View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of East Anglia digital repository

Weighted scores estimating equations and CL1 information criteria for longitudinal ordinal response*

Abstract

Available extensions of generalized estimating equations for longitudinal ordinal response require a conversion of the ordinal response to a vector of binary category indicators. That leads to a rather complicated working correlation structure and to large matrices when the number of categories and dimension of the clusters are large. Weighted scores estimating equations are constructed to overcome the aforementioned problems. Similar to generalized estimating equations which construct unbiased equations weighting the residuals, the weighted scores weight the univariate score functions. To specify the weight matrices, the weighted scores estimating equations use a working dependence model, namely the multivariate normal (MVN) copula model with univariate ordinal probit or logit regressions as the marginals. There is no need to convert the ordinal response to binary indicators, thus the weight matrices have smaller dimensions. Composite likelihood information criteria are further proposed as an intermediate step for selecting both the covariates in the mean function modelling and the structure of the latent correlation matrix induced by the MVN latent variables. The weighted scores estimating equations and composite likelihood information criteria are illustrated by analyzing a rheumatoid arthritis clinical trial. Our modelling framework is implemented in the package **weightedScores** within the open source statistical environment R.

Keywords: AIC/BIC; Composite likelihood; Correlation structure selection; Generalized estimating equations; Ordinal regression; Variable selection.

1 Introduction

The method of generalized estimating equations (Liang and Zeger, 1986; Zeger and Liang, 1986, GEE hereafter), which is popular in biostatistics, analyzes correlated data by assuming a generalized linear model (GLM) for the outcome variable, and a working correlation matrix to describe the pattern of association among the repeated measurements on each subject or cluster. The associations are treated as nuisance parameters; interest focuses on the statistical inference for the regression parameters and the method is based only on moments up to second order.

Extending GEE to ordinal response data, say with K categories, requires an alteration of the general theory because the first and second moments are not defined for ordinal observations. This modification is based on a conversion of the ordinal response to a vector of $K-1$ binary indicators of categories $1, \ldots, K-1$ (Lipsitz et al., 1994b; Heagerty and Zeger, 1996; Parsons et al., 2006; Touloumis et al., 2013). There are various options for choosing the binary variables and also various parts of associations which will eventually describe all of the

*Word count: 10497.

possible outcomes for the original ordinal responses. The first part is the association between the binary variables at one time point. The second is the association of the same coded binary variables across time, and the third and final part is the association of two differently coded binary variables across time (Nooraee et al., 2014). This leads to a rather complicated association structure and the introduction of large matrices when K and d are large, where d is the dimension of a "cluster" or "panel".

Parsons (2013) discussed many realistic examples, where repeated ordinal data with a large number of categories are involved, e.g., clinical scoring systems such as the Oxford Hip Score (Dawson et al., 1996), and used the Warwick Arthroplasty Trial (Achten et al., 2010) data with $K = 49$ categories. Parsons (2013) described these as long ordinal scores or composite ordinal scores, that result from complex surveys. When K or d is large the working correlation matrix, a square matrix of dimension $d(K-1)$, is very large, as it needs to account for correlations between the K-1 new binary scores at each time-point and between time-points. Matrix operations (e.g. inversion) required for parameter estimation can become very slow or even infeasible for large K or d. For long ordinal scores, with a large number of categories, this presents a problem, as many cut-point parameters would need to be estimated with presumably poor precision and likely convergence problems that are often particularly associated with models for repeated ordinals scores (Parsons, 2013). We don't necessarily associate lack of convergence for GEE (Lipsitz et al., 1994a; Parsons et al., 2006; Parsons, 2013; Touloumis et al., 2013; Nooraee et al., 2014) with large matrices. However, large matrices are a problem for GEE methods, in as much as they cause model fitting algorithms to run slowly.

In this paper, to overcome the aforementioned problems, we construct the weighted scores estimating equations for longitudinal ordinal data by extending the theory in Nikoloulopoulos et al. (2011). Nikoloulopoulos et al. (2011) developed the weighted scores method for regression models with longitudinal continuous or discrete response data. For concreteness, the theory was illustrated using two forms of marginal negative binomial regression as in Cameron and Trivedi (1998) to handle longitudinal and cluster count response data with overdispersion. Similar to GEE which constructs unbiased equations weighting the residuals, the weighted scores method weights the univariate score functions. To specify the weights, the weighted scores estimating equations use a working dependence model, namely the multivariate normal (MVN) copula model with univariate ordinal probit or logit regressions as the marginals to model the association across time-points. There is no need to convert the ordinal response to binary indicators, thus the weight matrices have smaller dimensions and it is not necessary to guess the associations of indicator variables for different categories.

Model selection is an important issue in longitudinal data analysis, since when conducting a GEE analysis, it is essential to carefully model the association parameters, in order to avoid a substantive loss in efficiency in the estimation of the regression parameters (Albert and McShane, 1995; Crowder, 1995; Wang and Carey, 2003; Shults et al., 2009). Hence, we further extend the composite likelihood information criteria in Nikoloulopoulos (2016) as an intermediate step for selecting both the covariates in the mean function modelling and the structure of the latent correlation matrix induced by the MVN latent variables. The proposed criteria have the similar attractive property with the quasi-likelihood under the independence model criterion (QIC) for GEE (Pan, 2001), i.e., they allow both covariate and working correlation structure selection.

The remainder of the paper proceeds as follows. Section 2 constructs the weighted scores estimating equations for longitudinal ordinal data and derives the composite likelihood information criteria in the context of longitudinal data analysis with an ordinal margin. Section 3 describes the simulation studies we performed to gauge the efficiency and robustness of the weighted scores method, and to assess the performance of the

composite likelihood information criteria for longitudinal ordinal data. We discuss an application example in Section 4 and conclude with some discussion in Section 5, followed by a brief section with software details and a technical Appendix.

2 Weighted scores estimating equations and CL1 information criteria

2.1 Weighted scores estimating equations

The weighted scores method does not depend on a constant cluster size d, but in what follows we assume for notational ease that all n clusters have the same dimension d. Let p be the number of covariates, that is, the dimension of a covariate vector $\mathbf{x}$. Let Z be a latent variable with cumulative distribution function (cdf) $\mathcal{F}$, such that $Y = y$ if $\alpha_{y-1} + \mathbf{x}^T\boldsymbol{\beta} \leq Z \leq \alpha_y + \mathbf{x}^T\boldsymbol{\beta}$, $y = 1, \ldots, K$, where K is the number of categories of Y (without loss of generality, we assume $\alpha_0 = -\infty$ and $\alpha_K = \infty$), and $\boldsymbol{\beta}$ is the p-dimensional regression vector. From this definition, the response Y is assumed to have probability mass function

$$f_1(y; \nu, \boldsymbol{\gamma}) = \mathcal{F}(\alpha_y + \nu) - \mathcal{F}(\alpha_{y-1} + \nu),$$

where $\nu = \mathbf{x}^T\boldsymbol{\beta}$ is a function of $\mathbf{x}$ and the p-dimensional regression vector $\boldsymbol{\beta}$, and $\boldsymbol{\gamma} = (\alpha_1, \ldots, \alpha_{K-1})$ is the q-dimensional vector of the univariate cutpoints ($q = K - 1$). Note that $\mathcal{F}$ normal leads to the probit model and $\mathcal{F}$ logistic leads to the cumulative logit model for ordinal response (Agresti, 2010, Section 3.3.2).

Suppose that the data are $(y_{ij}, \mathbf{x}_{ij})$, $j = 1, \ldots, d$, $i = 1, \ldots, n$, where i is an index for individuals or clusters, j is an index for the repeated measurements or within cluster measurements. The univariate marginal model for Y_{ij} is $f_1(y_{ij}; \nu_{ij}, \boldsymbol{\gamma})$, where $\nu_{ij} = \mathbf{x}_{ij}^\top\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of dimension q is the vector of univariate cutpoints. If for each i, $Y_{i1}, \ldots, Y_{id}$ are treated as independent, then the log-likelihood is

$$L_1 = \sum_{i=1}^{n}\sum_{j=1}^{d} \log f_1(y_{ij}; \nu_{ij}, \boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{d} \ell_1(\nu_{ij}, \boldsymbol{\gamma}, y_{ij}),$$

where $\ell_1(\cdot) = \log\ f_1(\cdot)$. The score equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are

$$\begin{pmatrix} \frac{\partial L_1}{\partial \boldsymbol{\beta}} \\ \frac{\partial L_1}{\partial \boldsymbol{\gamma}} \end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{d} \begin{pmatrix} \mathbf{x}_{ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_q \end{pmatrix} \begin{pmatrix} \frac{\partial \ell_1(\nu_{ij}, \boldsymbol{\gamma}, y_{ij})}{\partial \nu_{ij}} \\ \frac{\partial \ell_1(\nu_{ij}, \boldsymbol{\gamma}, y_{ij})}{\partial \boldsymbol{\gamma}} \end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{d} \begin{pmatrix} \mathbf{x}_{ij}\mathbf{1}_q \\ \mathbf{I}_q \end{pmatrix} \frac{\partial \ell_{1ij}(\boldsymbol{\gamma}_{ij}, y_{ij})}{\partial \boldsymbol{\gamma}_{ij}} = \mathbf{0}, \tag{1}$$

