Efficient and feasible inference for high-dimensional normal copula regression models

Aristidis K. Nikoloulopoulos *

Abstract

The composite likelihood (CL) is amongst the computational methods used for the estimation of high-dimensional multivariate normal (MVN) copula models with discrete responses. Its computational advantage, as a surrogate likelihood method, is that is based on the independence likelihood for the univariate marginal regression and non-regression parameters and pairwise likelihood for the correlation parameters. Nevertheless, the efficiency of the CL method for estimating the univariate regression and non-regression marginal parameters can be low. For a high-dimensional discrete response, weighted versions of the composite likelihood estimating equations and an iterative approach to determine good weight matrices are proposed. The general methodology is applied to the MVN copula with univariate ordinal regressions as the marginals. Efficiency calculations show that the proposed method is nearly as efficient as the maximum likelihood for fully specified MVN copula models. Illustrations include simulations and real data applications regarding longitudinal (low-dimensional) and time (high-dimensional) series ordinal response data with covariates. Our studies suggest that there is a substantial gain in efficiency via the weighted CL method.

Key Words: Composite likelihood; discrete time-series; estimating equations; ordered probit/logistic regression; multivariate probit.

*a.nikoloulopoulos@uea.ac.uk, School of Computing Sciences, University of East Anglia, Norwich NR47TJ, U.K.

1 Introduction

The multivariate normal (MVN) copula with discrete margins has been in use for a considerable length of time, e.g., Joe (1997), and much earlier in the biostatistics (Ashford and Sowden, 1970), psychometrics (Muthén, 1978), and econometrics (Hausman and Wise, 1978) literature. It is usually known as a multivariate or multinomial probit model. The multivariate probit model is a simple example of the MVN copula with univariate probit regressions as the marginals. The use of the MVN copula with logistic regression (or Poisson or negative binomial or ordinal regression) is just a special case of the general theory of dependence modelling with copulas (e.g., Joe 2014; He et al. 2018; Smith 2022).

The MVN copula is generated by the MVN distribution and thus inherits the useful properties of the latter (e.g., Li et al. 2017; He et al. 2018). Therefore, the MVN copula allows a wide range of dependence and overcomes the drawback of limited dependence inherent in other parametric copulas such as Archimedean and nested Archimedean (Nikoloulopoulos, 2013a,b). When the univariate margins are regression models for discrete response, then copula models can be more difficult to discriminate (Joe, 2014, page 242). Hence, the MVN copula with discrete margins or discretized MVN model not only provides a wide range of flexible dependence but also approximates other copula models.

Nevertheless, implementation 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; Panagiotelis et al., 2012) unless the correlation matrix is positive exchangeable or has an 1-factor structure (Johnson and Kotz, 1972). Hence, Joe (1997) and Song (2007) restricted on low dimensional regression modelling of dependent discrete data using the MVN copula. Accordingly, Nikoloulopoulos and Moffatt (2019) and Sun et al. (2020) used copula-based Markov models where bivariate or trivariate copula functions such as the bivariate or trivariate normal are used to construct the joint distribution function of the consecutive discrete responses.

Nikoloulopoulos (2013b, 2016b) and Masarotto and Varin (2012) proposed efficient simulated likelihood methods that can be used for estimation of MVN copula discrete regression models in higher dimensions, occurring with time series, spatial data, longer longitudinal studies, but there is an issue of computational burden as the dimension d and the sample size n increase. This is also the case for the efficient Bayesian data augmentation method of Pitt et al. (2006) or the data augmentation together with a parameter expansion approach of Murray et al. (2013) as the number of latent variables is of the same size as the data, i.e., a matrix of size $n \times d$ (e.g., Panagiotelis et al. 2012; Henn 2022).

Zhao and Joe (2005) proposed composite likelihood (CL) estimating equations to overcome the computational issues at the maximization routines for the MVN copula in a high-dimensional context by using the independence likelihood for the marginal parameters and pairwise likelihood for the correlation parameters. CL is a surrogate likelihood which leads to unbiased estimating equations (Varin, 2008; Varin et al., 2011) obtained by the derivatives of the composite univariate and bivariate log-likelihoods. As the estimation of univariate marginal parameters ignores the dependence, the efficiency of estimating the univariate regression and non-regression parameters is low.

