected on 1695 trees from eight different sectors between 2002 and 2007.

Before the felling, tree species and diameter at breast height (DBH at 1.3 m high) were recorded. Each tree was also classified according to two binary classifications: one for tree vigor and another for the potential product recovery (Majcen et al., 1990). The tree vigor classification aims at identifying trees that are thought to be at a high risk of dying before the next cutting cycle. It relies mostly on crown and bole defects. If more than one third of the crown showed evidence of dieback or damage, or if some defects could be observed (e.g., canker, decay, fungus or

Table 6 *Descriptive statistics for a forestry data set of size $n = 1695$*

	Paper birch (PB)	Yellow birch (YB)	Sugar maple (SM)
Number of trees	104	772	819
DBH range (cm)	[23.1, 49.6]	[23.1, 95.0]	[23.1, 79.3]
Vigorous trees	83	359	139
Potential sawlog trees	102	717	698
Trees with F1 grade	7	53	34
F2 grade	29	271	203
F3 grade	82	487	512
F4 grade	24	126	93
P grade	45	413	728

Table 7 *Parameter estimates of the marginal logistic regression models. Significant parameters at the 5% level are in boldface*

	P		F1		F2		F3		F4	
Variable	$\hat{\beta}_P$	SE	$\hat{\beta}_{F1}$	SE	$\hat{\beta}_{F2}$	SE	$\hat{\beta}_{F3}$	SE	$\hat{\beta}_{F4}$	SE
Intercept	0.66	0.37	**−7.84**	0.87	**−6.39**	0.48	0.45	0.27	**−1.90**	0.41
PB vs YB	0.03	0.22	0.50	0.44	0.00	0.25	0.15	0.26	0.24	0.26
SM vs YB	**1.81**	0.14	−0.03	0.24	−0.01	0.13	0.09	0.12	**−0.36**	0.16
DHP	**0.02**	0.01	**0.07**	0.01	**0.07**	0.01	**−0.04**	0.00	−0.01	0.01
Vigor	**−0.69**	0.13	**0.83**	0.25	**0.74**	0.13	**0.53**	0.13	0.20	0.16
Sawlog	**−1.08**	0.28	**1.51**	0.73	**2.36**	0.39	**1.69**	0.18	**0.67**	0.30

large open wounds), the tree was assigned to the non-vigorous class. Otherwise, it was deemed to be a vigorous tree. Likewise, depending on the straightness of the bole and external defects, trees were classified as pulpwood or having sawlog potential. See Martel et al. (2001) for a complete list of criteria used in both classifications.

After the felling, the selected trees were brought to the lumber mill and sawn into logs. The latter were then ranked according to the modified Petro classification (Petro and Calvert, 1976), which is commonly used in Québec to grade the different hardwood sawlogs. The classification has one grade of pulpwood logs (P) and four grades of sawlogs, labeled F1, F2, F3 and F4, in decreasing order of quality. The resulting data are summarized in Table 6.

This data set comprises $n = 1695$ observations, five dependent binary variables (F1–F4, P) and four explanatory variables, that is, DBH (cm), Vigor (vigorous = 1), Sawlog (potential sawlog = 1), and Species (YB, PB, SM). The parameter estimates of the marginal logistic regression models appear in Table 7, together with their standard errors. The boldface type identifies parameters that are significant at the 5% level. As can be seen, there are no differences between the PB and YB

Table 8 *Score tests of independence, composite likelihood estimates of the pairwise correlations and their standard errors in four meta-elliptical copula models*

