International Statistical Review
Revue Internationale de Statistique

International Statistical Review (2014), 82, 2, 275–295 doi:10.1111/insr.12035

A Unifying Framework for Marginalised Random-Intercept Models of Correlated Binary Outcomes

Bruce J. Swihart, Brian S. Caffo and Ciprian M. Crainiceanu

Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD 21205, USA
E-mail:bruce.swihart@gmail.com, bcaffo@jhsph.edu, ccrainic@jhsph.edu

Summary

We demonstrate that many current approaches for marginal modelling of correlated binary outcomes produce likelihoods that are equivalent to the copula-based models herein. These general copula models of underlying latent threshold random variables yield likelihood-based models for marginal fixed effects estimation and interpretation in the analysis of correlated binary data with exchangeable correlation structures. Moreover, we propose a nomenclature and set of model relationships that substantially elucidates the complex area of marginalised random-intercept models for binary data. A diverse collection of didactic mathematical and numerical examples are given to illustrate concepts.

Key words: Binary outcomes; copulas; marginal likelihood; multivariate logit; multivariate probit.

1 Introduction

Marginal regression models for correlated binary data have a link function that transforms marginal success probabilities into a linear predictor. We discuss a modelling framework that produces likelihood-based marginal inferences for exchangeably correlated binary outcomes and show that this framework unifies other likelihood-based techniques for clustered binary data of exchangeable correlation structures, particularly inclusive of marginalised random-intercept models (Heagerty, 1999; Heagerty & Zeger, 2000; Wang & Louis, 2003; Caffo & Griswold, 2006; Caffo *et al.*, 2007; Griswold *et al.*, 2013). There are two main novel contributions of this manuscript: (i) establishing the connection of popular existing marginalised multilevel models (MMMs) to copulas where this connection is not currently understood; and (ii) utilising the established connection to organise likelihood-based marginal models of clustered exchangeable binary outcomes into a taxonomy. The taxonomy starts broadly with any copula having fixed marginal distributions and narrows specifically to a model that gives simultaneously (approximate) marginal and conditional log-odds parameters and has a random intercept with the same distribution as the inverse marginal and conditional link. We advance no theory in the area of copulas, which are well known to be useful tools for analysing correlated binary data. Instead, we capitalise on the flexibility of copulas as a means to the end of a unified framework for MMM instances currently in the literature and introduce the Simult-Stable Model. In the process, we broadly survey marginalised random effect models for binary data and connect these popular models to copulas. Using copulas to model joint distributions and to

Table 1. *Visual impairment and race study data from an epidemiological study on visual impairment.*

Group	'Visual impairment' responses (1,1)	(0,1)	(1,0)	(0,0)	Total
B	108	102	94	2039	2343
W	103	135	117	2501	2756

Responses are the visual impairment of the eye (healthy=0, impaired=1) and the demographic covariates are Black (B) and White (W) people. Cell counts represent response-by-group specific counts.

generate samples of discrete random vectors is well established (Sklar, 1973; Schweizer, 1991; Meester & Mackay, 1994; Nelsen, 2006; Nikoloulopoulos & Karlis, 2008c; 2008a; 2008b; Bhat & Sener, 2009; Ruppert, 2010). Although copulas provide a convenient framework for analysis (with fixed marginals), the choice of a particular copula is usually based on convenience without much theoretical justification of such a choice. One intent of this article is to show that this convenience of choice can enable copula models with marginal and conditional inferential utility.

In the following motivating example, our aim is to provide insight for a known equivalence between latent threshold models and copulas. The equivalence will be useful when discussing MMMs that, as we establish, share with the latent threshold model an important and explicit quantity commonly called the transfer function and denoted by Δ in the MMM literature.

1.1 Motivating Example

In a study of eye disease and race, a log odds ratio is desired to estimate the effect of race (Black or White people) on the prevalence of visual impairment in an urban setting (Table 1) (Tielsch *et al.*, 1991; Liang *et al.*, 1992).

Let the pair (Y_{i1}, Y_{i2}) represent the presentation of eye disease of individual i, where $Y_{ij} = 1$ indicates disease in the eye j ($j = 1$, left; $j = 2$, right). The public health study is concerned with visual impairment, which is defined as an eye disease in either or both eyes. The prevalence of visual impairment is represented by the mariginal probability $P(Y_{ij} = 1)$. With D being the standard logistic distribution (i.e. D^{-1} is the logit link), H being the standard normal distribution (i.e. H^{-1} is the probit link), G being a normal distribution with mean 0 and variance τ^2, and F being a normal distribution with mean 0 and variance $1 + \tau^2$, consider three likelihood-based models with covariate $X_{i1} = X_{i2} = 1$ if individual i is a Black person and 0 if a White person:

$$P\left(Y_{ij} = 1\right) = D\left(\beta_0^M + X_{ij}\beta_1^M\right) \tag{1}$$

$$P\left(Y_{ij} = 1|u_i\right) = H\left(\beta_0^C + X_{ij}\beta_1^C + u_i\right) \tag{2}$$

$$\begin{aligned} P\left(Y_{ij} = 1|u_i\right) &= H\left(F^{-1}\left\{D(\beta_0^M + X_{ij}\beta_1^M)\right\} + u_i\right) \\ &= H\left(\Delta_{ij} + u_i\right) \end{aligned} \tag{3}$$

Model (1) is a marginal logistic regression with the desired race effect β_1^M as the marginal log odds ratio and ignores the clustering of the repeated measures (two eyes per individual). Model (2) is a conditional probit regression with the race effect β_1^C, which is of a subject-specific interpretation because of the dependence on random intercept u_i. Whereas β_1^M describes the difference in visual impairment prevalence between a Black population and a White population, β_1^C describes the difference in probability of visual impairment within the same individual as a Black and counterfactually as a White person. No one in the study was observed to change race (from Black to White person or the reverse) so there is no data to support this interpretation. Model (3) is a marginalised random-intercept model, with conditional link function H^{-1} and marginal link function D^{-1}. Model (3) is called 'marginalised' because it is based on a mixed (equivalently, conditional) model that has the random effects integrated out. That is, if the conditional probability $P(Y_{ij} = 1|u_i) = H(\Delta_{ij} + u_i)$ of Model (2) is marginalised by integrating with respect to G, the random-intercept distribution, we arrive at the desired relation of Model (1) between the marginal probability and β_1^M via F and a non-linear quantity in the conditional model, $\Delta_{ij} = F^{-1}\{D(\beta_0^M + X_{ij}\beta_1^M)\}$:

$$
\begin{aligned}
P(Y_{ij} = 1) = \int P(Y_{ij} = 1|u_i)dG = \int H(\Delta_{ij} + u_i)dG =: F(\Delta_{ij}) \\
= F\left(F^{-1}\left\{D(\beta_0^M + X_{ij}\beta_1^M)\right\}\right) = D\left(\beta_0^M + X_{ij}\beta_1^M\right)
\end{aligned}
$$

In Model (3), β_1^M has the desired marginal interpretation but a highly complicated, non-linear conditional interpretation in this example. This paper explores choices of D, F, H and G and how they restrict or modify conditional interpretation alongside the guaranteed marginal interpretation. Typically, Δ is not explicitly written in terms of the desired marginal distribution D and the resultant marginal distribution F because F rarely has a closed or known form, as established in the seminal MMM paper (Heagerty, 1999). Model (3) gives the desired marginal interpretation and accounts for the clustered, repeated measurements and thereby should be more efficient than Model (1).

The non-linear quantity Δ arose in the MMM literature but as we will see is present in any copula with fixed marginals. To demonstrate this, we first turn to latent threshold models and show that by choosing the threshold to be Δ a copula is induced, providing the same inference as the MMM in Model (3). To motivate the latent threshold model, consider the estimation of the joint probabilities $P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2})$, as opposed to $P(Y_{ij} = 1|u_i)$ or $P(Y_{ij} = 1)$. Such estimation must be done under the simplex constraint

$$
\begin{aligned}
P(Y_{i1} = 0, Y_{i2} = 0) + P(Y_{i1} = 1, Y_{i2} = 0) \\
+ P(Y_{i1} = 1, Y_{i2} = 1) + P(Y_{i1} = 0, Y_{i2} = 1) = 1.
\end{aligned}
$$

A flexible method for imposing simplex constraints is to use a latent threshold joint distribution that assigns values to each of the four joint probabilities. Let $(\tilde{Y}_{i1}, \tilde{Y}_{i2})$ be jointly distributed as F_2, a standard bivariate normal distribution with mean zero and marginal variance $1 + \tau^2$, so that it is characterised only by pairwise correlation $\rho = \frac{\tau^2}{1+\tau^2}$. The support of the distribution is the Cartesian cross of the real line, which can be visualised with $\tilde{Y}_{i1}$ taking values on the horizontal axis and $\tilde{Y}_{i2}$ on the vertical axis. The support can be partitioned into four quadrants by intersecting thresholds, the vertical $\tilde{Y}_{i1} = T_{i1}$ and horizontal $\tilde{Y}_{i2} = T_{i2}$ (Figure 1). Let the north-east, north-west, south-west and south-east quadrants with borders T_{i1} and T_{i2} be denoted $Q_i^{(0,0)}$, $Q_i^{(1,0)}$, $Q_i^{(1,1)}$ and $Q_i^{(0,1)}$, respectively. Then assign each joint probability of eye disease presentation to the volume measure under the density of F_2 over the partition of support:

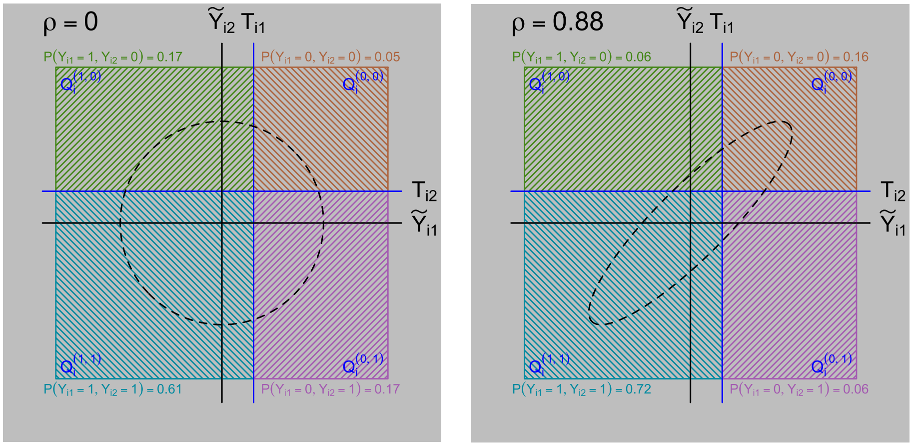

Figure 1. *In each graph is the 'floor' of a standard bivariate normal distribution with identical thresholds* (T_{i1}, T_{i2}). *The ellipse centred at* $(0, 0)$ *contains 95% of the distribution. Increasing the magnitude of* ρ *gathers the distribution along the diagonal of the support and changes the joint probability assignments. In this example, changing* ρ *from* 0 *to* 0.88 *increases* $P(Y_{i1} = 1, Y_{i2} = 1)$ *from* 0.61 *to* 0.72 *and decreases the probability of each discordant case from* 0.17 *to* 0.06.

$$
\begin{aligned}
P\left(Y_{i1} = 0, Y_{i2} = 0\right) &= P_{F_2}\left(\tilde{Y}_{i1} > T_{i1}, \tilde{Y}_{i2} > T_{i2}\right) = P_{F_2}\left(Q_i^{(0,0)}\right) \\
P\left(Y_{i1} = 1, Y_{i2} = 0\right) &= P_{F_2}\left(\tilde{Y}_{i1} \le T_{i1}, \tilde{Y}_{i2} > T_{i2}\right) = P_{F_2}\left(Q_i^{(1,0)}\right) \\
P\left(Y_{i1} = 1, Y_{i2} = 1\right) &= P_{F_2}\left(\tilde{Y}_{i1} \le T_{i1}, \tilde{Y}_{i2} \le T_{i2}\right) = F_2\left(T_{i1}, T_{i2}; \rho\right) \\
P\left(Y_{i1} = 0, Y_{i2} = 1\right) &= P_{F_2}\left(\tilde{Y}_{i1} > T_{i1}, \tilde{Y}_{i2} \le T_{i2}\right) = P_{F_2}\left(Q_i^{(0,1)}\right)
\end{aligned}
$$

The latent variable representation of $P(Y_{i1} = 1, Y_{i2} = 1) = P_{F_2}\left(\tilde{Y}_{i1} \le T_{i1}, \tilde{Y}_{i2} \le T_{i2}\right)$ is of special note in that all the inequalities are '$\le$', which allows the equivalent representation $P_{F_2}\left(Q_i^{(1,1)}\right) = F_2(T_{i1}, T_{i2}; \rho)$, the bivariate cumulative distribution function (CDF). This explicit CDF representation implies that $P(Y_{ij} = 1) = F(T_{ij})$, in which F is the univariate marginal distribution of F_2, a normal distribution with mean 0 and variance $1 + \tau^2$. Let D be the univariate CDF of the standard logistic distribution (mean=0, variance=$\pi^2/3$, scale=1). Consider thresholds $T_{ij} = \Delta_{ij} = F^{-1}\{D(\beta_0 + X_{ij}\beta_1)\}$ depending on the covariates X and regression parameter vector β. Such, T_{ij} imply $P(Y_{ij} = 1) = F(T_{ij}) = F[F^{-1}\{D(\beta_0 + X_{ij}\beta_1)\}] = D(\beta_0 + X_{ij}\beta_1)$, making β_1 the marginal log odds ratio of visual impairment of Black people as compared to White people, as in Model (3), the MMM specification.

Assuming independence across individuals, maximising the likelihood

$$
L(\beta_0, \beta_1, \rho | X, Y) = \Pi_i L_i(\beta_0, \beta_1, \rho | X_i, Y_i) = \Pi_i P_{F_2}\left(Q_i^{(y_{i1}, y_{i2})}\right),
$$

adjusts the joint probability $P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2})$ value assignments by moving the thresholds (estimating β_0 and β_1) and gathering/flattening the volume under the density of F_2 about the 45° ($\tilde{Y}_{i2} = \tilde{Y}_{i1}$) or −45° ($\tilde{Y}_{i2} = -\tilde{Y}_{i1}$) line (estimating positive or negative ρ, respectively). As seen in Figure 1, holding the thresholds constant with $\beta_0 = \beta_1 = 1$, as ρ moves from 0 to 0.88, approaching 1, the greater the probability of concordant cases of eye disease in both eyes $(1, 1)$ and no eye disease in either eye $(0, 0)$. The probability of the (1,1) case increases from 0.61 to 0.72, and the probability of the (0,0) case increases from 0.05 to 0.16,

Table 2. *Summary of example analyses, where $L(0,1)$ is the standard logistic cumulative distribution function (CDF), $B(0,\tau^2)$ is a bridge CDF, $\gamma_F = (1.2^{1.89} + \gamma_G^{1.89})^{1/1.89}$ and $S(a,c)$ is shorthand for a stable distribution parameterized $S(a,0,c,0;0)$.*

Example	D	F	H	G	β_0^M	β_1^M	ρ	τ^2 (γ_G)	ϕ
2.3	L(0,1)	N(0,1)	-	-	-2.43	0.079	0.74	-	-
2.4	L(0,1)	t(0,50)	-	-	-2.43	0.078	0.74	-	-
2.4	L(0,1)	t(0,7.3)	-	-	-2.44	0.072	0.71	-	-
2.6	L(0,1)	t(0,7.3)	-	-	-2.41	0.071	0.71	-	-
2.7	N(0,1)	N(0,1)	-	-	-1.40	0.039	0.74	-	-
3.2	L(0,1)	-	L(0,1)	N(0,τ^2)	-2.43	0.082	0.74	9.25	-
3.3	L(0,1)	N(0,1+τ^2)	N(0,1)	N(0,τ^2)	-2.43	0.079	0.74	2.88	-
3.5	L(0,1)	L(0,ϕ^{-1})	L(0,1)	B(0,τ^2)	-2.42	0.061	-	7.12	0.56
3.6	L(0,1)	L(0,ϕ^{-1})	B(0,τ^2)	L(0,1)	-2.44	0.080	-	1.52	0.82
3.11	N(0,1)	N(0,1+τ^2)	N(0,1)	N(0,τ^2)	-1.40	0.039	-	2.88	0.51
3.12	S(1.89,1.2)	S(1.89,γ_F)	S(1.89,1.2)	S(1.89,γ_G)	-2.44	0.067	-	(2.03)	0.50

reflecting the property of the density gathering around the 45° line. The value assignments to $P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2})$ are based on the measures of the resulting partition of the support, satisfying the simplex constraint.

A crucial insight is that the construct of thresholds $T_{ij} = \Delta_{ij} = F^{-1}\{D(\beta_0 + X_{ij}\beta_1)\}$ admits a copula formulation in which F-copula $C_F(a_1, a_2) = F_2(F^{-1}(a_1), F^{-1}(a_2))$ is evaluated at $(D(\beta_0 + X_{i1}\beta_1), D(\beta_0 + X_{i2}\beta_1))$. That is, by specifying the threshold of the latent threshold model as the transfer function from MMMs, the known equivalence between latent threshold models and copulas implicate that marginalising conditional random-intercept models induces a copula, a connection that is the focus of this paper. Note that F does not require an embedded conditional model specification. This insight is generalisable and will form the basis of our methods and ideas in the general case and in the case admitting embedded random effect models. To be thorough, we present the general case but note the case admitting embedded conditional models is of particular interest for analyses that wish to bridge conditional and marginal models for binary data, as in Wang & Louis (2003).

The paper is organised as follows: Modelling Framework, Modelling Framework Admitting Univariate Latent Random Intercepts, Modelling Framework Admitting Multivariate Random Effects, Model Fitting and Discussion. Select definitions, propositions and proofs are elaborated upon in the Appendix. The Baltimore Eye Survey data used in the motivating example is analysed in several illustrative examples, with results summarised in Table 2.

2 Modelling Framework

Herein, multivariate indicates a joint distribution for the clustered multiple instances of the same outcome (measured at different times or locations). Designate the vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ_i})$ to be the J_i correlated binary outcomes as measured on cluster i with corresponding $J_i \times K$ matrix $\mathbf{X}_i$ comprising of the stacked row vectors $\mathbf{X}_{ij} = (X_{ij1}, \ldots, X_{ijK})$ for K covariates. Let η_{ij} be the linear predictor $\mathbf{X}_{ij}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^\top$ is a column vector of regression coefficients. Allow latent variable $\tilde{\mathbf{Y}}_i = (\tilde{Y}_{i1}, \ldots, \tilde{Y}_{iJ_i}) \sim F_{J_i}(\tilde{y}_{i1}, \ldots, \tilde{y}_{iJ_i}; \rho)$, where F_{J_i} is a J_i-multivariate continuous CDF with pairwise correlation parameter $\rho = corr(\tilde{Y}_{ij}, \tilde{Y}_{ij'})$ and identical univariate marginal CDFs F such that $P(\tilde{Y}_{ij} \le T_{ij}) = F(T_{ij})$ (we suppress using the subscript '1' for the univariate case to reduce clutter). We refer to F_{J_i} as the facilitating distribution. Allow C_F to represent the F-copula distribution function as

$$C_F(a_1, \ldots, a_{J_i}) = F_{J_i}\left\{F^{-1}(a_1), \ldots, F^{-1}(a_{J_i}); \rho\right\}.$$