where $\boldsymbol{\gamma}_{ij} = (\alpha_1 + \nu_{ij}, \ldots, \alpha_{K-1} + \nu_{ij}) = (\gamma_{ij1}, \ldots, \gamma_{ijK-1})$, $\ell_{1ij}(\cdot) = \log f_{1ij}(\cdot)$, $f_{1ij}(\boldsymbol{\gamma}_{ij}, y) = \mathcal{F}(\gamma_{ijy}) - \mathcal{F}(\gamma_{ijy-1})$, $\mathbf{I}_q$ is an identity matrix of dimension q and $\mathbf{1}_q$ is a unit vector of size q. Let $\mathbf{X}_{ij}^T = \begin{pmatrix} \mathbf{x}_{ij}\mathbf{1}_q \\ \mathbf{I}_q \end{pmatrix}$ and $\mathbf{s}_{ij}^{(1)}(\mathbf{a}) = \frac{\partial \ell_{1ij}(\boldsymbol{\gamma}_{ij}, y_{ij})}{\partial \boldsymbol{\gamma}_{ij}}$, where $\mathbf{a}^\top = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top)$ is the column vector of all $r = p + q$ univariate parameters. The score equations (1) can be written as

$$\mathbf{g}_1 = \mathbf{g}_1(\mathbf{a}) = \frac{\partial L_1}{\partial \mathbf{a}} = \sum_{i=1}^{n}\sum_{j=1}^{d} \mathbf{X}_{ij}^\top\ \mathbf{s}_{ij}^{(1)}(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^\top\ \mathbf{s}_i^{(1)}(\mathbf{a}) = \mathbf{0}, \tag{2}$$

where $\mathbf{X}_i^\top = (\mathbf{X}_{i1}^\top, \ldots, \mathbf{X}_{id}^\top)$ and $\mathbf{s}_i^{(1)\top}(\mathbf{a}) = (\mathbf{s}_{i1}^{(1)\top}(\mathbf{a}), \ldots, \mathbf{s}_{id}^{(1)\top}(\mathbf{a}))$. The vectors $\mathbf{s}_{ij}^{(1)}(\mathbf{a})$ and $\mathbf{s}_i^{(1)}(\mathbf{a})$ have lengths q and dq respectively. The dimensions of $\mathbf{X}_{ij}$ and $\mathbf{X}_i$ are $q \times r$ and $dq \times r$ respectively.

For estimation of $\mathbf{a}$ when $Y_{i1}, \ldots, Y_{id}$ are dependent, we construct the weighted scores estimating equations by inserting weight matrices between the matrix of covariates $\mathbf{X}_i^\top$ and the vector of scores for regression and non-regression parameters $\mathbf{s}_i^{(1)}(\mathbf{a})$. The weight matrices are based on a working MVN copula model with cdf:

$$F(y_1, \ldots, y_d; \nu_1, \ldots, \nu_d, \boldsymbol{\gamma}, \mathbf{R}) = \Phi_d\left(\Phi^{-1}[F_1(y_1; \nu_1, \boldsymbol{\gamma})], \ldots, \Phi^{-1}[F_1(y_d; \nu_d, \boldsymbol{\gamma})]; \mathbf{R}\right),$$

where Φ_d denotes the standard MVN distribution function with correlation matrix $\mathbf{R} = (\rho_{jk} : 1 \leq j < k \leq d)$, Φ is cdf of the univariate standard normal, and $F_1(y; \nu, \boldsymbol{\gamma}) = \mathcal{F}(\alpha_y + \nu)$ is the univariate cdf for Y. The MVN copula inherits the dependence structure of the MVN distribution, but lacks a closed form cdf; this means likelihood inference of the MVN copula for discrete data is not easy, because the MVN distribution, as a latent model for discrete response, requires rectangle probabilities based on high-dimensional integrations or their approximations (Nikoloulopoulos and Karlis, 2009; Nikoloulopoulos, 2013b). However, the weight matrices depend on the covariances of the scores, hence only the bivariate marginal probabilities of Y_{ij} and Y_{ik}, $j \neq k$ will be needed for estimation.

The weighted scores estimating equations take the form:

$$\mathbf{g}_1^\star = \mathbf{g}_1^\star(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^T \, \mathbf{W}_{i,\text{working}}^{-1} \, \mathbf{s}_i^{(1)}(\mathbf{a}) = \mathbf{0}, \tag{3}$$

where $\mathbf{W}_{i,\text{working}}^{-1} = \Delta_i^{(1)}(\widetilde{\mathbf{a}})[\boldsymbol{\Omega}_i^{(1)}(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}})]^{-1}$ is based on the covariance matrix $\boldsymbol{\Omega}_i^{(1)}(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}})$ of $\mathbf{s}_i^{(1)}(\mathbf{a})$ computed from the fitted MVN copula model with estimated parameters $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$ and the symmetric $dq \times dq$ matrix $\Delta_i^{(1)} = \text{diag}(\Delta_{i1}^{(1)}, \ldots, \Delta_{id}^{(1)})$ with $\Delta_{ij}^{(1)} = -\mathrm{E}\,(\frac{\partial^2 \ell_{1ij}(\boldsymbol{\gamma}_{ij}, y_{ij})}{\partial \boldsymbol{\gamma}_{ij} \partial \boldsymbol{\gamma}_{ij}^\top})$. As bivariate normal cdf calculations are needed for the calculation of $\boldsymbol{\Omega}_i^{(1)}(\widetilde{\mathbf{a}}, \widetilde{\mathbf{R}})$ (different ones for different clusters), a nearly accurate approximation that can be quickly computed is important. The approximation we use is provided by Johnson and Kotz (1972, page 118) and is described in an Appendix.

The estimated parameters $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$ of the working MVN copula model can be easily obtained in a two-step approach, namely the CL1 method in Zhao and Joe (2005). Estimated $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$ are obtained by solving the CL1 univariate and bivariate composite score functions, respectively. The former are the same as the independent estimating equations (2), while the latter are given below:

$$\mathbf{g}_2 = \sum_{i=1}^{n} \mathbf{s}_i^{(2)}(\widetilde{\mathbf{a}}, \mathbf{R}) = \mathbf{0},$$

where $\mathbf{s}_i^{(2)}(\mathbf{a}, \mathbf{R}) = \frac{\partial \sum_{j<k} \log f_2(y_{ij}, y_{ik}; \nu_{ij}, \nu_{ik}, \boldsymbol{\gamma}, \rho_{jk})}{\partial \mathbf{R}}$ with $f_2(\cdot)$ the bivariate marginal probability of Y_{ij} and Y_{ik}, viz.

$$f_2(y_{ij}, y_{ik}; \nu_{ij}, \nu_{ik}, \boldsymbol{\gamma}, \rho_{jk}) = \int_{\Phi^{-1}[F_1(y_{ij}-1;\nu_{ij},\boldsymbol{\gamma})]}^{\Phi^{-1}[F_1(y_{ij};\nu_{ij},\boldsymbol{\gamma})]} \int_{\Phi^{-1}[F_1(y_{ik}-1;\nu_{ik}),\boldsymbol{\gamma}]}^{\Phi^{-1}[F_1(y_{ik};\nu_{ik},\boldsymbol{\gamma})]} \phi_2(z_j, z_k; \rho_{jk}) dz_j dz_k;$$

$\phi_2(\cdot; \rho)$ denotes the standard bivariate normal density with correlation ρ.

If the $\mathbf{W}_{i,\text{working}}$ are assumed fixed for the second stage of solving the weighted scores estimating equations, then the asymptotic covariance matrix of the solution $\widehat{\mathbf{a}}$ of (3) is

$$\mathbf{V}_1^\star = (-\mathbf{H}_{\mathbf{g}_1^\star})^{-1} \mathbf{J}_{\mathbf{g}_1^\star} (-\mathbf{H}_{\mathbf{g}_1^\star}^T)^{-1}$$

with

$$-\mathbf{H}_{\mathbf{g}_1^\star} = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_{i,\text{working}}^{-1} \Delta_i^{(1)} \mathbf{X}_i, \;\; \mathbf{J}_{\mathbf{g}_1^\star} \;\; = \;\; \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_{i,\text{working}}^{-1} \boldsymbol{\Omega}_{i,\text{true}}^{(1)} (\mathbf{W}_{i,\text{working}}^{-1})^T \mathbf{X}_i,$$

where $\boldsymbol{\Omega}_{i,\text{true}}^{(1)}$ is the "true covariance matrix" of $\mathbf{s}_i^{(1)}(\mathbf{a})$. The $\boldsymbol{\Omega}_{i,true}^{(1)}$ can be estimated by $\mathbf{s}_i^{(1)}(\widehat{\mathbf{a}})\mathbf{s}_i^{(1)^\top}(\widehat{\mathbf{a}})$. This estimate is similar to what is done in the "sandwich" covariance estimator in GEE.

2.2 CL1 information criteria

The CL1 estimating equations in Zhao and Joe (2005) have been used in the preceding subsection to estimate conveniently the univariate and latent correlation parameters of the MVN copula model in order to compute the working weight matrices and then solve the weighted scores equations in (3). The acronym CL1 also denotes the composite likelihood information criteria (Nikoloulopoulos, 2016), that are based on the CL1 estimating equations, for correlation structure and variable selection in the weighted scores estimating equations. In line with Nikoloulopoulos (2016), in this subsection we propose the CL1 information criteria as an intermediate step for selecting both the covariates in the mean function modelling and the structure of the latent correlation matrix induced by the MVN latent variables for longitudinal ordinal data.