To improve the efficiency of the CL method on estimating the MVN copula with discrete margins, we propose weighted versions of the CL estimating equations and an iterative approach to determine good weight matrices. Based on the matrix version of the Cauchy-Schwarz inequality (Chaganty, 1997; Chaganty and Joe, 2004), we determine the optimal weights for which the asymptotic efficiency of the proposed estimates with these weights is close to the asymptotic efficiency of the maximum likelihood (ML) estimates. Our intent is to develop an efficient estimation method for MVN copula regression models for discrete responses that can be used for the regression analysis of high-dimensional discrete response data with covariates observed in time series or spatial statistics. We call the proposed method weighted composite likelihood (WCL).

Nevertheless, its weights are not related with the contribution of particular pairs of variables as in other weighted composite likelihood methods, e.g., Pedeli and Varin (2020).

The remainder of the paper proceeds as follows. Section 2 introduces the general theory of weighted versions of the composite likelihood estimating equations and gives the details for the MVN copula with ordinal regressions as the marginals. Section 3 studies asymptotic efficiency of our method as compared to the 'gold standard' ML method. Section 4 studies the small-sample efficiency of the weighted composite score functions in both low- and high-dimension. Section 5 presents two applications of our methodology to analyze longitudinal (low-dimensional) and time (high-dimensional) series ordinal response data. We conclude this article with some discussion.

2 Weighted versions of the composite likelihood estimating equations

To illustrate the method of the weighted versions of the composite likelihood estimating equations to estimate the MVN copula parameters concretely, we use univariate ordinal probit/logit regressions as the marginals. The resulting multivariate discrete distribution is the multivariate ordinal probit/logit model.

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 and j is an index for the repeated measurements or within cluster measurements. The MVN copula model with ordinal probit regressions as the marginals has the following cumulative distribution function (cdf):

$$F_d(y_{i1}, \ldots, y_{id}; \nu_{i1}, \ldots, \nu_{id}, \boldsymbol{\gamma}, \mathbf{R}) = \Phi_d\left(\Phi^{-1}[F_1(y_{i1}; \nu_{i1}, \boldsymbol{\gamma})], \ldots, \Phi^{-1}[F_1(y_{id}; \nu_{id}, \boldsymbol{\gamma})]; \mathbf{R}\right),$$

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 $F_1(y; \nu, \boldsymbol{\gamma})$ is the univariate cdf for the ordinal variable Y. Let Z be a latent variable with cdf $\mathcal{F}$, such that $Y = y$ if $\alpha_{y-1} + \nu \leq Z \leq \alpha_y + \nu$, $y = 1, \ldots, K$, where K is the number of categories of Y (without loss of generality, we assume $\alpha_0 = -\infty$ and $\alpha_K = \infty$), $\boldsymbol{\gamma} = (\alpha_1, \ldots, \alpha_{K-1})$ is the q-dimensional vector of the univariate cutpoints ($q = K-1$) and $\nu = \mathbf{x}^\top\boldsymbol{\beta}$ is a function of $\mathbf{x}$ and the p-dimensional regression vector $\boldsymbol{\beta}$. From this definition, the ordinal variable Y is assumed to have cdf $F_1(y;\nu,\boldsymbol{\gamma}) = \mathcal{F}(\alpha_y+\nu)$. 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).

The MVN copula lacks a closed form cdf; hence implementation of the discretized MVN is feasible, but not easy, because the MVN distribution as a latent model for discrete response requires rectangle probabilities of the form

$$f_d(\mathbf{y}_i;\nu_{i1},\ldots,\nu_{id},\mathbf{R}) = \int_{\Phi^{-1}[F_1(y_{i1}-1;\nu_{i1})]}^{\Phi^{-1}[F_1(y_{i1};\nu_{i1})]}\cdots\int_{\Phi^{-1}[F_1(y_{id}-1;\nu_{id})]}^{\Phi^{-1}[F_1(y_{id};\nu_{id})]}\phi_d(z_1,\ldots,z_d;\mathbf{R})dz_1\cdots dz_d, \quad (1)$$

where ϕ_d denotes the standard d-variate normal density with correlation matrix $\mathbf{R}$. When the joint probability is too difficult to compute, as in the case of the discretized MVN model, composite likelihood is a good alternative (Varin, 2008; Varin et al., 2011).