Let C_{FD} be C_F evaluated at $\{D(\eta_{i1}), \ldots, D(\eta_{iJ_i})\}$:

$$C_{FD}(\eta_{i1}, \ldots, \eta_{iJ_i}) = C_F\{D(\eta_{i1}), \ldots, D(\eta_{iJ_i})\} = F_{J_i}\left[F^{-1}\{D(\eta_{i1})\}, \ldots, F^{-1}\{D(\eta_{iJ_i})\}; \rho\right],$$

and define threshold vector $\mathbf{T}_i = (T_{i1}, \ldots, T_{iJ_i}) = \left[F^{-1}\{D(\eta_{i1})\}, \ldots, F^{-1}\{D(\eta_{iJ_i})\}\right]$, where D, like F, is a univariate continuous CDF with real line support and symmetric about 0. Consider the threshold model given by

$$Y_{ij} = \mathbb{1}_{\{A_{ij}(\eta_{ij})\}}(\tilde{Y}_{ij}) := \begin{cases} 1 \text{ if } \tilde{Y}_{ij} \in A_{ij}(\eta_{ij}) \\ 0 \text{ otherwise} \end{cases}$$

where $A_{ij}(\eta_{ij}) = \{c \in \mathbb{R} : c \leq T_{ij}|\beta, X_i\}$. Define $Q_{ij}(\eta_{ij}, Y_{ij})$ as $A_{ij}(\eta_{ij})$ if $Y_{ij} = 1$ and as the complement set $\bar{A}_{ij}(\eta_{ij}) = \{c \in \mathbb{R} : c > T_{ij}|\beta, X_i\}$ if $Y_{ij} = 0$. Allow $Q_i = \otimes_j Q_{ij}(\eta_{ij}, Y_{ij})$, the Cartesian cross of all the sets for cluster i. The 2^{J_i} combinations of 1's and 0's in the random vector $\mathbf{Y}_i$ correspond to 2^{J_i} possible $Q_i = \otimes_j Q_{ij}(\eta_{ij}, Y_{ij})$, which collectively form the partition of the support for the latent variable $\tilde{\mathbf{Y}}_i$. The actualized $Q_i = \otimes_j Q_{ij}(\eta_{ij}, y_{ij})$ is a subset of the aforementioned partition. The measure of the actualized Q_i according to F_{J_i} corresponds to the probability of $\mathbf{Y}_i = \mathbf{y}_i$, the particular combination of 1s and 0s for cluster i (refer to Motivating Example to build intuition). The likelihood for I independent clusters, each with J_i measurements is

$$L(\beta, \rho|X, Y) = \Pi_i L_i(\beta, \rho|X_i, Y_i) = \Pi_i P_{F_{J_i}}(Q_i) = \Pi_i \int_{Q_i} dF_{J_i}. \tag{4}$$

Equivalently in terms of the joint CDF created with the F-copula, C_{FD}:

$$L(\beta, \rho|X, Y) = \Pi_i L_i(\beta, \rho|X_i, Y_i) = \Pi_i P_{C_{FD}}(Q_i') = \Pi_i \int_{Q_i'} dC_{FD}, \tag{5}$$

for analogous quantities $\tilde{\mathbf{Y}}_i' = (\tilde{Y}_{i1}', \ldots, \tilde{Y}_{iJ_i}') \sim C_{FD}(\tilde{y}_{i1}', \ldots, \tilde{y}_{iJ_i}')$, $\mathbf{T}_i' = (T_{i1}', \ldots, T_{iJ_i}') = (\eta_{i1}, \ldots, \eta_{iJ_i})$, $Y_{ij} = \mathbb{1}_{\{A_{ij}'(\eta_{ij})\}}(\tilde{Y}_{ij}')$, $A_{ij}'(\eta_{ij}) = \{c \in \mathbb{R} : c \leq T_{ij}'|\beta, X_i\}$, $Q_{ij}'(\eta_{ij}, Y_{ij})$ as $A_{ij}'(\eta_{ij})$ if $Y_{ij} = 1$ (as $\bar{A}_{ij}'(\eta_{ij})$ if $Y_{ij} = 0$), and $Q_i' = \otimes_j Q_{ij}'(\eta_{ij}, Y_{ij})$.

Definition 2.1 (Copulas over partitions model likelihood)**.** The likelihood contribution for cluster i having a general outcome $P(Y_{i1} = y_{i1}, \ldots, Y_{iJ_i} = y_{iJ_i})$ is equal to the probability measure $P_{C_{FD}}(Q_i') = \int_{Q_i'} dC_{FD}$. The quantity $C_{FD}(\eta_{i1}, \ldots, \eta_{iJ_i})$ is equal to the likelihood contribution for the unity outcome, $P(Y_{i1} = 1, \ldots, Y_{iJ_i} = 1)$.

The likelihood representation in (5) assigns values to joint probabilities $P(\mathbf{Y}_i)$ via the measure of a latent variable according to its distribution C_{FD} over sets defined by the thresholds of the untransformed linear predictor. Different cluster sizes are handled easily as the facilitating distribution is J_i-specific, and missing outcomes can be handled by setting the corresponding T_{ij} to ∞.

The aforementioned outlined protocol is called copulas over partitions (COP) modelling for binary outcomes. COP models serve as an alternative conceptualization to finite differencing in the copula literature (Nikoloulopoulos & Karlis, 2009). This naming draws from the construction of the following likelihood: assigning joint probabilities for all outcome profiles $\mathbf{Y}_i$ with a copula-based measure over subsets of the partition of the support. Maximising the likelihood for β and ρ allows the optimal assignment of values for the joint probabilities as per the data. Adjusting β moves the borders of the subsets in the partition of the support, and adjusting ρ

fluctuates the distribution itself over the subsets in the partition of the support to be measured (as shown in Figure 1).

2.1 Unrestricted Copulas Over Partition Models

Proposition 2.2 (Unrestricted COP model marginal probability)**.** In a COP model, the marginal probability $P(Y_{ij} = 1) = D(\eta_{ij})$.

By Definition 2.1 and Proposition 2.2 (elaborations in Appendix), the marginal estimation of β is likelihood-based, and the marginal interpretation of β would be determined by D because $D^{-1}\{P(Y_{ij} = 1)\} = \mathbf{X}_{ij}\beta$. That is, COP models produce marginal models for binary data with an arbitrarily specified marginal link function. The benefit of this model is that the facilitating distribution can be chosen to be a convenient multivariate distribution, whereas the marginal link distribution can be chosen for a desired interpretation. Thus, often D^{-1} is the logit link (hence D is the logistic distribution), whereas the two most obvious facilitating distributions are the multivariate normal and t distributions; an example of each is given in Examples 2.3 and 2.4.

Example 2.3 (Unrestricted COP: normal facilitating distribution, logit desired link)**.** Contrasted with the motivating example, there is no conditional model. Minimally, what needs to be specified is the facilitating distribution and the desired marginal link, leading to

$$
\begin{aligned}
F_2 &= MVN\left(\begin{pmatrix}0\\0\end{pmatrix}, \Sigma_F = \sigma_F^2\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)\\
F &= N\left(0, \sigma_F^2\right)\\
D &= logistic\left(0, \pi^2/3, 1\right)
\end{aligned}
$$

maximising $L(\beta, \rho|X, Y) = \Pi_i \int_{Q_i} dF_2$ yields $\beta_0 = -2.43$, $\beta_1 = 0.079$ and $\rho = 0.742$. The interpretation is that of a log odds model, where $\exp(\beta_0) = 0.088$ odds of White people having visual impairment and Black people having $\exp(\beta_1) = 1.082$ times the odds of visual impairment than White people.

Example 2.4 (Unrestricted COP: Student's t-test facilitating distribution, logit desired link)**.** O'Brien & Dunson (2004) selected F_{J_i} to be a multivariate t-distribution with degrees of freedom (ν) that was constructed from a standardised multivariate normal distribution so that the covariance matrix has ones on the diagonal and ρ occupying every off-diagonal. Thus, marginal F is the univariate t-distribution, and D is set to a standard logistic CDF. That is, $D = logistic(0, \pi^2/3, 1)$, $F = t(0, \nu)$, $F_2 = MVT(\mathbf{0}, \Sigma_F = R, \nu)$ and exchangeable correlation matrix R. Maximising $L(\beta, \rho|X, Y) = \Pi_i \int_{Q_i} dF_2$ yields $\beta_0 = -2.44$, $\beta_1 = 0.072$ and $\rho = 0.705$ for $\nu = 7.3$. The interpretation is that of a log odds model, where $\exp(\beta_0) = 0.087$ odds of White people having visual impairment and Black people having $\exp(\beta_1) = 1.075$ times the odds of visual impairment than White people. Maximising $L(\beta, \rho|X, Y) = \Pi_i \int_{Q_i} dF_2$ for $\nu = 50$ yields $\beta_0 = -2.43$, $\beta_1 = 0.078$ and $\rho = 0.737$. The selection of $\nu = 7.3$ is in anticipation of Example 2.6, whereas the selection of $\nu = 50$ is simply a sizeable number greater than 7.3 to demonstrate the estimates of the models approaching the MLEs of Example 2.3. That is, as the degrees of freedom increase the facilitating t-distribution approaches normality (see Table 2).

2.2 Restricted Copulas Over Partition Models

The unrestricted COP model is an extremely flexible marginal modelling approach for binary data. The key drawback of the approach is the non-linear relationship between F^{-1} and D. The non-linear relationship brings complexity to the likelihood computation and interpretation of the latent variable threshold model. A more computationally favourable setting has a linear relationship between F^{-1} and D, which would ease the computation of the likelihood and simplify the threshold model. Thus, we consider the subclass of models where symmetric-about-0 F and D are in the same scalar family.

Definition 2.5 (Restricted COP)**.** Cumulative distribution functions F and D of an unrestricted COP are restricted to be members of a class of scalar family distributions $\{D(c\,\cdot) : c \in \mathbb{R}_+\}$, with $F^{-1}\{D(a)\} = a\frac{S_F}{S_D} = a\phi^{-1}$. S_F and S_D are the scales of F and D, respectively.

Of course, in a restricted COP model, D remains the marginal link distribution. The benefit, however, is that the COP model simplifies as $C_{FD}(\eta_{i1}, \ldots, \eta_{iJ_i}) = F_{J_i}(\eta_{i1}\phi^{-1}, \ldots, \eta_{iJ_i}\phi^{-1})$. That is, a restricted COP model uses the facilitating distribution as a joint threshold model for correlated binary data, with the thresholds as a rescaling of the linear predictor. Thus, constructing restricted COP models that have marginal probit and t-quantile link functions for D and multivariate normal and t distributions for F_{J_i}, respectively, is convenient.