The CL1 versions of AIC and BIC criteria are defined as:

$$\text{CL1AIC} = -2L_2 + 2\text{tr}\Big(\mathbf{J_g}\mathbf{H_g}^{-1}\Big);$$

$$\text{CL1BIC} = -2L_2 + \log(n)\text{tr}\Big(\mathbf{J_g}\mathbf{H_g}^{-1}\Big),$$

where $L_2 = \sum_{i=1}^{n}\sum_{j<k}\log f_2(y_{ij}, y_{ik}; \nu_{ij}, \nu_{ik}, \boldsymbol{\gamma}, \rho_{jk})$ is the bivariate CL1 log-likelihood, $\mathbf{J_g}$ is the covariance or variability matrix, and $\mathbf{H_g}$ is the sensitivity or Hessian matrix of the CL1 estimating equations $\mathbf{g} = (\mathbf{g}_1, \mathbf{g}_2)^\top$. The form of $\mathbf{J_g}$ and $\mathbf{H_g}$ along with computational details are given in the Appendix.

3 Simulation study

In order to study the small-sample efficiency and robustness of the weighted scores estimating equations with a working MVN copula model, we will use various multivariate copula models as true models. We will compare the weighted scores estimating equations with the 'gold standard' maximum likelihood and also include in the comparison the local odds ratios GEE approach in Touloumis et al. (2013) as the current state of the art of the various GEE approaches for ordinal regression (Nooraee et al., 2014). We also assess the performance of the CL1 information criteria for longitudinal ordinal response data. Before that, the first subsection provides some background on copula models that might be suitable for clustered and longitudinal ordinal data.

3.1 Relevant background for copula models

A copula is a multivariate cdf with uniform $U(0,1)$ margins (Joe, 1997, 2014a; Nelsen, 2006). If $\mathcal{G}$ is a d-variate cdf with univariate margins $\mathcal{G}_1, \ldots, \mathcal{G}_d$, then Sklar's (1959) theorem implies that there is a copula C such that

$$\mathcal{G}(y_1, \ldots, y_d) = C\Big(\mathcal{G}_1(y_1), \ldots, \mathcal{G}_d(y_d)\Big).$$

Copulas enable you to break the model building process into two separate steps:

1. Choice of arbitrary marginal distributions $\mathcal{G}_1(y_1), \ldots, \mathcal{G}_d(y_d)$;
2. Choice of an arbitrary copula function C (dependence structure).

If one assumes a different copula, then a different multivariate distribution is constructed. The copula is unique if $\mathcal{G}_1, \ldots, \mathcal{G}_d$ are continuous, but not if some of the $\mathcal{G}_j$ have discrete components. If $\mathcal{G}$ is continuous and $(Y_1, \ldots, Y_d) \sim \mathcal{G}$, then the unique copula is the distribution of $(U_1, \ldots, U_d) = (\mathcal{G}_1(Y_1), \ldots, \mathcal{G}_d(Y_d))$ leading to

$$C(u_1, \ldots, u_d) = \mathcal{G}\Big(\mathcal{G}_1^{-1}(u_1), \ldots, \mathcal{G}_d^{-1}(u_d)\Big), \quad 0 \le u_j \le 1, j = 1, \ldots, d,$$

where $\mathcal{G}_j^{-1}$ are inverse cdfs. In particular, if $\mathcal{T}_d(\cdot;\nu,\mathbf{R})$ is the multivariate Student (MVT) cdf with correlation matrix $\mathbf{R} = (\rho_{jk} : 1 \le j < k \le d)$ and ν degrees of freedom, and $\mathcal{T}(\cdot;\nu)$ is the univariate Student t cdf with ν degrees of freedom, then the MVT copula is

$$C(u_1,\ldots,u_d) = \mathcal{T}_d\Big(\mathcal{T}^{-1}(u_1;\nu),\ldots,\mathcal{T}^{-1}(u_d;\nu);\nu,\mathbf{R}\Big).$$

For ordinal (discrete) random vectors, multivariate probabilities of the form $\pi_d(\mathbf{y}) = \Pr(Y_1 = y_1,\ldots,Y_d = y_d)$ involve 2^d finite differences of the joint cdf. Therefore likelihood inference for discrete data is straightforward for copulas with a computationally feasible form of the cdf. Archimedean (Joe, 1997) and mixtures of max-infinitely divisible (max-id) (Joe and Hu, 1996) copulas have closed form cdfs but have less range of dependence compared with the MVN or MVT copulas. However they provide enough structure to study the efficiency of the weighted scores method using the discretized MVN as a "working model". For example, the Archimedean copula is suitable for positive dependent clustered data with exchangeable dependence, while the mixture of max-id copula is suitable for more general positive dependence, including dependence that is decreasing with lag as in longitudinal data. More importantly, these copulas have different dependence properties than the "working" MVN copula. For example they provide reflection asymmetric tail dependence, while the MVN copula provides tail independence (Joe, 1997, 2014a). Hence they are suitable to study the robustness to dependence of the weighted scores method. These parametric families of copulas are briefly defined below:

- Multivariate Archimedean copulas have the form

$$C(u_1,\ldots,u_d\,;\,\theta) = \phi\left(\sum_{j=1}^{d}\phi^{-1}(u_j\,;\,\theta)\,;\,\theta\right),$$

  where $\phi(u\,;\,\theta)$ is the Laplace transform of a univariate family of distributions of positive random variables indexed by the parameter θ, such that $\phi(\cdot;\theta)$ and its inverse has closed form (Joe, 1997).

- Mixture of max-id copulas (Joe and Hu, 1996) have the form

$$C(u_1,\ldots,u_d;\theta,\theta_{jk} : 1 \le j < k \le d) = \phi\left(-\sum_{j<k}\log C_{jk}^{(m)}\big(e^{-\phi^{-1}(u_j;\theta)/(d-1)}, e^{-\phi^{-1}(u_k;\theta)/(d-1)};\theta_{jk}\big)\,;\,\theta\right).$$

  Since the mixing operation introduces dependence this new copula has a dependence structure that comes from the form of the max-id copula $C_{jk}^{(m)}(\cdot\,;\,\theta_{jk})$ and the form of Laplace transform $\phi(\cdot\,;\,\theta)$. Another interesting interpretation is that the Laplace transform ϕ introduces the smallest dependence between random variables (exchangeable dependence), while the copulas $C_{jk}^{(m)}$ add some pairwise dependence.

We consider a multivariate ordinal regression setting in which the $d \ge 2$ dependent ordinal variables $Y_1,\ldots,Y_d$ are observed together with a vector $\mathbf{x} \in \mathbb{R}^p$ of explanatory variables. If $C(\cdot)$ is any parametric family of copulas and $F_1(y_j,\nu_j,\boldsymbol{\gamma})$ is the parametric model for the jth univariate ordinal variable then

$$C\Big(F_1(y_1,\nu_1,\boldsymbol{\gamma}),\ldots,F_1(y_d,\nu_d,\boldsymbol{\gamma})\Big)$$

is a multivariate parametric model with univariate margins $F_1(y_1,\nu_1,\boldsymbol{\gamma}),\ldots,F_1(y_d,\nu_d,\boldsymbol{\gamma})$. For copula models, the response vector $\mathbf{Y} = (Y_1,\ldots,Y_d)$ can be discrete (Nikoloulopoulos, 2013a; Nikoloulopoulos and Joe, 2015).

3.2 Small sample efficiency and robustness of the weighted scores

We randomly generate $B = 10^4$ samples of size $n = 50, 100, 300$ from the above copula models with exchangeable and unstructured dependence. Note that AR(1)-like dependence is not used here since the local odds ratios GEE method in Touloumis et al. (2013) does not include this structure. For exchangeable dependence structure, the Gumbel copula in the Archimedean class with Laplace transform $\phi_G(t;\theta) = \exp(-t^{1/\theta})$ was used as the "true model". For unstructured dependence, the mixture of max-id copula with Laplace transform $\phi_G(\cdot;\theta)$ and the bivariate Gumbel copula for the $C_{jk}^{(m)}(\cdot;\theta_{jk})$ was used as the "true model". For simulation from Archimedean and mixture of max-id copulas we have used the algorithms in Joe (2014a, pp. 272–274). More specifically we used the functions `rmgum` and `rtmm1` in the R package **CopulaModel** (Joe, 2014b) to simulate from an Archimedean and mixture of max-id copula, respectively.

We use $d = 3$ and $K = 5$ ordinal categories (equally weighted) to imitate the real data application considered in Section 4. For the covariates and ordinal probit regression parameters, we use a combination of a time-stationary and a time-varying design, i.e., include covariates that are typically constant over time, and correlated over time. More specifically, we chose $p = 4, \mathbf{x}_{ij} = (x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij})$ with $x_{1ij} \in \{0, 1\}$ a group variable, x_{2ij} an i.i.d. from a trivariate Gumbel copula with dependence parameter $\theta = 2$, i.e., $d \times d$ Kendall's tau association matrix with off-diagonal elements equal to 0.5, and standard uniform margins, $x_{3ij} = x_{1ij} \times x_{2ij}$, and x_{4ij} a uniform random variable in the interval $[-1, 1]$; $\beta_1 = -\beta_2 = -\beta_3 = -0.5, \beta_4 = 0$. By considering the noise variable x_{4ij} we aim to check the Type I error rate for inference on $H_0 : \beta_4 = 0$ (see e.g., Larrabee et al., 2014) based on the weighted scores, local odds ratios GEE and ML methods.

Table 1 contains the parameter values, the bias, standard deviations (SD) and root mean square errors (RMSE) of the maximum likelihood (ML), weighted scores (WS) and GEE estimates, along with the average of their theoretical SDs (denoted with $\sqrt{\overline{V}}$). Note in passing, for the weighted scores estimates, V is the diagonal of $\mathbf{V}_1^\star$. The theoretical variance of the ML estimate is obtained via the gradients and the Hessian computed numerically during the maximization process. The GEE estimates and their theoretical variance are calculated with the function `ordLORgeeR` in the R package **multgee** (Touloumis, 2015). For the local odds ratios GEE approach we use the 'uniform' and the 'category exchangeability' structure for the exchangeable and unstructured case, respectively, as suggested by Touloumis et al. (2013).