		t_5		t_{10}		t_{15}		Gaussian	
Pair	z_{obs}	$\hat{\rho}$	SE	$\hat{\rho}$	SE	$\hat{\rho}$	SE	$\hat{\rho}$	SE
P–F1	−4.95	−0.30	0.08	−0.31	0.07	−0.32	0.07	−0.33	0.07
P–F2	−4.76	−0.23	0.05	−0.23	0.05	−0.24	0.05	−0.24	0.05
P–F3	−7.54	−0.37	0.05	−0.36	0.04	−0.36	0.04	−0.35	0.04
P–F4	−4.69	−0.21	0.05	−0.22	0.05	−0.23	0.05	−0.24	0.05
F1–F2	2.65	0.16	0.08	0.16	0.07	0.16	0.07	0.16	0.07
F1–F3	−2.60	−0.14	0.08	−0.15	0.07	−0.15	0.07	−0.16	0.07
F1–F4	−0.34	−0.17	0.10	−0.10	0.09	−0.07	0.09	−0.02	0.08
F2–F3	−5.26	−0.22	0.05	−0.22	0.04	−0.22	0.04	−0.22	0.04
F2–F4	0.32	−0.02	0.06	0.00	0.06	0.01	0.05	0.02	0.05
F3–F4	−0.21	0.02	0.05	0.01	0.05	0.00	0.05	−0.01	0.05

Table 9 *Average values of γ for each fitted model*

	t_5		t_{10}		t_{15}		Gaussian	
Pair	$\bar{\bar{\gamma}}$	SE	$\bar{\bar{\gamma}}$	SE	$\bar{\bar{\gamma}}$	SE	$\bar{\bar{\gamma}}$	SE
P–F1	−0.57	0.16	−0.58	0.11	−0.58	0.09	−0.57	0.06
P–F2	−0.38	0.14	−0.37	0.10	−0.37	0.09	−0.36	0.06
P–F3	−0.49	0.05	−0.50	0.04	−0.51	0.04	−0.51	0.05
P–F4	−0.40	0.12	−0.39	0.08	−0.38	0.07	−0.37	0.04
F1–F2	0.43	0.19	0.38	0.15	0.36	0.13	0.31	0.07
F1–F3	−0.31	0.14	−0.30	0.08	−0.30	0.06	−0.29	0.04
F1–F4	0.05	0.13	0.01	0.08	0.00	0.06	−0.03	0.01
F2–F3	−0.32	0.06	−0.32	0.03	−0.32	0.03	−0.32	0.05
F2–F4	0.09	0.15	0.07	0.09	0.06	0.06	0.03	0.01
F3–F4	−0.01	0.08	−0.01	0.04	−0.01	0.03	−0.02	0.00

species. Furthermore, the same predictors are significant for dependent variables F1–F3. It seems harder to predict F4, as none of the explanatory variables is significant at the 1% level. Nevertheless, all variables were kept in the model for prediction purposes.

Values of the Mantel–Haenszel statistic (3.2) are reported in Table 8 for the 10 pairs of variables. It appears from it that variable F4 is conditionally correlated with P only. Also given in Table 8 are composite likelihood estimates of the pairwise correlations, along with their standard errors assuming either a Gaussian copula or a Student t copula t_ν with $\nu = 5$, 10 or 15 degrees of freedom. As a reality check, alternative estimates of the standard errors were obtained via a nonparametric bootstrap procedure; the results (not included) were identical within bootstrap sampling error. In addition, Table 9 provides the value, averaged over all trees, of Goodman's gamma in the different models.

From Tables 8 and 9, it appears that while variables Vigor and Sawlog account for some of the dependence between P and F1–F4, substantial negative residual association remains between these two sets of variables. The multivariate copula models account for most of this residual dependence, but it is difficult to distinguish between them.

To compare the predictive power of the models, the probability of $\{Y_1 = y_1, \ldots, Y_5 = y_5\}$ was estimated for each combination of $y_1, \ldots, y_5 \in \{0, 1\}$ by averaging over all cases, viz.

$$\hat{h}(y_1, \ldots, y_5) = \frac{1}{n} \sum_{i=1}^{n} \widehat{\Pr}(Y_1 = y_1, \ldots, Y_5 = y_5 | \mathbf{x}_i).$$

Each summand was computed with formula (2.2) using the estimated parameters of the marginal logistic regressions and the copula. The results are reported in Table 10 for each model, as well as under independence. The joint probabilities for the Gaussian copula model were computed using the second order approximation of Joe (1995), while the randomized quasi Monte Carlo method of Genz and Bretz (1999, 2002) was used for the Student t copula models.