Example 2.6 (Restricted COP (approximation): Albert and Chib approximation)**.** Special case of Example 2.4, with $\nu = 7.3$ [as fully explained in Albert & Chib (1993) and utilised in O'Brien & Dunson (2004)]. This choice of ν invokes a known approximation of $F^{-1}D(a) = a\frac{1}{\sigma}$, where $\sigma^2 = \frac{\pi^2(\nu-2)}{3\nu} = 2.39$. The approximation allows the likelihood to be restated in terms of a scaled facilitating distribution with untransformed linear predictors as the thresholds.

$$\begin{aligned}
P(Y_{i1} = 1, Y_{i2} = 1) &= C_{FD}(\eta_{i1}, \eta_{i2}) \\
&= F_2\left[F^{-1}\{D(\eta_{i1})\}, F^{-1}\{D(\eta_{i2})\}; \Sigma_F = R\right] \\
&= F_2\left(\eta_{i1}\frac{1}{\sigma}, \eta_{i2}\frac{1}{\sigma}; \Sigma_F = R\right) \\
&= F_2\left(\eta_{i1}, \eta_{i2}; \Sigma_F = \sigma^2 R\right).
\end{aligned}$$

Example 2.7 (Restricted COP: Multivariate probit)**.** Let $D = N(0, \sigma_D^2)$, $F = N(0, \sigma_F^2)$, and $F_2 = MVN(\mathbf{0}, \Sigma_F = \sigma_F^2 R)$. This implies the probability assignment

$$\begin{aligned}
P(Y_{i1} = 1, Y_{i2} = 1) &= C_{FD}(\eta_{i1}, \eta_{i2}) \\
&= F_2\left[F^{-1}\{D(\eta_{i1}; \sigma_D); \sigma_F\}, F^{-1}\{D(\eta_{i2}; \sigma_D); \sigma_F\}; \Sigma_F = \sigma_F^2 R\right] \\
&= F_2\left(\eta_{i1}\frac{\sigma_F}{\sigma_D}, \eta_{i2}\frac{\sigma_F}{\sigma_D}; \Sigma_F = \sigma_F^2 R\right).
\end{aligned}$$

Without anything further specified, $\phi^{-1} = \frac{\sigma_F}{\sigma_D}$ is not identifiable, as easily seen by equivalently restating $P(Y_{i1} = 1, Y_{i2} = 1) = F_2(\eta_{i1}, \eta_{i2}; \Sigma_D = \sigma_D^2 R)$ with no dependency on σ_F. Setting $\sigma_D = \sigma_F = 1$ invokes the multivariate probit model (Chib & Greenberg, 1998).

Table 2 shows that the estimation of β_1^M is dependent on D and insensitive to other specifications concerning the facilitating distribution. The coefficients for the previous examples differ as expected between a probit–link (D = N(0,1)) and logit–link (D = L(0,1)) models.

Restricted COP models remove the fundamental non-linearity issue associated with unrestricted COP models, however require multivariate facilitating distributions with desirable marginal distributions. Therefore, for example, if one would like probit or t-quantile marginal link functions, the multivariate normal and t are available as the facilitating distribution in a restricted COP model. If one desires marginal logit models, no easily specified multivariate logistic distribution is available. However, an unrestricted COP model with multivariate normal or t facilitating distributions and logit (or any) marginal link distribution is readily available (Genest *et al.*, 2012).

Thus, we observe the inverse relationship between threshold model complexity and restrictions on the model space: more mathematically simple and analytically convenient models transmute an unlimited selection of F and D into a restricted class; conversely, expanding the restricted class of F and D to the unrestricted class introduces undesirable complexities as the cost of gained flexibility in the unrestricted selection of F and D. We explore this continuum and put forth distinct restrictions in the form of properties of marginalised random-intercept models. We find the intersection of two such properties gives rise to a class of models that admits latent random effects and contains the very familiar probit-probit-normal model (Example 3.11) and a specification to give log odds interpretations on both the marginal and conditional scales (Example 3.12).

3 Modelling Framework Admitting Univariate Latent Random Intercepts

Up to this point, our emphasis has been on using copula models for creating marginal likelihoods for correlated binary data having a specified marginal link distribution, D. These results are well established in the copula literature; their inclusion is to set the stage for the main contribution of this manuscript: a modelling framework admitting latent random intercepts.

If marginal interpretation is all that is required for analysis, then specifying F_J to be a multivariate normal or multivariate t-distribution and setting D^{-1} to be the desired link function would be convenient. Conditional random effect models, however, are sought frequently. Hence, we put forth a marginal modelling framework that are as follows: (i) has D^{-1} as the marginal link function; (ii) has an embedded conditionally specified link function with linear random effects; and (iii) has a fully specified likelihood. As we will show, many of the current models possessing properties (i)–(iii) produce marginal likelihoods that are either restricted or unrestricted COP models. Let H and G be univariate CDFs with real line support and symmetric about 0. We explore the two-stage hierarchical model specified by $H^{-1}\{P(Y_{ij} = 1|u_i)\} = \Delta_{ij} + u_i$ and random intercept $u_i \sim G$. Suppose that the conditional model is marginalised and the result is F:

$$P(Y_{ij} = 1) := F(\Delta_{ij}) = \int H(\Delta_{ij} + u_i)dG \tag{6}$$

We taxonomize the following properties of marginalised random-intercept models:

(1) $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\}$	(Transference)
(2) $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\} = \phi^{-1}\eta_{ij}$	(Bridging)
(3) $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\} = \phi^{-1}\eta_{ij}$ with $D = H$ exactly	(Simultaneity)
(4) $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\}$ where D, F, G, H are in same family	(Closure)
(5) D, F, G, H to be α-stable distributions	(Stability)
(6) D, F, G, H to be α-stable distributions with $D = H$ exactly	(Simult-stability)

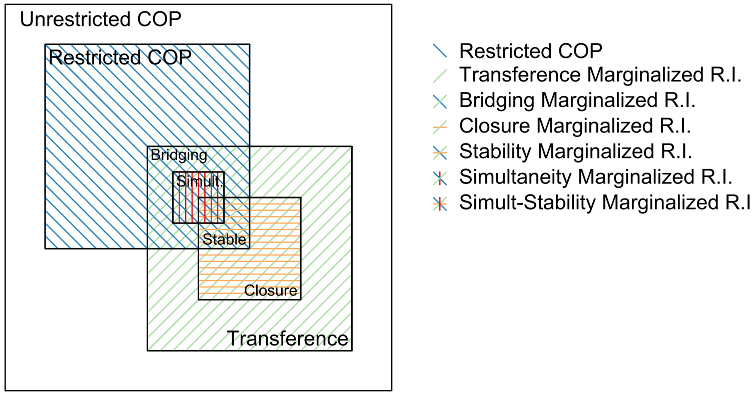

Figure 2. *A Venn diagram of how properties of marginalised random-intercept models restrict the model space and imply copulas over partitions (COP) models.*

Figure 2 gives a Venn diagram schematic of how the properties of marginalised random-intercept models inter-r elate and imply unrestricted and restricted COP models. *In our formulation, all marginalised random-intercept models exhibit transference.* Bridging, a special type of transference, places restrictions on F and D belonging to the same scale family. Simultaneity, a special type of bridging, further places the restriction of $D = H$ exactly. Closure requires D, F, G and H belong to the same family but does not require that family to be closed under scaling (Example 3.8). We now elaborate on the properties of marginalised random-intercept models and show how they imply COP models.

3.1 Transference

Deriving marginal likelihoods from conditional random-intercept models uses a transfer function Δ. The property of transference is herein defined as $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\}$, implying $D^{-1}\{P(Y_{ij} = 1)\} = \eta_{ij}$. That is, the transference property 'transfers' the marginal link distribution from F to D. In the succeeding texts, we state the previously unknown connection between the use of transference in MMMs for binary data and COP models.

Proposition 3.1 (Equivalence of likelihoods via transference)**.** A marginalised random-intercept model likelihood for binary data exhibiting transference imposes the likelihood of a unrestricted copula. A proof is in the Appendix.

Because application of the model requires calculating F, the marginal of F_J, we note that convenience dictates that H and G be in a family that is closed under scalar multiplication and convolution. In the succeeding texts, we discuss Examples 3.2 and 3.3; the latter has a closed form for F, whereas the former does not.

Example 3.2 (Transference, F not closed)**.** Heagerty (1999) selected $D = H = logistic(0, \pi^2/3, 1)$ and $G = N(0, \tau^2)$. The marginalisation is not closed or of a common parametric distribution form, so F must be evaluated numerically, using the R function `logit.normal.mle()` of Comstock & Heagerty (1999).

Example 3.3 (Transference, F closed)**.** Specifying $H = N(0,1)$ and the random effects distribution $G = N(0,\tau^2)$, Caffo & Griswold (2006) utilised this choice of H and G to make $F = N(0, 1+\tau^2)$, easing computation. D is left to be arbitrarily specified. For $D = logistic(0, \pi^2/3, 1)$, we have an unrestricted COP model of the type in Example 2.3 with $\sigma_F^2 = \sigma_H^2 + \sigma_G^2 = 1 + \tau^2$ and $\rho = \sigma_G^2/\sigma_F^2 = \tau^2/(1+\tau^2)$. As Table 2 shows, the estimates and interpretations are the same as Example 2.3 with the gain of τ^2, which is an estimate of heterogeneity in the population.

An unfortunate consequence of most transference models is that the conditional model is distorted by use of the transfer function. Specifically, the fixed effects are often not linear on the scale of the conditional link function. Moreover, the random effects are not added linearly on the same scale as the fixed effects (Larsen *et al.*, 2000; Lee & Nelder, 2004). This disagreement of scale is especially troubling if one interprets random effects as collections of unobserved cluster-specific deviations around the fixed effects. Bridging specifically circumvents these issues by restricting the transfer function to be linear.

3.2 Bridging

Bridging is a restricted type of transference so that $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\} = \eta_{ij}\phi^{-1}$.

Proposition 3.4 (Equivalence of likelihoods via bridging)**.** A marginalised random-intercept model likelihood for binary data exhibiting bridging is equivalent to the likelihood of a restricted copula. Furthermore, $\phi = \frac{S_D}{S_F}$, where S_D is the scale of D, S_F the scale of F and S_F will be a function of S_H and S_G, the scales of H and G, respectively. Also, the estimation of conditional model regression coefficients is enabled: $\eta_{ij}^M\phi^{-1} = \mathbf{X}_{ij}\beta^M\phi^{-1} = \mathbf{X}_{ij}\beta^C = \eta_{ij}^C$.

Example 3.5 (Bridging random-intercept distribution)**.** Wang & Louis (2003) derived a 'logit-logit-bridge' model, where $D = H = logistic(0, \pi^2/3, 1)$ is specified and G is 'bridged' in the sense that it analytically satisfies $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\} = \eta_{ij}\phi^{-1}$, where $\phi = \left(\frac{\pi^2/3}{\pi^2/3+\tau^2}\right)^{1/2}$, and τ^2 is the variance of the random-intercept bridge distribution.