Zhao and Joe (2005) proposed the CL method to overcome the computational issues at the maximization routines for the MVN copula in a high-dimensional context. Estimation of the model parameters can be approached by solving the estimating equations obtained by the derivatives of the sums of univariate and bivariate log-likelihoods.

The sum of univariate log-likelihoods is

$$L_1 = \sum_{i=1}^{n}\sum_{j=1}^{d}\log f_1(y_{ij};\nu_{ij},\boldsymbol{\gamma}) = \sum_{i=1}^{n}\sum_{j=1}^{d}\ell_1(\nu_{ij},\boldsymbol{\gamma},\, y_{ij}), \quad (2)$$

where $f_1(y;\nu,\boldsymbol{\gamma}) = \mathcal{F}(\alpha_y+\nu) - \mathcal{F}(\alpha_{y-1}+\nu)$ and $\ell_1(\cdot) = \log\ f_1(\cdot)$. The score equations for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are

$$\begin{pmatrix}\frac{\partial L_1}{\partial\boldsymbol{\beta}}\\ \\ \frac{\partial L_1}{\partial\boldsymbol{\gamma}}\end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{d}\begin{pmatrix}\mathbf{x}_{ij} & \mathbf{0}\\ \\ \mathbf{0} & \mathbf{I}_q\end{pmatrix}\begin{pmatrix}\frac{\partial\ell_1(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\nu_{ij}}\\ \\ \frac{\partial\ell_1(\nu_{ij},\boldsymbol{\gamma},y_{ij})}{\partial\boldsymbol{\gamma}}\end{pmatrix} = \sum_{i=1}^{n}\sum_{j=1}^{d}\begin{pmatrix}\mathbf{x}_{ij}\mathbf{1}_q\\ \\ \mathbf{I}_q\end{pmatrix}\frac{\partial\ell_{1ij}(\boldsymbol{\gamma}_{ij},y_{ij})}{\partial\boldsymbol{\gamma}_{ij}} = \mathbf{0}, \quad (3)$$

where $\boldsymbol{\gamma}_{ij} = (\alpha_1+\nu_{ij},\ldots,\alpha_{K-1}+\nu_{ij}) = (\gamma_{ij1},\ldots,\gamma_{ij,K-1})$, $\ell_{1ij}(\cdot) = \log f_{1ij}(\cdot)$, $f_{1ij}(\boldsymbol{\gamma}_{ij},y) = \mathcal{F}(\gamma_{ijy}) - \mathcal{F}(\gamma_{ij,y-1})$, $\mathbf{I}_q$ is an identity matrix of dimension q and $\mathbf{1}_q$ is a vector of units 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 marginal parameters. The score equations in (3) can be written as

$$\mathbf{g}_1 = \mathbf{g}_1(\mathbf{a}) = \frac{\partial L_1}{\partial \mathbf{a}} = \sum_{i=1}^{n}\sum_{j=1}^{d} \mathbf{X}_{ij}^\top\, \mathbf{s}_{ij}^{(1)}(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^\top\, \mathbf{s}_i^{(1)}(\mathbf{a}) = \mathbf{0}, \tag{4}$$

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 dimensions q and dq respectively. The dimensions of $\mathbf{X}_{ij}$ and $\mathbf{X}_i$ are $q \times r$ and $dq \times r$ respectively. The CL estimate of $\mathbf{a}$, denoted by $\widetilde{\mathbf{a}}$, is the solution of $\mathbf{g}_1(\mathbf{a}) = \mathbf{0}$.