As a coarse measure of goodness of fit, the observed relative frequencies $\mathcal{O}_{\mathbf{y}}$ were compared to the expected or model estimated frequencies $\mathcal{E}_{\mathbf{y}} = \hat{h}(y_1, \ldots, y_5)$ via the statistic

$$\chi^2 = \sum_{\mathbf{y}} \frac{(\mathcal{O}_{\mathbf{y}} - \mathcal{E}_{\mathbf{y}})^2}{\mathcal{E}_{\mathbf{y}}},$$

where the sum is taken over the 32 combinations of $\mathbf{y} = (y_1, \ldots, y_5)$. Table 10 confirms that the use of copulas results in a dramatic decrease in the value of the χ^2 statistic. For most categories, the observed and predicted frequencies are very close. An exception, however, is the combination $y_1 = \cdots = y_5 = 0$. Under the independence assumption, this cell's contribution to the lack of fit is $\{(0.0047 - 0.0475)^2/0.0475\}/0.0954 \approx 40\%$ of the total. This is reduced, for example, to approximately 26% for the t_5 copula. The difficulty in predicting this particular cell leads to the speculation that, in this study, such trees might have been undersampled because they have no commercial value.

7 Conclusion

This paper provides additional evidence in favour of copula modelling in the context of multivariate logistic regression. It was seen that formulating alternatives to independence in terms of copulas leads to useful tests for detecting pairwise dependence between binary responses. When residual dependence is present, it can then be accounted for with relative ease through a model of the form (1.1). Meta-elliptical copulas are especially flexible in this regard, as they provide a wide

Table 10 *Observed (obs.) frequencies of the dependent variables and average estimated probabilities under the fitted models and independence (ind.), along with values of the goodness-of-fit statistic* χ^2

P	F1	F2	F3	F4	Obs.	Ind.	t_5	t_{10}	t_{15}	Gaussian
0	0	0	0	0	0.0047	0.0475	0.0157	0.0160	0.0161	0.0164
0	0	0	0	1	0.0059	0.0089	0.0056	0.0057	0.0057	0.0058
0	0	0	1	0	0.1209	0.1131	0.1158	0.1152	0.1150	0.1145
0	0	0	1	1	0.0366	0.0247	0.0343	0.0345	0.0345	0.0345
0	0	1	0	0	0.0248	0.0241	0.0194	0.0195	0.0195	0.0196
0	0	1	0	1	0.0083	0.0045	0.0066	0.0065	0.0065	0.0064
0	0	1	1	0	0.0572	0.0492	0.0581	0.0583	0.0584	0.0585
0	0	1	1	1	0.0124	0.0105	0.0164	0.0163	0.0162	0.0162
0	1	0	0	0	0.0024	0.0025	0.0022	0.0022	0.0022	0.0022
0	1	0	0	1	0.0012	0.0005	0.0007	0.0006	0.0006	0.0006
0	1	0	1	0	0.0094	0.0054	0.0063	0.0066	0.0067	0.0068
0	1	0	1	1	0.0018	0.0012	0.0015	0.0015	0.0015	0.0016
0	1	1	0	0	0.0088	0.0030	0.0064	0.0063	0.0062	0.0061
0	1	1	0	1	0.0018	0.0005	0.0015	0.0015	0.0015	0.0015
0	1	1	1	0	0.0035	0.0042	0.0081	0.0080	0.0080	0.0079
0	1	1	1	1	0.0006	0.0008	0.0016	0.0016	0.0017	0.0016
1	0	0	0	0	0.1917	0.1632	0.1754	0.1755	0.1755	0.1756
1	0	0	0	1	0.0171	0.0211	0.0208	0.0209	0.0210	0.0210
1	0	0	1	0	0.2667	0.2555	0.2798	0.2792	0.2790	0.2786
1	0	0	1	1	0.0360	0.0413	0.0340	0.0341	0.0342	0.0343
1	0	1	0	0	0.0690	0.0612	0.0790	0.0789	0.0789	0.0788
1	0	1	0	1	0.0118	0.0089	0.0105	0.0104	0.0103	0.0102
1	0	1	1	0	0.0737	0.0954	0.0656	0.0661	0.0663	0.0668
1	0	1	1	1	0.0077	0.0158	0.0076	0.0075	0.0075	0.0074
1	1	0	0	0	0.0035	0.0063	0.0058	0.0058	0.0058	0.0056
1	1	0	0	1	0.0000	0.0009	0.0007	0.0006	0.0006	0.0007
1	1	0	1	0	0.0053	0.0097	0.0043	0.0044	0.0045	0.0047
1	1	0	1	1	0.0000	0.0016	0.0004	0.0004	0.0004	0.0002
1	1	1	0	0	0.0100	0.0082	0.0111	0.0109	0.0109	0.0107
1	1	1	0	1	0.0012	0.0011	0.0009	0.0009	0.0010	0.0010
1	1	1	1	0	0.0047	0.0083	0.0037	0.0037	0.0037	0.0037
1	1	1	1	1	0.0012	0.0013	0.0002	0.0002	0.0003	0.0003
					χ^2	0.0954	0.0290	0.0285	0.0282	0.0280