Example 3.6 (Bridging conditional distribution)**.** One can reverse the assignment of distributions to H and G in Example 3.5 to provide an interesting twist on the Wang & Louis (2003) model. Consider a 'logit-(bridge-link)-logistic' model, where H is bridged, analytically satisfying $\Delta_{ij} = F^{-1}\{D(\eta_{ij})\} = \eta_{ij}\phi^{-1}$ and $D = G = logistic(0, \pi^2/3, 1)$. The scalar attenuation factor $\phi = \left(\frac{\pi^2/3}{\pi^2/3+\tau^2}\right)^{1/2}$ is invariant to the switching of G and H, and τ^2 is the variance of H, whose inverse serves as the conditional link.

We believe that Example 3.6 is the first instance of bridging the conditional link distribution, rather than bridging the random effect distribution [as in Wang & Louis (2003)]. Bridging the conditional link distribution is not entirely a theoretical exercise, as bridged link functions can be used in generalised linear mixed models (GLMM) software that fix the random effect distribution as Gaussian yet allow a user-specified conditional link function. Because of bridging, the random effects are on the same scale, ameliorating concerns raised by Lee & Nelder (2004). However, for simultaneous interpretations of β^M and β^C, the links D^{-1} and H^{-1} need to be identical while retaining bridging.

3.3 Simultaneity

Simultaneity is a restricted type of bridging such that $F^{-1}\{D(\eta_{ij})\} = \eta_{ij}\phi^{-1}$ and $H = D$ exactly. As for bridging, a simultaneity model yields conditional regression coefficients β^C, marginal regression coefficients β^M and the scalar attenuation factor ϕ that relates marginal and conditional effects according to $\beta^M = \phi\beta^C$. The additional feature of simultaneity over bridging is due to $H = D$, which gives rise to β^M and β^C affecting change on the same CDF curve. The parameter β^C is still a cluster-specific interpretation, but given u_i adjusts the slope of the same CDF as β^M.

Example 3.7 (Simultaneity recap)**.** The logit-logit-bridge model of Example 3.5 exhibits bridging and simultaneity. The logit-(bridge-link)-logistic model of Example 3.6 exhibits bridging yet lacks simultaneity. Heagerty's logit-logit-normal model of Example 3.2 is neither bridged nor simultaneous.

3.4 Closure

Closure is a sub-type of transference with D, F, G and H of the same family.

Example 3.8 (Closure)**.** Caffo *et al.* (2007) discussed a class of models, where D, G, H and F all belong to the mixture of normal distributions family. Connecting our notation to theirs:

$$
\begin{aligned}
D = H &:= F_w = \Sigma_{l=1}^{L_w} \pi_{w,l} \Phi\left(\frac{w - \mu_{w,l}}{\sigma_{w,l}}\right) \\
G &:= F_u = \Sigma_{l=1}^{L_u} \pi_{u,l} \Phi\left(\frac{u - \mu_{u,l}}{\sigma_{u,l}}\right) \\
F = \int H dG &:= F_q = \Sigma_{l=1}^{L_w} \Sigma_{l'=1}^{L_u} \pi_{w,l}\pi_{u,l'} \Phi\left(\frac{q - \mu_{w,l} - \mu_{u,l'}}{(\sigma_{w,l}^2 + \sigma_{u,l'}^2)^{1/2}}\right) \\
\Delta_{ij} &:= F_q^{-1}\left\{F_w(\eta_{ij}^M)\right\}
\end{aligned}
$$

Despite $D = H$, this class of model exhibits neither simultaneity nor bridging.

The main benefit of closure is the substantial simplification of the calculation of F. However, as Example 3.8 illustrates, one still may require non-linear transfer functions to obtain a desired marginal interpretation, even if D is in the same distributional family as F, G and H. Requiring closure and bridging to linearize Δ under one distributional family can only be satisfied with α-stable distributions.

3.5 Stability

Stability is a restricted type of transference that exhibits both closure and bridging. Our result is that the model is both closed and bridged if and only if D, G and H are of the same α-stable family.

Proposition 3.9 (Model equivalence between closed and bridged COP models and α-stable distributions)**.** A model will exhibit both closure and bridging if and only if D, G and H are of the same α-stable family. A proof is in the Appendix.

Example 3.10 (Stability COP model (general, $0 < \alpha \leq 2$)).

$$\begin{aligned} D &:= S(\alpha, 0, \gamma_D, 0; 0) \\ H &:= S(\alpha, 0, \gamma_H, 0; 0) \\ G &:= S(\alpha, 0, \gamma_G, 0; 0) \\ F = \int H dG &:= S(\alpha, 0, (\gamma_H^\alpha + \gamma_G^\alpha)^{1/\alpha}, 0; 0) \\ \Delta_{ij} &:= F^{-1}\left\{D(\eta_{ij}^M)\right\} = \eta_{ij}\phi^{-1}, \qquad \text{where } \phi = \frac{\gamma_D}{(\gamma_H^\alpha + \gamma_G^\alpha)^{1/\alpha}} \end{aligned}$$

Refer to Nolan (2010) for the parameterization of stable distributions.

If $\gamma_D = \gamma_H$ is enforced, then simultaneity is gained in addition to closure and bridging.

3.6 Simult-Stability

Example 3.11 (Simult-Stability: probit-probit-normal). Take Example 3.10, with $\alpha = 2$, $\gamma_D = \gamma_H = \frac{1}{\sqrt{2}}$, and $\gamma_G = \frac{\tau}{\sqrt{2}}$. This example is equivalent to Example 2.7, where $\sigma_F^2 = (1 + \tau^2)$, $\rho = \frac{\tau^2}{\sigma_F^2}$ and $\sigma_D^2 = 1$.

This example is also equivalent to that of Example 3.3 with D specified as $N(0, 1)$ instead of $logistic(0, \pi^2/3, 1)$. The gain is not only in conditional regression coefficients being estimable, $(\beta_0^C, \beta_1^C) = \phi^{-1}(\beta_0^M, \beta_1^M) = (-2.75, .078)$, but also in the interpretation of β_1^M and β_1^C (for a cluster i) being the slopes of the same CDF.

Note that ϕ was only identifiable analytically via τ^2 showing that all the estimates we needed were available from Examples 2.7 and 3.3, but the theoretical result of bridging allowed the information to render conditional estimates.

Example 3.12 (Simult-Stability: (approx-logit)-(approx-logit)-stable). Take Example 3.10, setting $\alpha = 1.89$ and $\gamma_D = \gamma_H = 1.2$. To obtain F_2, Monte Carlo sampling with the `fBasics` R package was utilised, and the likelihood was maximised via simulated annealing in `optim()` (Wuertz, 2009).

The stable distribution $D = H = S(1.89, 0, 1.2, 0; 0)$ approximates a standard logistic distribution well (Figure 3). The (approx-logit)-(approx-logit)-stable model would give rise to MLEs of β^M and γ_G, allowing β^C to be obtained. The regression coefficients β^M and β^C would have approximately a log odds interpretation on the population and cluster-specific metric, respectively, and the scale (not variance) γ_G could provide some analogue of heterogeneity typical to the variance of the random-intercept models.

Example 3.13 (Models with $D = H = logistic(0, \pi^2/3, 1)$ recap). The (approx-logit)-(approx-logit)-stable model of Example 3.12 is Simult-Stable and therefore exhibits bridging, simultaneity and closure as well as a log odds interpretation for both the conditional and marginal regression coefficients. The logit-logit-bridge model of Example 3.5 exhibits bridging, simultaneity and log odds interpretations of both the conditional and marginal regression coefficients but not closure. The logit-(bridge-link)-logistic model of Example 3.6 exhibits bridging and log odds interpretation for the marginal regression coefficients yet lacks simultaneity, closure and log odds interpretation for the conditional regression coefficients. Heagerty's

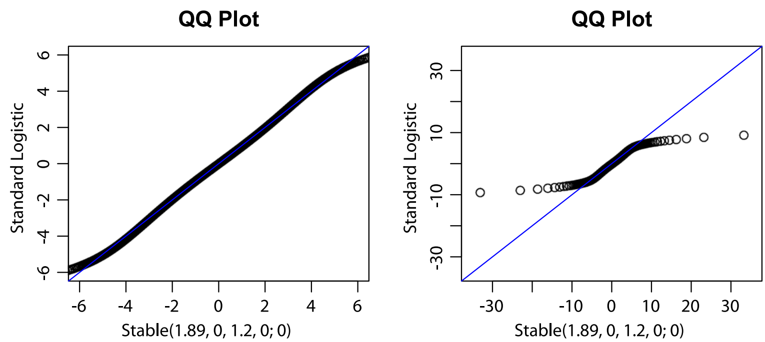

Figure 3. *The stable distribution can approximate the logistic distribution well. A quantile of (-5.6, 5.6) corresponds to a probability of (0.004, 0.996), respectively, showing that for all but the most extreme data of overwhelming 0 or 1 cases the stable distribution can approximate the logistic distribution.*

logit-logit-normal model of Example 3.2 exhibits log odds interpretations of the marginal regression coefficients but does not provide bridging, simultaneity, closure or log odds interpretation for the conditional regression coefficients.

4 Modelling Framework Admitting Multivariate Random Effects

Whereas the previous section admits one random intercept per cluster, this section generalises results for multivariate random effects. Consider $\mathbf{u}_i \sim G_Q$, a Q-variate distribution and $\mathbf{Z}_{ij}\mathbf{u}_i \sim G_{ij}$, a univariate distribution where stacked row vectors $\mathbf{Z}_{ij} = (Z_{ij1}, \ldots, Z_{ijQ})$ produce the $J_i \times Q$ matrix $\mathbf{Z}_i$ whose Q columns are a subset of the K columns of $\mathbf{X}_i$.

The property of transference in $H^{-1}\{P(Y_{ij} = 1|u_{ij})\} = \Delta_{ij} + \mathbf{Z}_{ij}\mathbf{u}_i$ has $\Delta_{ij} = F_{ij}^{-1}\{D(\eta_{ij})\}$ implying a COP model likelihood if the convolution $F_{ij} = H * G_{ij}$ are the marginal distributions of the facilitating distribution F_{J_i} where $\Sigma(Z_i)$ is a Z_i-dependent covariance structure and $C_{FD}(\eta_{i1}, \ldots, \eta_{iJ_i}) = F_{J_i}\left[F_{i1}^{-1}\{D(\eta_{i1})\}, \ldots, F_{iJ_i}^{-1}\{D(\eta_{iJ_i})\}; \Sigma(Z_i)\right]$.