The sum of bivariate log-likelihoods is

$$L_2 = \sum_{i=1}^{n}\sum_{j<k} \log f_2(y_{ij}, y_{ik}; \nu_{ij}, \nu_{ik}, \boldsymbol{\gamma}, \rho_{jk}), \tag{5}$$

where

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

$\phi_2(\cdot; \rho)$ denotes the standard bivariate normal density with correlation ρ. Differentiating L_2 with respect to $\mathbf{R}$ leads to the bivariate composite score function:

$$\mathbf{g}_2 = \sum_{i=1}^{n} \mathbf{s}_i^{(2)}(\widetilde{\mathbf{a}}, \mathbf{R}) = \mathbf{0}, \tag{6}$$

where $\mathbf{s}_i^{(2)}(\mathbf{a}, \mathbf{R}) = \frac{\partial \sum_{j<k} \log f_2(y_{ij}, y_{ik}; \nu_{ij}, \nu_{ik}, \boldsymbol{\gamma}, \rho_{jk})}{\partial \mathbf{R}}$. The CL estimate of $\mathbf{R}$, denoted by $\widetilde{\mathbf{R}}$, is the solution of $g_2(\widetilde{\mathbf{a}}, \mathbf{R}) = \mathbf{0}$.

In what follows, we form weighted versions of the CL estimating equations and an iterative approach to determine good weight matrices. At the first stage we reform the univariate estimating composite function by inserting weight matrices between the matrix of the covariates $\mathbf{X}_i$ and the vector of univariate scores for regression and non-regression parameters in the univariate composite score function in (4). The resulting estimation function is

$$\mathbf{g}_1^\star = \mathbf{g}_1^\star(\mathbf{a}) = \sum_{i=1}^{n} \mathbf{X}_i^T\, [\mathbf{W}_i^{(1)}]^{-1}\, \mathbf{s}_i^{(1)}(\mathbf{a}), \tag{7}$$

where $\mathbf{W}_i^{(1)}$ are invertible $d(1+q) \times d(1+q)$ matrices. The WCL estimate of $\mathbf{a}$, denoted by $\hat{\mathbf{a}}$, is the solution of $\mathbf{g}_1^\star(\mathbf{a}) = \mathbf{0}$.

With $\mathbf{a}$ fixed, as estimated a the first stage of the method, we also include weight matrices in the bivariate composite score function in (6) at the second stage of the method. The resulting estimation function is

$$g_2^\star = g_2^\star(\hat{\mathbf{a}}, \mathbf{R}) = \sum_{i=1}^{n} [\mathbf{W}_i^{(2)}]^{-1} s_i^{(2)}(\hat{\mathbf{a}}, \mathbf{R}), \tag{8}$$

where $\mathbf{W}_i^{(2)}$ are invertible $\binom{d}{2} \times \binom{d}{2}$ matrices. We estimate $\mathbf{R}$ by $\widehat{\mathbf{R}}$, a solution of $g_2^\star(\hat{\mathbf{a}}, \mathbf{R}) = \mathbf{0}$.

The estimators of $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{R})$ that solve the weighted scores estimating equations $\mathbf{g}^\star = (\mathbf{g}_1^\star, \mathbf{g}_2^\star)^\top$, under the usual regularity conditions on the log-likelihood of univariate and bivariate margins as $n \to \infty$, are asymptotically normal viz.,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to N\Big(\mathbf{0}, \mathbf{G}_{g^\star}^{-1}(\boldsymbol{\theta})\Big),$$

where $\mathbf{G}_{g^\star}(\boldsymbol{\theta})$ is the Godambe information matrix (Godambe, 1991). The asymptotic covariance matrix for the estimators $\hat{\boldsymbol{\theta}}$ that solve the weighted scores estimating equations $\mathrm{g}^\star$ viz.,

$$\mathbf{G}_{g^\star}^{-1} = (-\mathbf{H}_{\mathbf{g}^\star})^{-1} \mathbf{J}_{\mathbf{g}^\star} (-\mathbf{H}_{\mathbf{g}^\star}^{T})^{-1}, \tag{9}$$

is used to obtain optimal choices for $\mathbf{W}_i^{(1)}$ and $\mathbf{W}_i^{(2)}$. The covariance matrix $\mathbf{J}_{\mathbf{g}}^\star$ of the estimating functions $\mathrm{g}^\star$ is