range of pairwise dependence when compared with other parametric families such as Archimedean copulas. Although their lack of closed form makes estimation somewhat challenging, this problem can be overcome by resorting to a composite likelihood method. When dealing with large data sets, added convenience is also provided by the variance estimator of the dependence parameters proposed here.

While little difference was observed between the Gaussian and Student t copulas in the forestry data, the meta-elliptical class of copulas represents a vast extension of the multivariate Gaussian copula model. This added flexibility may prove useful in other applications.

Appendix

Fix $k, \ell \in \{1, \dots, d\}$ with $k \neq \ell$ and write $\rho = \rho_{k\ell}$. To determine the value of $I_{\rho\rho} = \operatorname{var}\{s_\rho(\rho, \beta_k, \beta_\ell)\}$, first observe that the 4×1 random vector

$$\mathbf{Z}_i = \big((1 - Y_{ik})(1 - Y_{i\ell}), (1 - Y_{ik})Y_{i\ell}, Y_{ik}(1 - Y_{i\ell}), Y_{ik}Y_{i\ell}\big)^\top$$

has a multinomial distribution with $\mathrm{E}(\mathbf{Z}_i) = \mathbf{p}_i = (p_{i1}, p_{i2}, p_{i3}, p_{i4})^\top$, where

$$p_{i1} = C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}), \qquad p_{i2} = \bar{\pi}_{i\ell} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}),$$
$$p_{i3} = \bar{\pi}_{ik} - C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}), \qquad p_{i4} = 1 - \bar{\pi}_{ik} - \bar{\pi}_{i\ell} + C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}).$$

Letting $\mathbf{q}_i = (1/p_{i1}, -1/p_{i2}, -1/p_{i3}, 1/p_{i4})^\top$, one may rewrite (3.1) as

$$s_\rho(\rho, \beta_k, \beta_\ell) = \sum_{i=1}^{n} \dot{C}_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\mathbf{q}_i^\top \mathbf{Z}_i.$$

Now $\operatorname{var}(\mathbf{Z}_i) = \operatorname{diag}(\mathbf{p}_i) - \mathbf{p}_i \mathbf{p}_i^\top$ and, hence,

$$\operatorname{var}(\mathbf{q}_i^\top \mathbf{Z}_i) = \mathbf{q}_i^\top \operatorname{var}(\mathbf{Z}_i)\mathbf{q}_i = 1/p_{i1} + 1/p_{i2} + 1/p_{i3} + 1/p_{i4}.$$

As $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are mutually independent, formula (5.1) ensues.