The property of bridging is not readily generalisable, even in the Simult-Stable case. To see this, take H and D to be the univariate standard normal CDF, and G_Q to be $MVN(0, \Sigma)$, which implies G_{ij} is $N(0, Z_{ij}\Sigma Z_{ij}^{\mathrm{T}})$ and F_{ij} is $N(0, 1 + Z_{ij}\Sigma Z_{ij}^{\mathrm{T}})$. Therefore, ϕ is now dependent on i and j, so that $\phi_{ij} = \frac{1}{\sqrt{1+Z_{ij}\Sigma Z_{ij}^{\mathrm{T}}}}$, which disrupts the calculation of β^C as β^M times a scalar independent of i and j.

For bridging to carry over in the multivariate random effect setting, the distribution of $\mathbf{Z}_{ij}\mathbf{u}_i$ must necessarily be independent of i and j. In the most recently considered Simult-Stable case, restrict Σ to have 0 correlation, variance σ_1^2 for the random intercept, variance σ_2^2 for the random slope and $\mathbf{Z}_{ij} = (Z_{ij1}, Z_{ij2}) = (1, -1)$ or $(1, 1)$ (possibly representing a random-intercept and random treatment effect). Straightforwardly, $\mathbf{Z}_{ij}\mathbf{u}_i \sim N(0, \sigma_1^2 + \sigma_2^2)$, which is independent of i and j, allowing the bridging coefficient $\phi = \frac{1}{\sqrt{1+\sigma_1^2+\sigma_2^2}}$. Thus, bridging can generalise to uncorrelated heteroskedastic elliptical multivariate random effect distributions that are transformed by $\mathbf{Z}_{ij}$ that satisfy both conditions $\forall i, j$: $\sum_q Z_{ijq}^2 = L$ and $|Z_{ijq}| = |Z_{i'j'q}|$. Bridging can generalise to uncorrelated homoskedastic (that is, spherical) multivariate random effect distributions if transformed by $\mathbf{Z}_{ij}$ that satisfy the first condition. Rotational matrix rows are included in the L=1 case, but not all L=1 cases represent rows from rotational matrices. Nevertheless, geometrically speaking, rotations of spherical multivariate random effect distributions allow for bridging to hold in the multivariate Stability setting.

5 Model Fitting

Fitting COP models remains a difficult task, as well outlined in the copula literature (Joe, 2005; Nikoloulopoulos & Karlis, 2009; Genest *et al.*, 2012). For large J_i, it may be useful to exploit the random effect structure and fit the model as MMM as opposed to finite differencing or in the COP model formulation. For random-intercept models, the log-likelihood has the form

$$\sum_i \log\left[\int \prod_{j=1}^{J_i} H(\Delta_{ij}+u_i)^{Y_{ij}}\{1-H(\Delta_{ij}+u_i)\}^{1-Y_{ij}}\,dG(u_i)\right],$$

a univariate integral in u_i, hence numerical integration can be performed. Gauss/Hermite quadrature is often useful if orthogonal polynomials for G exists (as is the case for most standard distributions). Because of COP models being likelihood-based, Bayesian approaches are possible. Specifying normal priors for the fixed effects, the appropriate random-intercept distribution for the intended model (Table 2) and a prior for the covariance structure as a whole or the individual components is a feasible consideration.

Fitting COP models without interest in the random effects is reasonable for smaller J_i. If the facilitating multivariate distribution function is manageable, then calculating the log-likelihood is easy, and maximisation using quasi-Newton algorithms can be done. For example, if the facilitating distribution is multivariate Gaussian or t, then accurate numerical approximations to the distribution function exist (Genz *et al.*, 2008; Genz & Bretz, 2009). However, these approximations tend to work well for smaller dimensions of J_i. For moderately large J_i, Monte Carlo approximations may need to be calculated. Specifically, let $\tilde{Y}_{i,k} = [\tilde{Y}_{i1,k}, \ldots, \tilde{Y}_{iJ_i,k}]$ be the k^{th} simulation from F_{J_i}. Then $\tilde{Y}'_{i,k} = [D^{-1}\{F(\tilde{Y}_{i1,k})\}, \ldots, D^{-1}\{F(\tilde{Y}_{iJ_i,k})\}]$ is a draw from C_{FD}. Therefore, a Monte Carlo approximation to the log-likelihood is given by $\sum_i \log\left\{\frac{1}{K}\sum_{k=1}^{K} \mathbb{1}_{\{Q'_i\}}(\tilde{Y}'_{i,k})\right\}$, where the quantity inside the logarithm is the Monte Carlo approximation of $\int_{Q'_i} dC_{FD}$; with a similar approach and numerical ease, the Monte Carlo approximation of $\int_{Q_i} dF$ could be calculated as in Example 3.12. One can then use Monte Carlo Maximum Likelihood (Geyer, 1992) to obtain parameter estimates. We note that for large J_i, the calculation of multivariate probabilities using *i.i.d.* Monte Carlo may be inefficient, necessitating Markov chain calculations in large problems (Robert, 1995). Motivated by the difficulty of MCMC methods, the facilitating distribution of a K-finite normal mixtures yielding a finite normal mixture copula has produced a numerically tractable option for model fitting (Nikoloulopoulos & Karlis, 2009). Such a method only requires inverting a numerically evaluated linear function of K univariate Gaussian CDFs as opposed to evaluating a J_i-dimensional integral.

6 Discussion

COP models are a unifying likelihood-based framework for correlated binary data, encapsulating multivariate probit and marginalised random-intercept models (Figure 4). Via a taxonomy

Model	a-Stable	Closure	Bridging $\beta^C=\phi\beta^M$	Force same marginal and conditional link	Simultaneity	Marginal likelihood equivalent to restricted COP	non-trivial Δ
(α-γ_D-stabit)-(α-γ_D-stabit)-(α-γ_G-stable)	✔	✔	✔	✔	✔	✔	-
probit-probit-normal	✔	✔	✔	✔	✔	✔	-
cauchit-cauchit-cauchy	✔	✔	✔	✔	✔	✔	-
comploglog-comploglog-bridge	-	-	✔	✔	✔	✔	-
logit-logit-bridge	-	-	✔	✔	✔	✔	-
logit-(bridge-link)-logistic	-	-	✔	-	-	✔	-
(mix-probit)-(mix-probit)-(mix-normal)	-	✔	-	✔	-	-	✔
link-link-normal	-	-	-	✔	-	-	✔
link-probit-normal	-	-	-	-	-	-	✔
logit-logit-normal	-	-	-	✔	-	-	✔
logit-probit-normal	-	-	-	-	-	-	✔

Model	D	H	G	F	ϕ	Papers/ References
(α-γ_D-stabit)-(α-γ_D-stabit)-(α-γ_G-stable)	$S(\alpha, 0, \gamma_D, 0; 0)$	$S(\alpha, 0, \gamma_D, 0; 0)$	$S(\alpha, 0, \gamma_G, 0; 0)$	$S(\alpha, 0, (\gamma_D^{\alpha}+\gamma_G^{\alpha})^{1/\alpha}, 0; 0)$	$\gamma_D/(\gamma_D^{\alpha}+\gamma_G^{\alpha})^{1/\alpha}$	
probit-probit-normal	$N(0, 1)$	$N(0, 1)$	$N(0, \tau^2)$	$N(0, 1+\tau^2)$	$1/(1+\tau^2)^{1/2}$	
cauchit-cauchit-cauchy	$C(0, 1)$	$C(0, 1)$	$C(0, \gamma_G)$	$C(0, 1+\gamma_G)$	$1/(1+\gamma_G)$	WL2003
comploglog-comploglog-bridge	$Gom(0, \pi^2/6, 1)$	$Gom(0, \pi^2/6, 1)$	$Bridge(0, \tau^2)$		$\{(\pi^2/6)/(\pi^2/6+\tau^2)\}^{\{1/2\}}$	WL2003
logit-logit-bridge	$L(0, \pi^2/3, 1)$	$L(0, \pi^2/3, 1)$	$Bridge(0, \tau^2)$	$L(0, (\pi^2/3+\tau^2), 1)$	$\{(\pi^2/3)/(\pi^2/3+\tau^2)\}^{\{1/2\}}$	WL2003
logit-(bridge-link)-logistic	$L(0, \pi^2/3, 1)$	$Bridge(0, \tau^2)$	$L(0, \pi^2/3, 1)$	$L(0, (\pi^2/3+\tau^2), 1)$	$\{(\pi^2/3)/(\pi^2/3+\tau^2)\}^{\{1/2\}}$	WL2003
(mix-probit)-(mix-probit)-(mix-normal)	Fw	Fw	Fu	Fq	-	CAR2005
link-link-normal	D	D	$N(0, \tau^2)$	-	-	H1999, HZ2000
link-probit-normal	D	$N(0, 1)$	$N(0, \tau^2)$	$N(0, 1+\tau^2)$	-	CG2005
logit-logit-normal	$L(0, \pi^2/3, 1)$	$L(0, \pi^2/3, 1)$	$N(0, \tau^2)$	-	-	H1999
logit-probit-normal	$L(0, \pi^2/3, 1)$	$N(0, 1)$	$N(0, \tau^2)$	$N(0, 1+\tau^2)$	-	CG2005

Figure 4. *A table showing properties of marginalised random-intercept models in the literature. We commandeer the hyphenated phrase terminology '$\boldsymbol{D}^{-1} - \boldsymbol{H}^{-1} - \boldsymbol{G}$' to represent models, with $\boldsymbol{D}^{-1}$ always being the marginal link and $\boldsymbol{H}^{-1}$ only being a conditional link when bridging is present. If bridging is not present, $\boldsymbol{H}^{-1}$ does not link to a linear predictor, and is by definition not a conditional link, despite $\boldsymbol{H}$ being specified in the integrand of the marginalisation equation. Reference key: H1999 is Heagerty (1999), HZ2000 is Heagerty and Zeger (2000), WL2003 is Wang and Louis (2003), CAR2005 is Caffo, An, and Rohde (2005), and CG2005 is Caffo and Griswold (2005).*

of properties, we arrived at COP models producing not only marginal estimates but also measures of heterogeneity and conditional model estimates. The COP model class includes many marginalised random-intercept models where the random-intercept distribution is heavier tailed than a normal distribution, which deviates from traditional methods but may have benefits (Lee & Thompson, 2007).