From the results, it is apparent there is additional small-sample bias for the weighted scores method when n is small, say $n <= 50$, but it decreases and becomes negligible as the sample size n increases, say $n >= 300$. Nevertheless, biases, SDs and RMSEs do not differ significantly among the competing methods. Hence, we can conclude that the weighted scores and the local odds ratios GEE method are robust to dependence and nearly as efficient as maximum likelihood for fully specified copula models. This agrees with the conclusions of Nikoloulopoulos et al. (2011) and Touloumis et al. (2013) for the weighted scores and local odds ratios approach, respectively. When the MVN copula is used as the working model, the weighted scores estimating equations yield consistent estimates when the latent correlation parameters is such that the MVN copula model is quite close to the true copula model in Kullback-Leibler distance. This is elucidated further in Nikoloulopoulos et al. (2011) where theoretical efficiency and robustness studies have also been performed for a negative binomial margin. Same results are to be expected for any other margin.

Furthermore, Table 2 contains the observed level of the two-sided test for three common nominal levels for inference on $H_0 : \beta_4 = 0$ based on the weighted scores, local odds ratios GEE and ML methods.The observed

Table 1: Small sample of sizes $n = 50, 100, 300$ simulations (10^4 replications) and resultant biases, root mean square errors (RMSE), and standard deviations (SD), along with average theoretical SDs ($\sqrt{\bar{V}}$), for the ML estimates of the regression parameters for the trivariate Gumbel copula (exchangeable) or the mixture of max-id copula with Laplace transform ϕ_G and the bivariate Gumbel copula for the $C_{jk}^{(m)}(\cdot;\theta_{jk})$ (unstructured) model and ordinal probit regression, and the weighted scores (WS) and local odds ratios GEE with an exchangeable or unstructured correlation matrix.

			$\beta_1 = -0.5$			$\beta_2 = 0.5$			$\beta_3 = 0.5$			$\beta_4 = 0$		
			WS	GEE	ML	WS	GEE	ML	WS	GEE	ML	WS	GEE	ML
Exch	$n = 50$	Bias	-0.026	-0.015	-0.019	0.031	0.014	0.018	0.032	0.017	0.020	-0.001	-0.001	0.000
$\theta = 3$		SD	0.418	0.409	0.387	0.405	0.392	0.369	0.580	0.562	0.530	0.084	0.083	0.078
		RMSE	0.418	0.409	0.388	0.406	0.392	0.370	0.581	0.562	0.531	0.084	0.083	0.078
		$\sqrt{\bar{V}}$	0.394	0.386	0.375	0.373	0.362	0.356	0.534	0.520	0.513	0.083	0.082	0.078
	$n = 100$	Bias	-0.014	-0.008	-0.008	0.015	0.007	0.008	0.016	0.008	0.011	0.000	0.000	0.000
		SD	0.284	0.283	0.265	0.271	0.267	0.250	0.388	0.384	0.358	0.058	0.058	0.054
		RMSE	0.284	0.283	0.266	0.271	0.268	0.250	0.388	0.384	0.359	0.058	0.058	0.054
		$\sqrt{\bar{V}}$	0.273	0.271	0.260	0.256	0.254	0.244	0.368	0.365	0.350	0.057	0.057	0.053
	$n = 300$	Bias	-0.004	-0.001	-0.001	0.007	0.004	0.006	0.003	0.000	-0.001	-0.001	-0.001	-0.001
		SD	0.157	0.158	0.148	0.149	0.150	0.140	0.212	0.214	0.199	0.032	0.032	0.030
		RMSE	0.157	0.158	0.148	0.149	0.150	0.140	0.212	0.214	0.199	0.032	0.032	0.030
		$\sqrt{\bar{V}}$	0.156	0.156	0.148	0.145	0.146	0.138	0.209	0.210	0.198	0.032	0.032	0.030
Exch	$n = 50$	Bias	-0.032	-0.017	-0.018	0.033	0.013	0.016	0.042	0.023	0.024	-0.001	-0.001	-0.001
$\theta = 5$		SD	0.391	0.376	0.361	0.323	0.303	0.290	0.460	0.431	0.413	0.062	0.060	0.057
		RMSE	0.392	0.376	0.361	0.325	0.303	0.290	0.462	0.432	0.414	0.062	0.060	0.057
		$\sqrt{\bar{V}}$	0.370	0.354	0.351	0.295	0.274	0.279	0.424	0.392	0.404	0.061	0.058	0.056
	$n = 100$	Bias	-0.015	-0.008	-0.007	0.020	0.010	0.011	0.017	0.009	0.010	0.000	0.000	0.000
		SD	0.262	0.258	0.247	0.209	0.203	0.193	0.299	0.290	0.276	0.041	0.041	0.039
		RMSE	0.262	0.258	0.247	0.210	0.203	0.193	0.300	0.290	0.276	0.041	0.041	0.039
		$\sqrt{\bar{V}}$	0.255	0.249	0.243	0.199	0.192	0.190	0.285	0.275	0.273	0.041	0.040	0.039
	$n = 300$	Bias	-0.005	-0.002	-0.002	0.006	0.003	0.004	0.004	0.001	0.001	0.000	0.000	0.000
		SD	0.146	0.146	0.139	0.113	0.113	0.107	0.163	0.162	0.154	0.023	0.023	0.022
		RMSE	0.146	0.146	0.139	0.113	0.113	0.108	0.163	0.162	0.154	0.023	0.023	0.022
		$\sqrt{\bar{V}}$	0.144	0.144	0.138	0.111	0.111	0.107	0.160	0.159	0.154	0.023	0.023	0.022
Unstr	$n = 50$	Bias	-0.021	-0.009	-0.019	0.023	0.004	0.028	0.021	0.002	0.017	0.000	0.000	0.001
		SD	0.433	0.423	0.430	0.507	0.489	0.501	0.716	0.692	0.711	0.141	0.135	0.138
		RMSE	0.434	0.423	0.430	0.508	0.489	0.501	0.716	0.692	0.711	0.141	0.135	0.138
		$\sqrt{\bar{V}}$	0.406	0.405	0.415	0.471	0.470	0.480	0.667	0.666	0.685	0.131	0.130	0.129
$\theta = 1.2$	$n = 100$	Bias	-0.007	-0.001	-0.005	0.016	0.007	0.020	0.007	-0.003	0.002	0.000	0.000	-0.001
$\theta_{12} = 1.5$		SD	0.298	0.295	0.295	0.348	0.342	0.340	0.490	0.482	0.482	0.094	0.092	0.090
$\theta_{13} = 1.1$		RMSE	0.299	0.295	0.295	0.348	0.342	0.340	0.490	0.482	0.482	0.094	0.092	0.090
$\theta_{23} = 2.7$		$\sqrt{\bar{V}}$	0.287	0.287	0.284	0.333	0.333	0.326	0.472	0.472	0.467	0.092	0.092	0.089
	$n = 300$	Bias	-0.005	-0.003	-0.004	0.002	-0.001	0.004	0.007	0.004	0.005	0.000	0.000	0.000
		SD	0.169	0.169	0.164	0.196	0.195	0.187	0.276	0.275	0.265	0.053	0.053	0.050
		RMSE	0.169	0.169	0.164	0.196	0.195	0.187	0.276	0.275	0.265	0.053	0.053	0.050
		$\sqrt{\bar{V}}$	0.166	0.166	0.161	0.193	0.192	0.184	0.273	0.273	0.263	0.053	0.053	0.050

Table 2: Empirical Type I error rates for inference on $H_0 : \beta_4 = 0$ based on the weighted scores (WS), local odds ratios GEE and ML methods.

		$\alpha = 0.01$ †			$\alpha = 0.05$			$\alpha = 0.1$		
		WS	GEE	ML	WS	GEE	ML	WS	GEE	ML
Exch, $\theta = 3$	$n = 50$	0.014	0.014	0.008	0.060	0.063	0.048	0.112	0.114	0.097
	$n = 100$	0.013	0.013	0.010	0.057	0.056	0.053	0.114	0.112	0.105
	$n = 300$	0.011	0.011	0.010	0.050	0.050	0.051	0.101	0.101	0.099
Exch, $\theta = 5$	$n = 50$	0.015	0.019	0.008	0.059	0.068	0.049	0.113	0.120	0.102
	$n = 100$	0.013	0.013	0.011	0.053	0.060	0.050	0.106	0.110	0.099
	$n = 300$	0.012	0.012	0.010	0.054	0.053	0.051	0.105	0.107	0.102
Unstr §	$n = 50$	0.020	0.015	0.024	0.072	0.060	0.068	0.132	0.117	0.121
	$n = 100$	0.012	0.011	0.014	0.056	0.050	0.056	0.111	0.102	0.110
	$n = 300$	0.010	0.009	0.009	0.050	0.049	0.051	0.101	0.099	0.105

†: α denotes the significance level; §: the true copula parameters are $\{\theta, \theta_{12}, \theta_{13}, \theta_{23}\} = \{1.2, 1.5, 1.1, 2.7\}$.

levels are close to nominal levels and hence demonstrate that the tests from all the competing approaches are reliable.

Finally in order to study the relative performance of the weighted scores over the local odds radios GEE method as the dimension d or the number of categories increase we randomly generated $B = 20$ samples of size $n = 100$ from the Gumbel copula model with exchangeable dependence for $d, K \in \{10, 15, 20, 25\}$. The link function, model parameters and covariates are set as before. The simulations were carried out on an Intel(R) Xeon(R) CPU X5650 2.67GHz.