To compute $I_{\rho\beta_j}$, first note that if $i \in \{1, \dots, n\}$ and $j \in \{k, \ell\}$, one has

$$\frac{\partial}{\partial \beta_j} C_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell}) = -\dot{C}_{\rho j}(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\pi_{ik}\bar{\pi}_{i\ell}\mathbf{x}_i. \tag{A.1}$$

Accordingly,

$$\begin{aligned} I_{\rho\beta_j} &= -\mathrm{E}\left\{\frac{\partial}{\partial \beta_j} s_\rho(\rho, \beta_k, \beta_\ell)\right\} \\ &= -\sum_{i=1}^{n}\left\{\frac{\partial}{\partial \beta_j}\dot{C}_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\mathrm{E}(\mathbf{q}_i^\top \mathbf{Z}_i) + \dot{C}_\rho(\bar{\pi}_{ik}, \bar{\pi}_{i\ell})\mathrm{E}\left(\frac{\partial}{\partial \beta_j}\mathbf{q}_i^\top \mathbf{Z}_i\right)\right\}. \end{aligned}$$

Formula (5.2) then obtains, upon observing that $\mathrm{E}(\mathbf{q}_i^\top \mathbf{Z}_i) = 0$ and exploiting (A.1) for the evaluation of the second summand.

Now the standard asymptotic theory of maximum likelihood estimation in the logistic regression model (e.g., Cox and Hinkley, 1974, Chap. 9) yields

$$\hat{\beta}_j - \beta_j = I(\beta_j)^{-1}\sum_{i=1}^{n}\{(1 - Y_{ij}) - \bar{\pi}_{ij}\}\mathbf{x}_i + o_p(n^{-1/2}).$$

To compute $\Sigma_{\beta\beta}$, it then suffices to use the fact that

$$\operatorname{cov}\{(1-Y_{ik})-\bar{\pi}_{ik},(1-Y_{i\ell})-\bar{\pi}_{i\ell}\}=\begin{cases}\pi_{ik}\bar{\pi}_{ik} & \text{if } k=\ell,\\ C_\rho(\bar{\pi}_{ik},\bar{\pi}_{i\ell})-\bar{\pi}_{ik}\bar{\pi}_{i\ell} & \text{if } k\neq\ell.\end{cases}$$

Finally, one has $\Sigma_{\rho\beta_k}=0$ because $\mathrm{E}(\mathbf{q}_i^\top\mathbf{Z}_i)=0$ and, hence,

$$\operatorname{cov}\{\mathbf{q}_i^\top\mathbf{Z}_i,(1-Y_{ik})-\bar{\pi}_{ik}\}=\mathrm{E}\{\mathbf{q}_i^\top\mathbf{Z}_i(1-Y_{ik})\}=(p_{i1},p_{i2},0,0)\mathbf{q}_i=0.$$

Acknowledgment

Research funded in part by the Canada Research Chairs Program, the Natural Sciences and Engineering Research Council of Canada, and the Fonds québécois de la recherche sur la nature et les technologies.

References

Abdous, B., Genest, C. and Rémillard, B. (2005). Dependence properties of meta-elliptical distributions. In *Statistical Modelling and Analysis for Complex Data Problems* (P. Duchesne and B. Rémillard, eds.) 1–15. Dordrecht, The Netherlands: Kluwer. MR2189528

Agresti, A. (1980). Generalized odds ratios for ordinal data. *Biometrics* **36**, 59–67. MR0672139

Bonney, G. E. (1987). Logistic regression for dependent binary observations. *Biometrics* **43**, 951–973.

Chib, S. and Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**, 347–361.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837

Denuit, M. and Lambert, P. (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* **93**, 40–57. MR2119763

Fang, H.-B., Fang, K.-T. and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* **82**, 1–16. [Corr.: *Journal of Multivariate Analysis* **94** (2005) 222–223.] MR1918612

Gauvreau, K. and Pagano, M. (1997). The analysis of correlated binary outcomes using multivariate logistic regression. *Biometrical Journal* **39**, 309–325.