We have connected MMMs to copular modelling with fixed marginals, indicating that copular approaches can have workable, embedded conditional models. Some view a conditional GLMM random-intercept model (which is likelihood based) as an alternative to any marginal model. However, in some instances, if the effect being modeled does not change within cluster (e.g. race, sex and treatment), then the conditional model's fixed effect regression coefficients take on a counterfactual interpretation (Neuhaus *et al.*, 1991). If the goal of the analysis is to compare the effect between groups of clusters (Black versus White populations, male versus female populations, placebo versus active drug populations, etc) then a marginal inference is needed. Undoubtedly, the most popular marginal modelling approach utilises generalised estimating equations (GEEs) modelling (Liang & Zeger, 1986; Zeger *et al.*, 1988; Liang *et al.*, 1992). As the name suggests, GEE methods do not specify a full likelihood (Lee & Nelder, 2009). Instead, GEEs restrict the first two moments of the data and possess asymptotic robustness properties on the second moment. The GEE framework, however, does not produce random effects that can be estimated, which prevents use for cluster-specific prediction. Until recently, the GEE approach precluded the benefits of full likelihood-based analysis of data, such as weaker restrictions on missing data processes for valid inferences (Robins *et al.*, 1995; Scharfstein *et al.*, 1999). Recent work has lifted those limitations, potentially making the GEE a close competitor in situations of missing data in marginal models (Seaman & Copas, 2009; Birhanu *et al.*, 2011; Beunckens *et al.*, 2008). Whether or not a GEE model is consistent with any likelihood-based model is a hard and open problem [for example the discussion of Fréchet bounds (Chaganty & Joe, 2004; 2006)].

Copulas over partition models with embedded conditional models still run the risk of misspecifying the random effects distribution and how such misspecification propagates to the marginal inference warrants further study. Heagerty (1999) chided that the random-intercept variance having dependence on covariates or serially correlated random effects can induce slight bias in the marginal coefficients. MMMs and the equivalent COPs inherit the considerations of the conditional models at their core when it comes to model fit assessment and random effects distribution misspecification: the interplay and untestable assumptions of choices of D, H and G, as well as the covariate modelling choices still stand as an area of challenging research (Agresti *et al.*, 2004; Lee & Thompson, 2007; Caffo *et al.*, 2007). Assessing model fit is also another challenging area of research for MMMs, but one can now utilise tools for copular model assessment methods, which include copula selection by certain deviation measures of the observed data from the assumed copula family and goodness-of-fit testing involving the assessment of the significance of the observed deviation measure against its null distribution under the assumed copula family (Berg, 2009; Yan, 2007).

Further extensions of the proposed COP model would be worthwhile and should be investigated in future work. The taxonomy of properties outlined is due to the restrictions on D, F, G and H and serves as a structure for the focused case of clustered exchangeable binary outcomes. The relaxing of restrictions will allow more flexible correlation structures to be represented. A crucial step in building upon these promising but intentionally restricted findings of this manuscript is to formulate the model in terms of conditions on the class of GLMMs and hierarchical generalised linear models (Lee & Nelder (1996)) to yield sub-classes that imply COP models.

Two natural extensions exist with regard to the COP model taxonomy: relaxing the exchangeable correlation structure and allowing more general multilevel models (admitting random slopes in addition to random intercepts). These two extensions are related, as the exchangeable structure is implied by the conditional single random-intercept model. The exchangeable structure can be relaxed and the changes to the taxonomy explored. Concerning specifically the Wang-Louis model in Example 3.5, Parzen *et al.* (2011) and Boehm *et al.* (2012) deviated from exchangeability through the use of multiple random intercepts, allowing for correlation to decrease over time and spatially correlated structures, respectively. The implementation utilises a copula for the random effects distribution to get a multivariate distribution with bridge marginals. In that vein, (6) could be updated to $P(Y_{ij} = 1) := F(\Delta_{ij}) = \int H(\Delta_{ij} + u_{ij}) dG_{J_i}$ with multiple random intercepts $u_i = (u_{i1}, u_{i2}, \ldots, u_{iJ_i}) \sim G_{J_i}$. Here, G_{J_i} is a multivariate distribution with marginals that have the same properties as G specified in the taxonomy thereby preserving the taxonomy and allowing for non-exchangeable structures. That is, having a copula-assisted random-intercepts distribution can still induce a COP model, one with AR(1) or spatially correlated structures. The taxonomy would need to be modified in light of bridging depending on more than one ϕ, as illustrated by Boehm *et al.* (2012). Not to be overlooked is the wealth of copula literature that put dependence modelling at the forefront of the analysis as opposed to GEE's treatment of them as a nuisance. Directions for navigating away from exchangeable to more flexible structures are carefully outlined (Nikoloulopoulos & Karlis, 2009; Genest *et al.*, 2012).

The limitation to random intercepts can be relaxed for the consideration of general multilevel models that maintain at least the bridging property. Section 4 entertained spherical random effect $\mathbf{u}_i$ distributions under rotations. Admittedly, the practical scope is then limited to zero-correlated random effects, but the exercise geometrically motivates potentially focusing on elliptical distributions that can be rendered spherical. Elliptical distributions (non-zero-correlated random effects), which include multivariate symmetric-about-zero α-stable distributions, can be made spherical via linear transformations $\mathbf{Z}_i$ of $\mathbf{u}_i$ (Rachev, 2003). The challenge will be formulating the appropriate $\mathbf{Z}_i$ to be ρ-dependent so that appropriate rotations are informed by the data via estimation, as well as a procedure of back transformation for the purposes of interpretation as a random slope.

Only binary outcomes were considered; therefore, formulating COP models for clustered polytomous outcomes is a topic of future work. Literature exists and is actively researching extensions of the bridge density for continuous and discrete data as well as missing data (Lin *et al.*, 2010; Li *et al.*, 2011; Su *et al.*, 2011). Bayesian castings of COP model likelihoods should also be considered.

7 Supplementary Materials

A subset of the examples' R code is provided online.

Acknowledgements

Dr Caffo was supported by Grant Number R01EB012547 from the National Institute of Biomedical Imaging And Bioengineering and Grant Number R01NS060910 from the National Institute of Neurological Disorders and Stroke. Dr Crainiceanu was supported by the National Institute of Neurological Disorders and Stroke (R01NS060910).

References

Agresti, A., Caffo, B. & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput. Statist. Data. Anal.*, **47**(3), 639–653.

Albert, J.H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.*, **88**(422), 669–679.

Berg, D. (2009). Copula goodness-of-fit testing: An overview and power comparison. *Eur. J. Finance*, **15**(7-8), 675–701.

Beunckens, C., Sotto, C. & Molenberghs, G. (2008). A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data. *Comput. Statist. Data Anal.*, **52**(3), 1533–1548.

Bhat, C.R. & Sener, I.N. (2009). A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *J. Geogr. Syst.*, **11**(3), 243–272.

Birhanu, T., Molenberghs, G., Sotto, C. & Kenward, M.G. (2011). Doubly robust and multiple-imputation-based generalized estimating equations. *J. Biopharm. Stat.*, **21**(2), 202–225.

Boehm, L., Reich, B. & Bandyopadhyay, D. (2012). Bridging conditional and marginal inference for spatially-referenced binary data. *Biometrics*, **69**, 545–554.

Caffo, B., An, M.W. & Rohde, C. (2007). Flexible random intercept models for binary outcomes using mixtures of normals. *Comput. Statist. Data Anal.*, **51**(11), 5220–5235.

Caffo, B. & Griswold, M. (2006). A user-friendly tutorial on link-probit-normal models. *Amer. Statist.*, **60**, 139–145.

Chaganty, N.R. & Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *J. Roy. Stat. Soc. Ser. B*, **66**(4), 851–860.

Chaganty, N.R. & Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, **93**(1), 197–206.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**(2), 347–361.

Comstock, B.A. & Heagerty, P.J. (1999). *Intro to lnMLE: An R Package for Marginally Specified Logistic-Normal Models for Longitudinal Binary Data*. Available from: http://faculty.washington.edu/heagerty/Software/LDA/MLV/lnMLEhelp.pdf [Accessed on 15 August 2013].

Genest, C., Nikoloulopoulos, A.K., Rivest, L.P. & Fortin, M. (2012). Predicting dependent binary outcomes through logistic regressions and meta-ellliptical copulas. *Braz. J. Probab. Statist.*, **27**(3), 265–284.

Genz, A. & Bretz, F. (2009). *Computation of Multivariate Normal and t Probabilities.* Berlin Heidelberg: Springer Verlag.

Genz, A., Bretz, F., Hothorn, T., Miwa, T., Mi, X., Leisch, F. & Scheipl, F. (2008). *mvtnorm: Multivariate Normal and t Distribution*. R package version 0.9-2, Available from: http://CRAN.R-project.org/package=mvtnorm [Accessed on 15 August 2013].

Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Stat. Sci.*, **7**(4), 473–483.

Griswold, M.E., Swihart, B.J., Caffo, B.S & Zeger, S.L. (2013). Practical marginalized multilevel models. *Stat*, **2**(1), 129–142.

Heagerty, P.J. (1999). Marginally specified logistic-normal models for longitudinal binary data. *Biometrics*, **55**(3), 688–698.

Heagerty, P.J. & Zeger, S.L. (2000). Marginalized multilevel models and likelihood inference (with comments and a rejoinder by the authors). *Stat. Sci.*, **15**(1), 1–26.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *J. Multivariate Anal.*, **94**(2), 401–419.

Larsen, K., Petersen, J.H., Budtz-Jørgensen, E. & Endahl, L. (2000). Interpreting parameters in the logistic regression model with random effects. *Biometrics*, **56**(3), 909–914.

Lee, Y. & Nelder, J.A. (1996). Hierarchical generalized linear models. *J. Roy. Stat. Soc. B Met.*, **58**(4), 619–678.

Lee, Y. & Nelder, J.A. (2004). Conditional and marginal models: Another view. *Stat. Sci.*, **19**(2), 219–228.