Table 3 summarizes the computing times (averaged over 20 replications) in seconds. Clearly the local odds ratios GEE approach requires a much higher computing time for large d or K. Note in passing that for large d or K memory up to 60GB was required for the local odds ratios GEE approach. Hence it is demonstrated that large matrices are a problem for GEE methods, in as much as they cause model fitting algorithms to run slowly. Note in passing that for larger (than the ones in Table 3) values of K or d the local odds ratios GEE implementation (Touloumis, 2015) is infeasible.

3.3 Model selection criteria

We perform simulation studies to examine the reliability of using CL1AIC and CL1BIC for latent correlation structure and covariate selection in Subsection 3.3.1 and Subsection 3.3.2, respectively. For exchangeable, AR(1), and unstructured dependence, the Gumbel copula, the mixture of max-id copula with Laplace transform $\phi_G(\cdot;\theta)$ and the bivariate Gumbel copula for the $C_{jk}^{(m)}(\cdot;\theta_{jk})$, and the MVT copula were used as the "true models", respectively.

3.3.1 Correlation structure selection

We randomly generate $B = 10^3$ samples of size $n = 50, 100, 300$ with $d = 3$ and ordinal probit regression with $p = 3, \mathbf{x}_{ij} = (1, x_{1ij}, j-1)^T$ where x_{1ij} are taken as Bernoulli random variables with probability of success $1/2$, and $\beta_0 = 0.25 = -\beta_1 = -\beta_2$.

In Table 4, we present the number of times that different working correlation structures are chosen over 1000 simulation runs under each true correlation structure. If the true correlation structure is exchangeable or AR(1),

Table 3: Small sample of size $n = 100$ simulations (20 replications) from the Gumbel copula model with exchangeable dependence for $d, K \in \{10, 15, 20, 25\}$ and computing times (averaged over 20 replications) in seconds of the weighted scores (WS) over the local odds ratios GEE approach.

d	K	θ	time(WS)	time(GEE)	$\frac{\text{time(GEE)}}{\text{time(WS)}}$
10	10	3	88.1	35.3	0.4
		5	95.4	69.7	0.7
10	15	3	184.2	126.7	0.7
		5	193.9	182.2	0.9
10	20	3	363.2	444.9	1.2
		5	422.8	640.0	1.5
10	25	3	760.8	1559.4	2.0
		5	716.1	1490.9	2.1
15	10	3	208.4	254.1	1.2
		5	219.6	388.2	1.8
15	15	3	350.4	1267.6	3.6
		5	371.1	1566.2	4.2
15	20	3	636.8	6057.0	9.5
		5	706.1	6738.5	9.5
15	25	5	1524.1	23221.8	15.2
		3	1380.5	25139.2	18.2
20	10	3	334.4	1289.0	3.9
		5	360.7	1646.5	4.6
20	15	3	562.4	8165.8	14.5
		5	602.9	9196.7	15.3
20	20	3	1287.1	48270.2	37.5
		5	1268.0	50030.8	39.5
20	25	3	2131.3	153707.2	72.1
		5	2243.1	138304.0	61.7

CL1BIC is better than CL1AIC. If the true correlation structure is unstructured, CL1AIC performs extremely well, especially for a small sample size, which is typical of medical studies. The difference between the correct identification rate of CL1AIC and that of CL1BIC becomes small when the sample size increases to 100 or 300. The CL1AIC tends to choose the unstructured correlation structure more often than CL1BIC does, since AIC is more likely to result in an overparametrized model than BIC in parametric settings (Chen and Lazar, 2012; Nikoloulopoulos, 2016).

3.3.2 Variable selection

We randomly generate $B = 10^3$ samples of size $n = 50, 100, 300$ with $d = 3$ and ordinal probit regression with $p = 5, \mathbf{x}_{ij} = (1, x_{1ij}, j-1, x_{3ij}, x_{4ij})^\top$ where $x_{1ij}, \beta_0, \beta_1, \beta_2$ are as before, x_{3ij}, x_{4ij} are independent uniform random variables in the interval $[-1, 1]$ (and independent of x_{1ij}), and $\beta_3 = \beta_4 = 0$. We consider the same candidate models, with various subsets of covariates, and include all the aforementioned parametric correlation structures as true correlation structures. The subsets of covariates that we consider are the following:

- $\mathbf{x}_1 = (1, x_{1ij})^\top$.
- $\mathbf{x}_{12} = (1, x_{1ij}, j-1)^\top$ (the true subset of covariates).
- $\mathbf{x}_{123} = (1, x_{1ij}, j-1, x_{3ij})^\top$.

Table 4: Frequencies of the correlation structure and the set of the variables identified using CL1AIC and CL1BIC from 1000 simulation runs in each setting. The first column indicates the true correlation structure. The number of times that the true correlation structure and subset of covariates are chosen by each criterion are bold faced.

Latent correlation structure selection										
		$n = 50$			$n = 100$			$n = 300$		
		Exch	AR(1)	Unstr	Exch	AR(1)	Unstr	Exch	AR(1)	Unstr
Exch	CL1AIC	**732**	123	145	**779**	63	158	**814**	3	183
$\theta = 2$	CL1BIC	**838**	129	33	**907**	71	22	**979**	13	8
AR(1)	CL1AIC	117	**727**	156	34	**794**	172	0	**821**	179
	CL1BIC	136	**821**	43	40	**927**	33	4	**981**	15
Unstr	CL1AIC	64	13	**923**	64	13	**923**	0	0	**1000**
	CL1BIC	249	34	**717**	51	4	**945**	0	0	**1000**

Variable selection													
		$n = 50$				$n = 100$				$n = 300$			
		$\mathbf{x}_1$	$\mathbf{x}_{12}$	$\mathbf{x}_{123}$	$\mathbf{x}_{1234}$	$\mathbf{x}_1$	$\mathbf{x}_{12}$	$\mathbf{x}_{123}$	$\mathbf{x}_{1234}$	$\mathbf{x}_1$	$\mathbf{x}_{12}$	$\mathbf{x}_{123}$	$\mathbf{x}_{1234}$
Exch	CL1AIC	1	**675**	162	162	0	**701**	152	147	0	**692**	173	135
	CL1BIC	3	**862**	88	47	0	**884**	84	32	0	**939**	49	12
AR(1)	CL1AIC	3	**653**	183	161	0	**667**	189	144	0	**673**	168	159
	CL1BIC	22	**819**	107	52	2	**876**	87	35	0	**924**	55	21
Unstr	CL1AIC	93	**512**	191	204	93	**512**	191	204	0	**599**	199	202
	CL1BIC	230	**586**	116	68	62	**801**	85	52	0	**902**	76	22

For exchangeable, AR(1), and unstructured dependence the true parameters are $\theta = 2$, $\{\theta, \theta_{12}, \theta_{13}, \theta_{23}\} = \{1.5, 4, 1, 4\}$, and $\{\rho_{12}, \rho_{13}, \rho_{23}, \nu\} = \{-0.5, -0.3, 0.3, 5\}$, respectively; $\mathbf{x}_1 = (1, x_{1ij})^\top$, $\mathbf{x}_{12} = (1, x_{1ij}, j-1)^\top$ (the true subset of covariates), $\mathbf{x}_{123} = (1, x_{1ij}, j-1, x_{3ij})^\top$, and $\mathbf{x}_{1234} = (1, x_{1ij}, j-1, x_{3ij}, x_{4ij})^\top$.

- $\mathbf{x}_{1234} = (1, x_{1ij}, j-1, x_{3ij}, x_{4ij})^\top$.

In Table 4, we present the number of times that different subsets of covariates are chosen over 1000 simulation runs under each true correlation structure. For all the true correlation structures, CL1BIC performs better than CL1AIC, and its performance improves as the sample size increases.

4 The rheumatoid arthritis data

We illustrate the weighted scores estimating equations and composite likelihood information criteria by analysing the rheumatoid arthritis data-set (Bombardier et al., 1986). These data have previously been used as an example for other methodological papers on GEE for ordinal regression (Ware and Lipsitz, 1986; Lipsitz et al., 1994b; Touloumis et al., 2013). The data were taken from a randomized clinical trial designed to evaluate the effectiveness of the treatment Auranofin versus a placebo therapy for the treatment of rheumatoid arthritis. The repeated ordinal response is the self-assessment of arthritis, classified on a five-level ordinal scale ($1 =$ poor, $\ldots, 5 =$ very good). Patients ($n = 303$) were randomized into one of the two treatment groups after baseline self-assessment followed during five months of treatment with measurements taken at the first month and every two months during treatment resulting in a maximum of 3 measurements per subject (unequal cluster sizes). The covariates are time, baseline-assessment, age in years at baseline, sex and treatment. We treat time and baseline-assessment as categorical variables following Touloumis et al. (2013). However, instead of testing for differences to the reference outcome category we look at differences between adjacent outcome categories (see, e.g., Tutz and Gertheiss, 2016). To this end we followed the coding scheme for ordinal independent variables in Walter et al. (1987).

To select the appropriate correlation structure, we use the proposed model selection criteria in the weighted scores estimating equations, based on the full model with all covariates (Table 5, correlation structure selection). Further, both logit and probit links are used for the ordinal regressions. According to CL1AIC the selected correlation structure is the unstructured, while according to the CL1BIC, it is exchangeable. In this example we will prefer CL1BIC since one can easily distinguish between the various structures, as their difference in magnitude is large. This is not the case for the CL1AIC, where the differences are rather small. This was also the finding in our simulation studies, where it has been revealed that CL1AIC is more prone to select the unstructured case. Further, ordinal logistic regression is slightly better than ordinal probit regression based on CL1BIC.