Genest, C. and Nešlehová, J. (2007). A primer on copulas for count data. *The Astin Bulletin* **37**, 475–515. MR2422797

Genest, C., Quessy, J.-F. and Rémillard, B. (2006). Local efficiency of a Cramér–von Mises test of independence. *Journal of Multivariate Analysis* **97**, 274–294. MR2208854

Genz, A. and Bretz, F. (1999). Numerical computation of multivariate t-probabilities with application to power calculation of multiple contrasts. *Journal of Statistical Computation and Simulation* **63**, 361–378. MR1718625

Genz, A. and Bretz, F. (2002). Comparison of methods for the computation of multivariate t probabilities. *Journal of Computational and Graphical Statistics* **11**, 950–971. MR1944269

Glonek, G. F. V. and McCullach, P. (1995). Multivariate logistic models. *Journal of the Royal Statistical Society, Ser. B* **57**, 533–546.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics* **9**, 861–869. MR0619289

Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association* **49**, 732–764.

Gueorguieva, R. V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102–1112. MR1947258

Hult, H. and Lindskog, F. (2002). Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability* **34**, 587–608. MR1929599

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association* **90**, 957–964. MR1354012

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall. MR1462613

Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis* **57**, 240–265. MR1391171

Majcen, Z., Richard, Y., Ménard, M. and Grenier, Y. (1990). Choix des tiges à marquer pour le jardinage d'érablières inéquiennes. Guide technique. Mémoire no 96, Direction de la recherche forestière. Ministère de l'Énergie et des Ressources naturelles du Québec, Québec, Canada.

Martel, J., Bergeron, C., Demers, G., Fortin, Y. and Hénaire, F. (2001). Méthode d'échantillonnage pour les suivis des interventions forestières, Exercice 2001–02. Direction de l'assistance technique, Ministère des Ressources naturelles du Québec, Québec, Canada.

Meester, S. G. and MacKay, R. J. (1994). A parametric model for cluster correlated categorical data. *Biometrics* **50**, 954–963.

Mesfioui, M. and Tajar, A. (2005). On the properties of some nonparametric concordance measures in the discrete case. *Journal of Nonparametric Statistics* **17**, 541–554. MR2141361

Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* **89**, 633–644.

Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed. New York: Springer. MR2197664

Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. *Journal of Multivariate Analysis* **98**, 544–567. MR2293014

Nikoloulopoulos, A. K., Joe, H. and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12**, 129–148. MR2515644

Nikoloulopoulos, A. K. and Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Statistics in Medicine* **27**, 6393–6406. MR2655124

Nikoloulopoulos, A. K. and Karlis, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference* **139**, 3878–3890. MR2553774

Papathomas, M. and O'Hagan, A. (2005). Updating beliefs for binary variables. *Journal of Statistical Planning and Inference* **135**, 324–338. MR2200472

Petro, F. J. and Calvert, W. W. (1976). *La classification des billes de bois franc destinées au sciage*. Ottawa, Canada: Service canadien des forêts, Ministère des Pêches et de l'Environnement.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association* **60**, 516–522. MR0183042

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033–1048. MR0980998

Song, P. X.-K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics* **27**, 305–320. MR1777506

Song, P. X-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Application*. New York: Springer. MR2377853

Trégouët, D. A., Ducimetière, P., Bocquet, V., Visvikis, S., Soubrier, F. and Tiret, L. (1999). A parametric copula model for analysis of familial binary data. *American Journal of Human Genetics* **64**, 886–893.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* **33**, 335–356. MR2193979

C. Genest
Department of Mathematics and Statistics
McGill University
805, rue Sherbrooke ouest
Montréal (Québec)
Canada H3A 0B9
E-mail: cgenest@math.mcgill.ca

A. K. Nikoloulopoulos
School of Computing Sciences
University of East Anglia
Norwich NR4 7TJ
United Kingdom
E-mail: A.Nikoloulopoulos@uea.ac.uk

L.-P. Rivest
Département de mathématiques et de statistique
Université Laval
1045, avenue de la Médecine
Québec (Québec)
Canada G1V 0A6
E-mail: Louis-Paul.Rivest@mat.ulaval.ca

M. Fortin
Laboratoire d'étude des ressources forêt-bois
UMR INRA–AgroParisTech 1092
Centre INRA de Nancy
54280 Champenoux
France
E-mail: mfortin@nancy.inra.fr