Lee, Y. & Nelder, J.A. (2009). Likelihood inference for models with unobservables: Another view. *Stat. Sci.*, **24**(3), 255–269.

Lee, K.J. & Thompson, S.G. (2007). Flexible parametric models for random-effects distributions. *Stat. Med.*, **27**(3), 418–434.

Li, X., Bandyopadhyay, D., Lipsitz, S. & Sinha, D. (2011). Likelihood methods for binary responses of present components in a cluster. *Biometrics*, **67**(2), 629–635.

Liang, K.Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**(1), 13–22.

Liang, K.Y., Zeger, S.L. & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *J. Roy. Stat. Soc. B Met.*, **54**(1), 3–40.

Lin, L., Bandyopadhyay, D., Lipsitz, S. & Sinha, D. (2010). Association models for clustered data with binary and continuous responses. *Biometrics*, **66**(1), 287–293.

Meester, S.G. & Mackay, J. (1994). A parametric model for cluster correlated categorical data. *Biometrics*, **50**(4), 954–963.

Nelsen, R.B. (2006). *An Introduction to Copulas.* New York: Springer Verlag.

Neuhaus, J.M., Kalbfleisch, J.D. & Hauck, W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int. Stat. Rev.*, **59**(1), 25–35.

Nikoloulopoulos, A.K. & Karlis, D. (2008). Copula model evaluation based on parametric bootstrap. *Comput. Statist. Data Anal.*, **52**(7), 3342–3353.

Nikoloulopoulos, A.K. & Karlis, D. (2008). Fitting copulas to bivariate earthquake data: The seismic gap hypothesis revisited. *Environmetrics*, **19**(3), 251–269.

Nikoloulopoulos, A.K. & Karlis, D. (2008). Multivariate logit copula model with an application to dental data. *Stat. Med.*, **27**(30), 6393–6406.

Nikoloulopoulos, A.K. & Karlis, D. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *J. Statist. Plann. Inference*, **139**(11), 3878–3890.

Nolan, J.P. (2010). *Stable Distributions–Models for Heavy Tailed Data.* Boston: Birkhäuser.

O'Brien, S.M. & Dunson, D.B. (2004). Bayesian multivariate logistic regression. *Biometrics*, **60**(3), 739–746.

Parzen, M., Ghosh, S., Lipsitz, S., Sinha, D., Fitzmaurice, G.M., Mallick, B.K. & Ibrahim, J.G. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *Ann. Appl. Stat.*, **5**(1), 449–467.
Rachev, S.T. (2003). *Handbook of Heavy Tailed Distributions in Finance.* Amsterdam: Elsevier.
Robert, C.P. (1995). Simulation of truncated normal variables. *Stat. Comput.*, **5**(2), 121–125.
Robins, J.M., Rotnitzky, A. & Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.*, **90**(429), 106–121.
Ruppert, D. (2010). *Statistics and Data Analysis for Financial Engineering.* New York: Springer Verlag.
Scharfstein, D.O., Rotnitzky, A. & Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *J. Amer. Statist. Assoc.*, **94**(448), 1096–1120.
Schweizer, B. (1991). Thirty years of copulas. In *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, pp. 13–50. Dordrecht: Springer.
Seaman, S & Copas, A. (2009). Doubly robust generalized estimating equations for longitudinal data. *Stat. Med.*, **28**(6), 937–955.
Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika*, **9**, 449–460.
Su, L., Tom, B.DM & Farewell, V.T. (2011). A likelihood-based two-part marginal model for longitudinal semi-continuous data. *Stat. Methods Med. Res.,* DOI:10.1177/0962280211414620.
Tielsch, J.M., Sommer, A., Katz, J., Quigley, H. & Ezrine, S. (1991). Socioeconomic status and visual impairment among urban Americans. *Arch. Ophthalmol.*, **109**(5), 637–641.
Wang, Z. & Louis, T.A. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika*, **90**(4), 765–775.
Wuertz, D. (2009). *fbasics: Rmetrics–Markets and Basic Statistics R package version 2100.78*. http://CRAN.R-project.org/package=fBasics [Accessed on 15 August 2013].
Yan, J. (2007). Enjoy the joy of copulas: With a package copula. *J. Stat. Softw.*, **21**(4), 1–21.
Zeger, S.L., Liang, K.Y. & Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, **44**(4), 1049–1060.

Appendix A

Construction of Definition 2.1:

Let $\tilde{Y}'_{i1}, \ldots, \tilde{Y}'_{iJ} \sim P_{C_{FD}}$ and $Y_{i1}, \ldots, Y_{iJ}$ be potentially correlated binary outcomes with associated covariates $x_{i1}, \ldots, x_{iJ_i}$. Consider the model $Y_{ij} = \mathbb{1}_{\{A'_{ij}(\eta_{ij})\}}(\tilde{Y}_{ij})$ where $A'_{ij}(\eta_{ij}) = \{c : c \leq \eta_{ij} | \beta, X_i\}$. Define $Q'_{ij}(\eta_{ij}, Y_{ij})$ as $A'_{ij}(\eta_{ij})$ if $Y_{ij} = 1$ and $\bar{A}'_{ij}(\eta_{ij})$ if $Y_{ij} = 0$. Allow $Q'_i = \otimes_j Q'_{ij}(\eta_{ij}, Y_{ij})$, the Cartesian cross of all the sets $A'_{ij}(\eta_{ij})$ and $\bar{A}'_{ij}(\eta_{ij})$. Q'_i corresponds to all the particular combinations of 1s and 0s in the vector $\mathbf{Y}_i$ and is the subset of the partition of the support for the latent variable, $\tilde{\mathbf{Y}}_i$.

Hence, assuming independence across clusters and setting $\eta_{ij} = \eta^M_{ij} = x_{ij}\beta^M$, we have a likelihood contribution for cluster i, $L_i(\beta^M, \Sigma(\rho))$, equal to

$$P(Y_{i1} = y_{i1}, \ldots, Y_{iJ_i} = y_{iJ_i}) = P_{C_{FD}}\left(\otimes_j Q'_{ij}(\eta^M_{ij}, y_{ij})\right) \quad = P_{C_{FD}}(Q'_i) = \int_{Q'_i} dC_{FD} \tag{A1}$$

Using (A1), the unity case makes Q'_i the Cartesian cross of all the sets with the '$\leq$' operator, allowing us to write the integral in terms of the multivariate CDF, C_{FD}:

$$P(Y_{i1} = 1, \ldots, Y_{iJ_i} = 1) = \int_{Q'_i} dC_{FD} \quad = \int_{A'_{i1}} \cdots \int_{A'_{iJ_i}} dC_{FD} = C_{FD}(\eta_{i1}, \ldots, \eta_{iJ_i}) \tag{A2}$$

Appendix B

Proof of Proposition 2.2:

Using (A2) and the properties of copulas, $P(Y_{ij} = 1)$ is equivalent to

$$P_{C_{FD}}(\mathbb{R} \otimes \ldots \otimes \mathbb{R} \otimes A'_{ij}(\eta^M_{ij}) \otimes \mathbb{R} \otimes \ldots \otimes \mathbb{R}) = \lim_{e \to \infty} C_{FD}(e, \ldots, e, \eta_{ij}, e, \ldots, e) \ = D(\eta_{ij})$$

Appendix C

Proof of Proposition 3.1 :

Allow random variables $\tilde{Y}_{ij} \sim F$, $W_{ij} \sim H$, $u_i \sim G$. By the convolution equation (6), $\tilde{Y}_{ij} = W_{ij} - u_i$, and vector $\tilde{\mathbf{Y}}_i = \mathbf{W}_i - u_i \mathbb{1}_{\{J_i \times 1\}} \sim F_{J_i}$. Utilising two assumptions, (1) $Y_{ij} = y_{ij}|u_i, X_i$ are independent (which implies $W_{ij}|u_i, X_i$ are independent), and (2) $u_i \overset{\text{iid}}{\sim} G(u_i|X_i)$:

$$\begin{aligned}
L_i &= \int \prod_{j=1}^{J_i} H(\Delta_{ij} + u_i)^{y_{ij}} \{1 - H(\Delta_{ij} + u_i)\}^{1-y_{ij}} dG(u_i|X_i) \\
&= \int \prod_{j=1}^{J_i} P(W_{ij} \leq \Delta_{ij} + u_i|u_i, X_{ij})^{y_{ij}} \{P(W_{ij} > \Delta_{ij} + u_i|u_i, X_{ij})\}^{1-y_{ij}} dG(u_i|X_i) \\
&= \int P_{W_i}(\otimes_j Q'_{ij}(\Delta_{ij} + u_i, y_{ij})|u_i, X_{ij}) dG(u_i|X_i) \\
&= \int P_{Y_i}(\otimes_j Q'_{ij}(\Delta_{ij}, y_{ij})|u_i, X_{ij}) dG(u_i|X_i) \\
&= \int P_{Y_i}(\otimes_j Q'_{ij}(\Delta_{ij}, y_{ij})|X_{ij}) dG(u_i|X_i) \\
&= P_{Y_i}(\otimes_j Q'_{ij}(\Delta_{ij}, y_{ij})|X_{ij}) \\
&= P_{Y_i}(\otimes_j Q_{ij}(\eta_{ij}, y_{ij})|X_{ij}) \\
&= P_{Y_i}(Q_i|X_{ij}) \\
&= \int_{Q_i} dF_{J_i}
\end{aligned}$$

Appendix D

Proof of Proposition 3.9:

The '$\Leftarrow$' direction is trivial given scaling and convolution properties of α-stable distributions.
Model exhibiting closure and bridging $\Rightarrow D, F, G, H$ α-stable:
Let $X_D \sim D$, $X_F \sim F$, $X_g \sim G$, $X_H \sim H$, and with $F = H * G$, we have $X_F = X_H - X_G$ with a distribution from the same family as X_H and X_G because of closure. Bridging implies $X_F = k_D X_D$ follows a distribution from the same scalar family as D, and by closure, H and G are in that scalar family. Thus, setting the two representations of X_F equal, we can write $X_H - X_G = k_D X_D$ as $\frac{1}{k_H} X_F - \frac{1}{k_G} X_F = X_F$, which is a special case of the definition of a stable distribution (Nolan, 2010).

[*Received July 2012, accepted July 2013*]