Under the preferred exchangeable structure, we fit different models with different subsets of covariates, and find that the model with time, baseline-assessment, treatment and age, has the smallest CL1AIC and CL1BIC. Note that in Touloumis et al. (2013) age and sex have not been considered.

Table 5: The values of the different criteria for correlation structure selection at the full model and variable selection for the exchangeable structure for the arthritis data. The smallest value of each criterion is boldfaced.

Link	Probit		Logit	
Correlation structure selection				
	CL1AIC	CL1BIC	CL1AIC	CL1BIC
Exchangeable	4280.92	**4357.81**	4275.09	**4351.41**
AR(1)	4298.97	4374.26	4292.42	4367.20
Unstructured	**4279.97**	4362.37	**4273.87**	4355.72
Variable selection				
time trt baseline age sex	4280.92	4357.81	4275.09	4351.41
time trt baseline age	**4277.91**	4348.03	**4273.14**	**4342.77**
time trt baseline sex	4287.24	4357.54	4282.32	4352.04
time trt baseline	4284.22	**4347.71**	4279.78	4342.81
trt baseline	4305.09	4363.89	4298.98	4357.27
time trt	4491.15	4529.31	4497.76	4535.93
time	4515.26	4546.00	4517.43	4548.24
trt	4511.37	4545.76	4517.11	4551.46

Finally, Table 6 gives the estimates and standard errors of the model parameters obtained using the weighted scores estimating equations and local odds ratio GEE under the optimal exchangeable correlation structure, set of covariates, and logit link. Clearly a "true" copula model cannot be known for this (or any other) example, hence copula models (e.g., Gumbel, mixture of max-id, elliptical) are not assumed and used for ML estimation. This is precisely a reason why our method is superior compared with the ML method. Our estimating equations based on a "working" MVN copula-based model are robust, while on the other hand ML estimates could be biased if the univariate model is correct but dependence is modelled incorrectly. The goal of this paper is not to compare copula models for a best fit, as that type of research has already been done elsewhere; see e.g., Nikoloulopoulos and Karlis (2008).

Our analysis shows that the estimates of all the parameters and their corresponding standard errors obtained from the weighted scores method are nearly the same as those obtained from the local odds ratios GEE approach. In fact, the columns of p-values for the two methods agree very closely and the same factors are found to be significant or insignificant for $\alpha = 10\%$. This example also shows that if the correlation structure and the variables in the mean function modelling are correctly specified, then there is no loss in efficiency in GEE.

Table 6: Weighted scores and GEE estimates (Est.), along with their standard errors (SE) under the optimal correlation structure and set of covariates for the arthritis data.

	Weighted scores				GEE			
	Est.	SE	Z	p-value	Est.	SE	Z	p-value
α_1	-2.050	0.638	-3.215	0.001	-2.081	0.637	-3.268	0.001
α_2	0.058	0.607	0.096	0.924	0.028	0.606	0.046	0.963
α_3	2.021	0.612	3.305	0.001	1.994	0.610	3.268	0.001
α_4	4.329	0.653	6.634	< 0.001	4.307	0.650	6.625	< 0.001
$I(\text{time} = 2, 3)$	-0.007	0.121	-0.059	0.953	0.003	0.122	0.021	0.984
$I(\text{time} = 3)$	-0.370	0.113	-3.267	0.001	-0.365	0.113	-3.220	0.001
trt	-0.511	0.168	-3.037	0.002	-0.507	0.168	-3.023	0.003
$I(\text{baseline} = 2, 3, 4, 5)$	-0.620	0.380	-1.631	0.103	-0.650	0.380	-1.710	0.087
$I(\text{baseline} = 3, 4, 5)$	-0.567	0.226	-2.510	0.012	-0.548	0.227	-2.418	0.016
$I(\text{baseline} = 4, 5)$	-1.369	0.236	-5.790	< 0.001	-1.395	0.236	-5.921	< 0.001
$I(\text{baseline} = 5)$	-1.417	0.403	-3.519	< 0.001	-1.389	0.406	-3.424	0.001
age	0.013	0.008	1.656	0.098	0.014	0.008	1.736	0.083

In fact, if a 'time exchangeability' or a homogenous Goodman's row and column effects ('RC') structure is assumed in the local odds ratios GEE approach the age effect is statistically insignificant for $\alpha = 10\%$ (results available upon request).

5 Discussion

In this article, we have constructed weighted scores estimating equations using a working MVN copula model with univariate ordinal probit or logit regressions as the marginals. Since the univariate margins are regression models for ordinal response, copula models are more difficult to discriminate (Joe, 2014a, page 242). Hence, the advantage of the MVN copula is that does not only provide a wide range of flexible dependence but also approximates other copula models. By using a working MVN copula model with univariate ordinal probit or logit regressions as the marginals, the resulting weight matrices should be close to optimal. That is, assuming the choice of dependence structure (e.g., exchangeable, AR or unstructured) is acceptable, the correlation parameters used in the weight matrices are good choices when the working model is a good approximation.

The current state of the art of the various GEE approaches for ordinal regression is the local odds ratios GEE proposed by Touloumis et al. (2013). The local odds ratios GEE stemmed from recent criticism of the GEE (Crowder, 1995; Chaganty and Joe, 2004, 2006). Similarly to the local odds ratio GEE, the weighted scores estimating equations overcome the GEE pitfalls pointed out by the aforementioned papers. The weighted scores method is based on a working MVN copula model, that is actually a proper multivariate model, and the nuisance parameters are interpretable as latent or polychoric (Olsson, 1979) correlation parameters that are not restricted by the marginal model specification. Larger absolute values of the latent correlation parameters indicate stronger dependence regardless of the response scale or marginal model. The latent correlation matrix induced by the MVN has a proper definition and relates to the probability distribution of the response vector, thus causing no breakdown of the asymptotic properties of the estimation procedure. With this formal definition, the latent correlation matrix has direct mathematical relationship to the covariance of the response vector, and in the presence of a proper underlying probability distribution assertions of consistency are valid; law of large numbers assumes that there is an underlying probability distribution (Lee and Nelder, 2009).

In an extensive simulation study, we have compared the proposed equations with the local odds ratios GEE. We have demonstrated that both approaches are nearly as efficient as maximum likelihood and have the merit of robustness to different dependence models if the variable selection in the mean function modelling and the structured correlation matrix are correctly specified. With the composite likelihood function available as an intermediate step, we have further proposed the CL1 information criteria for selecting the structure of the latent correlation matrix and the explanatory variables. The local odds ratios GEE approach has no likelihood function available, hence there is no means of comparing the relative fit of competing parameter specifications. Therefore, an advantage of our framework is the variable/correlation structure selection using the CL1 information criteria, which are well-grounded in likelihood theory (Varin et al., 2011) and cannot be used in GEE methods, which are based on moments with no defined likelihood (Lindsey and Lambert, 1998). We have further shown that large matrices are a problem for local odds ratio GEE, in as much as they cause model fitting algorithms to run slowly.

It is worth mentioning that the range of possible applications of our methods goes beyond biometric, disease or health data and is of interest also in other fields such as finance. A typical example are panel data of corporate credit ratings (e.g., Hirk et al. 2019). A credit rating agency such as the Standard & Poor's assigns ratings on an ordinal scale with a considerable large number of categories and the dimension of the cluster is usually large.

Software

R functions to implement the weighted scores method and the CL1 information criteria for longitudinal ordinal data have been implemented in the package **weightedScores** (Nikoloulopoulos and Joe, 2020) within the open source statistical environment R (Team, 2020). The data and code used in Section 4 are given as data and code examples in the package, respectively.

Appendix

The approximation for the bivariate normal cdf in Johnson and Kotz (1972)

The following approximation for the bivariate normal cdf is from Johnson and Kotz (1972, page 118). Let $\Phi_2(x_1, x_2; \rho) = \Pr(Z_1 \leq x_1,\, Z_2 \leq x_2)$, where (Z_1, Z_2) is bivariate normal with means 0, variances 1 and correlation ρ. Let Φ, ϕ be the univariate standard normal cdf and density respectively. Then an expansion, due to Pearson (1901), is

$$\Phi_2(x_1, x_2; \rho) \;=\; \Phi(x_1)\Phi(x_2) + \phi(x_1)\phi(x_2) \sum_{j=1}^{\infty} \frac{\rho^j}{j!}\, \psi_j(x_1)\psi_j(x_2)$$

where

$$\psi_j(z) = (-1)^{j-1} \frac{d^{j-1}\phi(z)}{dz^{j-1}}.$$

Since

$$\begin{aligned}
\phi'(z) &= -z\phi(z) \\
\phi''(z) &= (z^2 - 1)\phi(z) \\
\phi'''(z) &= [2z - z(z^2 - 1)]\phi(z) = (3z - z^3)\phi(z) \\
\phi^{(4)}(z) &= [3 - 3z^2 - z(3z - z^3)]\phi(z) = (3 - 6z^2 + z^4)\phi(z)
\end{aligned}$$

we have

$$\begin{aligned}\Phi_2(x_1,x_2;\rho) &= \Phi(x_1)\Phi(x_2)+\phi(x_1)\phi(x_2)\left[\rho+\rho^2x_1x_2/2\right.\\ &\quad +\rho^3(x_1^2-1)(x_2^2-1)/6\\ &\quad +\rho^4(x_1^3-3x_1)(x_2^3-3x_2)/24\\ &\quad \left.+\rho^5(x_1^4-6x_1^2+3)(x_2^4-6x_2^2+3)/120+\cdots\right]\end{aligned}$$

A nearly accurate approximation is obtained truncating the series at ρ^3 term for $|\rho|\leq 0.4$, and at ρ^5 term for $0.4<|\rho|\leq 0.7$. Higher order terms may be required for $|\rho|>0.7$.

Details in the calculation of the matrices that form the CL1 criteria

The covariance matrix $\mathbf{J_g}$ of the composite score functions $\mathbf{g}$ is given as below

$$\mathbf{J_g}=\mathrm{Cov}(\mathbf{g})=\begin{pmatrix}\mathrm{Cov}(\mathbf{g}_1) & \mathrm{Cov}(\mathbf{g}_1,\mathbf{g}_2)\\ \mathrm{Cov}(\mathbf{g}_2,\mathbf{g}_1) & \mathrm{Cov}(\mathbf{g}_2)\end{pmatrix}=\frac{1}{n}\sum_i\begin{pmatrix}\mathbf{X}_i^\top\boldsymbol{\Omega}_i^{(1)}\mathbf{X}_i & \mathbf{X}_i^\top\boldsymbol{\Omega}_i^{(1,2)}\\ \boldsymbol{\Omega}_i^{(2,1)}\mathbf{X}_i & \boldsymbol{\Omega}_i^{(2)}\end{pmatrix},$$

where

$$\begin{pmatrix}\boldsymbol{\Omega}_i^{(1)} & \boldsymbol{\Omega}_i^{(1,2)}\\ \boldsymbol{\Omega}_i^{(2,1)} & \boldsymbol{\Omega}_i^{(2)}\end{pmatrix}=\begin{pmatrix}\mathrm{Cov}\Big(\mathbf{s}_i^{(1)}(\mathbf{a})\Big) & \mathrm{Cov}\Big(\mathbf{s}_i^{(1)}(\mathbf{a}),\mathbf{s}_i^{(2)}(\mathbf{a},\mathbf{R})\Big)\\ \mathrm{Cov}\Big(\mathbf{s}_i^{(2)}(\mathbf{a},\mathbf{R}),\mathbf{s}_i^{(1)}(\mathbf{a})\Big) & \mathrm{Cov}\Big(\mathbf{s}_i^{(2)}(\mathbf{a},\mathbf{R})\Big)\end{pmatrix}.$$

To define the Hessian matrix of the CL1 estimating equations, first set $\boldsymbol{\theta}=(\mathbf{a},\mathbf{R})^\top$, then

$$-\mathbf{H_g}=E\Big(\frac{\partial\mathbf{g}}{\partial\boldsymbol{\theta}}\Big)=\begin{pmatrix}E\Big(\frac{\partial\mathbf{g}_1}{\partial\mathbf{a}}\Big) & E\Big(\frac{\partial\mathbf{g}_1}{\partial\mathbf{R}}\Big)\\ E\Big(\frac{\partial\mathbf{g}_2}{\partial\mathbf{a}}\Big) & E\Big(\frac{\partial\mathbf{g}_2}{\partial\mathbf{R}}\Big)\end{pmatrix}=\begin{pmatrix}-\mathbf{H}_{\mathbf{g}_1} & \mathbf{0}\\ -\mathbf{H}_{\mathbf{g}_{2,1}} & -\mathbf{H}_{\mathbf{g}_2}\end{pmatrix},$$

where $-\mathbf{H}_{\mathbf{g}_1}=\frac{1}{n}\sum_i^n\mathbf{X}_i^\top\boldsymbol{\Delta}_i^{(1)}\mathbf{X}_i$, $-\mathbf{H}_{\mathbf{g}_{2,1}}=\frac{1}{n}\sum_i^n\boldsymbol{\Delta}_i^{(2,1)}\mathbf{X}_i$, and $-\mathbf{H}_{\mathbf{g}_2}=\frac{1}{n}\sum_i^n\boldsymbol{\Delta}_i^{(2,2)}$. The dimensions of the matrices above are given in Table 7.

Table 7: The dimensions of various matrices involved in the calculation of CL1 information criteria. Note that $t=p+\binom{d}{2}+q$ and $r=p+q$.

Matrix	$\mathbf{J_g}$	$\mathbf{J}_\mathbf{g}^{(1)}$	$\mathbf{J}_\mathbf{g}^{(1,2)}$	$\mathbf{J}_\mathbf{g}^{(2,1)}$	$\mathbf{J}_\mathbf{g}^{(2)}$	$\boldsymbol{\Omega}_i^{(1)}$	$\boldsymbol{\Omega}_i^{(1,2)}$	$\boldsymbol{\Omega}_i^{(2,1)}$	$\boldsymbol{\Omega}_i^{(2)}$
Dimensions	$t\times t$	$r\times r$	$r\times\binom{d}{2}$	$\binom{d}{2}\times r$	$\binom{d}{2}\times\binom{d}{2}$	$dq\times dq$	$dq\times\binom{d}{2}$	$\binom{d}{2}\times dq$	$\binom{d}{2}\times\binom{d}{2}$
Matrix	$\mathbf{H_g}$	$\mathbf{H}^{(1)}$	$\mathbf{H}^{(1,2)}$	$\mathbf{H}^{(2,1)}$	$\mathbf{H}^{(2)}$	$\boldsymbol{\Delta}_i^{(1)}$	$\boldsymbol{\Delta}_i^{(1,2)}$	$\boldsymbol{\Delta}_i^{(2,1)}$	$\boldsymbol{\Delta}_i^{(2)}$
Dimensions	$t\times t$	$r\times r$	$r\times\binom{d}{2}$	$\binom{d}{2}\times r$	$\binom{d}{2}\times\binom{d}{2}$	$dq\times dq$	$dq\times\binom{d}{2}$	$\binom{d}{2}\times dq$	$\binom{d}{2}\times\binom{d}{2}$

The matrices involved in the calculation of the sensitivity matrix $\mathbf{H_g}$ of the CL1 estimating functions $\mathbf{g}$ take the form:

$$-\mathbf{H}_{\mathbf{g}_1}=\mathbf{X}_i^\top E\begin{pmatrix}\frac{\partial\mathbf{s}_{i1}^{(1)}(\mathbf{a})}{\partial\boldsymbol{\gamma}_{i1}} & \dots & 0 & 0 & \dots & 0\\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots\\ 0 & \dots & \frac{\partial\mathbf{s}_{ij-1}^{(1)}(\mathbf{a})}{\partial\boldsymbol{\gamma}_{ij-1}} & 0 & \dots & \dots\\ 0 & \dots & 0 & \frac{\partial\mathbf{s}_{ij}^{(1)}(\mathbf{a})}{\partial\boldsymbol{\gamma}_{ij}} & \dots & \dots\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & \dots & 0 & 0 & \dots & \frac{\partial\mathbf{s}_{id}^{(1)}(\mathbf{a})}{\partial\boldsymbol{\gamma}_{id}}\end{pmatrix}\mathbf{X}_i;$$

$$-\mathbf{H}_{\mathbf{g}_{2,1}}^{\top} = \mathbf{X}_i^{\top} E \begin{pmatrix} \frac{\partial \mathbf{s}_{i,12}^{(2)}(\mathbf{a},\rho_{12})}{\partial \boldsymbol{\gamma}_{i1}} & \dots & 0 & \dots & 0 \\ \frac{\partial \mathbf{s}_{i,12}^{(2)}(\mathbf{a},\rho_{12})}{\partial \boldsymbol{\gamma}_{i2}} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & \frac{\partial \mathbf{s}_{i,j-1j}^{(2)}(\mathbf{a},\rho_{j-1j})}{\partial \boldsymbol{\gamma}_{ij-1}} & \dots & 0 \\ 0 & \dots & \frac{\partial \mathbf{s}_{i,j-1j}^{(2)}(\mathbf{a},\rho_{j-1j})}{\partial \boldsymbol{\gamma}_{ij}} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \frac{\partial \mathbf{s}_{i,d-1d}^{(2)}(\mathbf{a},\rho_{d-1d})}{\partial \boldsymbol{\gamma}_{id-1}} \\ 0 & \dots & 0 & \dots & \frac{\partial \mathbf{s}_{i,d-1d}^{(2)}(\mathbf{a},\rho_{d-1d})}{\partial \boldsymbol{\gamma}_{id}} \end{pmatrix};$$

$$-\mathbf{H}_{\mathbf{g}_2} = E \begin{pmatrix} \frac{\partial \mathbf{s}_{i,12}^{(2)}(\mathbf{a},\rho_{12})}{\partial \rho_{12}} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & \frac{\partial \mathbf{s}_{i,jk}^{(2)}(\mathbf{a},\rho_{jk})}{\partial \rho_{jk}} & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \frac{\partial \mathbf{s}_{i,d-1d}^{(2)}(\mathbf{a},\rho_{d-1d})}{\partial \rho_{d-1d}} \end{pmatrix}.$$

The elements of these matrices are calculated as below:

$$-E\Big(\frac{\partial \mathbf{s}_{i,jk}^{(2)}(\mathbf{a},\rho_{jk})}{\partial \rho_{jk}}\Big) = -E\Big(\frac{\partial^2 \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \rho_{jk}^2}\Big) = E\Big(\big(\frac{\partial \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \rho_{jk}}\big)^2\Big),$$

where $\frac{\partial \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \rho_{jk}} = \frac{\partial f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \rho_{jk}} / f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})$;

$$\begin{aligned} -E\Big(\frac{\partial \mathbf{s}_{i,jk}^{(2)}(\mathbf{a},\rho_{jk})}{\partial \mathbf{a}^{\top}}\Big) &= -E\Big(\frac{\partial^2 \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \mathbf{a}^{\top}\partial \rho_{jk}}\Big) \\ &= E\Big(\frac{\partial \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \mathbf{a}^{\top}} \frac{\partial \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \rho_{jk}}\Big), \end{aligned}$$

where

$\frac{\partial \log f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \mathbf{a}^{\top}} = \frac{\partial f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \mathbf{a}^{\top}} / f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})$,

$\frac{\partial f_2(y_{ij},y_{ik};\nu_{ij},\nu_{ik},\boldsymbol{\gamma},\rho_{jk})}{\partial \mathbf{a}^{\top}} = \frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \boldsymbol{\gamma}_{ij}}\mathbf{X}_{ij} + \frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \boldsymbol{\gamma}_{ik}}\mathbf{X}_{ik}$,

$\frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \boldsymbol{\gamma}_{ij}} = \frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \Phi^{-1}\big(F_1(y_{ij};\boldsymbol{\gamma}_{ij})\big)} \frac{\partial \Phi^{-1}\big(F_1(y_{ij};\boldsymbol{\gamma}_{ij})\big)}{\partial \boldsymbol{\gamma}_{ij}} + \frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \Phi^{-1}\big(F_1(y_{ij}-1;\boldsymbol{\gamma}_{ij})\big)} \frac{\partial \Phi^{-1}\big(F_1(y_{ij}-1;\boldsymbol{\gamma}_{ij})\big)}{\partial \boldsymbol{\gamma}_{ij}}$,

$\frac{\partial \Phi^{-1}\big(F_1(y_{ij};\boldsymbol{\gamma}_{ij})\big)}{\partial \boldsymbol{\gamma}_{ij}} = \sum_1^{y_{ij}} \frac{\partial f_1(y_{ij};\boldsymbol{\gamma}_{ij})}{\partial \boldsymbol{\gamma}_{ij}} / \phi\Big(\Phi^{-1}\big(F_1(y_{ij};\boldsymbol{\gamma}_{ij})\big)\Big)$, where $\frac{\partial f_1(y_{ij};\boldsymbol{\gamma}_{ij})}{\partial \boldsymbol{\gamma}_{ij}} = f_1(y_{ij};\boldsymbol{\gamma}_{ij}) \frac{\partial \ell_{1ij}(\boldsymbol{\gamma}_{ij},y_{ij})}{\partial \boldsymbol{\gamma}_{ij}}$.

At the above formulas

$$f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk}) = \int_{\Phi^{-1}[F_1(y_{ij}-1;\boldsymbol{\gamma}_{ij})]}^{\Phi^{-1}[F_1(y_{ij};\boldsymbol{\gamma}_{ij})]} \int_{\Phi^{-1}[F_1(y_{ik}-1;\boldsymbol{\gamma}_{ik})]}^{\Phi^{-1}[F_1(y_{ik};\boldsymbol{\gamma}_{ik})]} \phi_2(z_j,z_d;\rho_{jk}) dz_j dz_k,$$

where $F_{1ij}(y;\boldsymbol{\gamma}_{ij}) = \mathcal{F}(\gamma_{ijy})$, while the derivatives $\frac{\partial f_2(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \rho_{jk}}$ and $\frac{\partial f_{2ijk}(y_{ij},y_{ik};\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \Phi^{-1}\big(F_1(y_{ij};\boldsymbol{\gamma}_{ij})\big)}$ are computed with the R functions `exchmvn.deriv.rho` and `exchmvn.deriv.margin`, respectively, in the R package **mprobit** (Joe, 1995; Joe et al., 2011).

References

Achten, J., Parsons, N., Edlin, R., Griffin, D., and Costa, M. (2010). A randomised controlled trial of total hip arthroplasty versus resurfacing arthroplasty in the treatment of young patients with arthritis of the hip joint. *BMC Musculoskeletal Disorders*, 11.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*. John Wiley & Sons.

Albert, P. S. and McShane, L. M. (1995). A generalized estimating equations approach for spatially correlated binary data: Applications to the analysis of neuroimaging data. *Biometrics*, 51(2):627–638.

Bombardier, C., Ware, J. H., and Russell, I. J. (1986). Auranofin therapy and quality of life in patients with rheumatoid arthritis. *American Journal of Medicine*, 81:565–578.

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of the Royal Statistical Society, Series B*, 66:851–860.

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

Chen, J. and Lazar, N. A. (2012). Selection of working correlation structure in generalized estimating equations via empirical likelihood. *Journal of Computational and Graphical Statistics*, 21(1):18–41.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, 82(2):407–410.

Dawson, J., Fitzpatrick, R., Carr, A., and Murray, D. (1996). Questionnaire on the perceptions of patients about total hip replacement. *Journal of Bone and Joint Surgery – Series B*, 78(2):185–190.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, 91(435):1024–1036.

Hirk, R., Hornik, K., and Vana, L. (2019). Multivariate ordinal regression models: an analysis of corporate credit ratings. *Statistical Methods & Applications*, 28(3):507–539.

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association*, 90(431):957–964.

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

Joe, H. (2014a). *Dependence Modeling with Copulas*. Chapman & Hall, London.

Joe, H. (2014b). ***CopulaModel**: Dependence Modeling with Copulas*. R package version 0.6. URL: https://copula.stat.ubc.ca.

Joe, H., Chou, L. W., and Zhang, H. (2011). ***mprobit**: Multivariate Probit Model for Binary/Ordinal Response*. R package version 0.9-3. URL: http://CRAN.R-project.org/package=mprobit.

Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57(2):240–265.

Johnson, N. L. and Kotz, S. (1972). *Continuous Multivariate Distributions*. Wiley, New York.

Larrabee, B., Scott, H. M., and Bello, N. M. (2014). Ordinary least squares regression of ordered categorical data: Inferential implications for practice. *Journal of Agricultural, Biological, and Environmental Statistics*, 19(3):373–386.

Lee, Y. and Nelder, J. A. (2009). Likelihood inference for models with unobservables: another view. *Stastical Science*, 24(3):255–269.

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

Lindsey, J. K. and Lambert, P. (1998). On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, 17(4):447–469.

Lipsitz, S., Fitzmaurice, G., Orav, E., and Laird, N. (1994a). Performance of generalized estimating equations in practical situations. *Biometrics*, 50(1):270–278.

Lipsitz, S. R., Kim, K., and Zhao, L. (1994b). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, 13(11):1149–1163.

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

Nikoloulopoulos, A. K. (2013a). Copula-based models for multivariate discrete response data. In Durante, F., Härdle, W., and Jaworski, P., editors, *Copulae in Mathematical and Quantitative Finance*, pages 231–249. Springer.

Nikoloulopoulos, A. K. (2013b). On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. *Journal of Statistical Planning and Inference*, 143:1923–1937.

Nikoloulopoulos, A. K. (2016). Correlation structure and variable selection in generalized estimating equations via composite likelihood information criteria. *Statistics in Medicine*, 35:2377–2390.

Nikoloulopoulos, A. K. and Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, 80:126–150.

Nikoloulopoulos, A. K. and Joe, H. (2020). ***weightedScores**: Weighted Scores Method for Regression Models with Dependent data*. R package version 0.9.5.3. URL: http://CRAN.R-project.org/package=weightedScores.

Nikoloulopoulos, A. K., Joe, H., and Chaganty, N. R. (2011). Weighted scores method for regression models with dependent data. *Biostatistics*, 12:653–665.

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

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.

Nooraee, N., Molenberghs, G., and van den Heuvel, E. R. (2014). GEE for longitudinal ordinal data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Computational Statistics & Data Analysis*, 77:70–83.

Olsson, F. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44:443–460.

Pan, W. (2001). Akaike's information criterion in generalized estimating equations. *Biometrics*, 57:120–125.

Parsons, N. R. (2013). Proportional-odds models for repeated composite and long ordinal outcomescales. *Statistics in Medicine*, 32(18):3181–3191.

Parsons, N. R., Edmondson, R. N., and Gilmour, S. G. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(4):507–524.

Pearson, K. (1901). Mathematical contributions to the theory of evolution-VII. on the correlation of characters not quantitatively measureable. *Philosophical Transactions of the Royal Society of London, Series A*, 195:1–47.

Shults, J., Sun, W., Tu, X., Kim, H., Amsterdam, J., Hilbe, J. M., and Ten-Have, T. (2009). A comparison of several approaches for choosing between working correlation structures in generalized estimating equation analysis of longitudinal binary data. *Statistics in Medicine*, 28(18):2338–2355.

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Touloumis, A. (2015). R package **multgee**: A generalized estimating equations solver for multinomial responses. *Journal of Statistical Software*, 64(8):1–14.

Touloumis, A., Agresti, A., and Kateri, M. (2013). GEE for multinomial responses using a local odds ratios parameterization. *Biometrics*, 69(3):633–640.

Tutz, G. and Gertheiss, J. (2016). Regularized regression for categorical data. *Statistical Modelling*, 16(3):161–200.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21:5–42.

Walter, S., Feinstein, A., and Wells, C. (1987). Coding ordinal independent variables in multiple regression analyses. *American Journal of Epidemiology*, 125:319–323.

Wang, Y.-G. and Carey, V. (2003). Working correlation structure misspecification, estimation and covariate design: Implications for generalised estimating equations performance. *Biometrika*, 90(1):29–41.

Ware, J. H. and Lipsitz, S. R. (1986). Statistical methods for the analysis of repeated categorical outcomes. *Proceedings of the ASA Biopharmaceutical Section*, 81:254–260.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42:121–130.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, 33(3):335–356.