$$\begin{aligned} \mathbf{J}_{\mathbf{g}^\star} &= \mathrm{Cov}(\mathbf{g}^\star) = \begin{pmatrix} \mathrm{Cov}(\mathbf{g}_1^\star) & \mathrm{Cov}(\mathbf{g}_1^\star, \mathbf{g}_2^\star) \\ \mathrm{Cov}(\mathbf{g}_2^\star, \mathbf{g}_1^\star) & \mathrm{Cov}(\mathbf{g}_2^\star) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^n \mathbf{X}_i^\top [\mathbf{W}_i^{(1)}]^{-1} \boldsymbol{\Omega}_i^{(1)} [\mathbf{W}_i^{(1)\top}]^{-1} \mathbf{X}_i & \sum_{i=1}^n \mathbf{X}_i^\top [\mathbf{W}_i^{(1)}]^{-1} \boldsymbol{\Omega}_i^{(1,2)} [\mathbf{W}_i^{(2)\top}]^{-1} \\ \sum_{i=1}^n [\mathbf{W}_i^{(2)}]^{-1} \boldsymbol{\Omega}_i^{(2,1)} [\mathbf{W}_i^{(1)\top}]^{-1} \mathbf{X}_i & \sum_{i=1}^n [\mathbf{W}_i^{(2)}]^{-1} \boldsymbol{\Omega}_i^{(2)} [\mathbf{W}_i^{(2)\top}]^{-1} \end{pmatrix} \end{aligned} \tag{10}$$

where

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

The Hessian matrix $-\mathbf{H}_{\mathbf{g}^\star}$ of the estimating functions $\mathbf{g}^\star$ is

$$
\begin{aligned}
-\mathbf{H}_{\mathbf{g}^\star} &= \begin{pmatrix} E\Big(\frac{\partial \mathbf{g}_1^\star}{\partial \mathbf{a}}\Big) & E\Big(\frac{\partial \mathbf{g}_1^\star}{\partial \mathbf{R}}\Big) \\ E\Big(\frac{\partial \mathbf{g}_2^\star}{\partial \mathbf{a}}\Big) & E\Big(\frac{\partial \mathbf{g}_2^\star}{\partial \mathbf{R}}\Big) \end{pmatrix} = \begin{pmatrix} -\mathbf{H}_{\mathbf{g}_1^\star} & \mathbf{0} \\ -\mathbf{H}_{\mathbf{g}_{2,1}^\star} & -\mathbf{H}_{\mathbf{g}_2^\star} \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n \mathbf{X}_i^\top [\mathbf{W}_i^{(1)}]^{-1} \mathbf{\Psi}_i^{(1)} & 0 \\ & \\ \sum_i^n [\mathbf{W}_i^{(2)}]^{-1} \mathbf{\Psi}_i^{(2,1)} & \sum_i^n [\mathbf{W}_i^{(2)}]^{-1} \mathbf{\Delta}_i^{(2)} \end{pmatrix}, \qquad (11)
\end{aligned}
$$

where $\mathbf{\Psi}_i^{(1)} = E\Big(\frac{\partial \mathbf{s}_i^{(1)}(\mathbf{a})}{\partial \mathbf{a}}\Big)$, $\mathbf{\Psi}_i^{(2,1)} = E\Big(\frac{\partial s_i^{(2)}(\mathbf{a},\mathbf{R})}{\partial \mathbf{a}}\Big)$ and $\mathbf{\Delta}_i^{(2)} = E\Big(\frac{\partial s_i^{(2)}(\mathbf{a},\mathbf{R})}{\partial \mathbf{R}}\Big)$.

The matrix Cauchy-Schwarz inequality (Chaganty, 1997; Chaganty and Joe, 2004), shows that the optimal choices of $\mathbf{W}_i^{(1)}$ and $\mathbf{W}_i^{(2)}$ satisfy $\mathbf{X}_i^\top [\mathbf{W}_i^{(1)}]^{-1} = \mathbf{\Psi}_i^{(1)\top} [\mathbf{\Omega}_i^{(1)}]^{-1}$ and $[\mathbf{W}_i^{(2)}]^{-1} = \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1}$, respectively, leading to

$$
\mathbf{J}_{\mathbf{g}^\star} = \begin{pmatrix} \sum_{i=1}^n \mathbf{\Psi}_i^{(1)\top} [\mathbf{\Omega}_i^{(1)}]^{-1} \mathbf{\Psi}_i^{(1)} & \sum_{i=1}^n \mathbf{\Psi}_i^{(1)\top} [\mathbf{\Omega}_i^{(1)}]^{-1} \mathbf{\Omega}_i^{(1,2)} \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1} \\ & \\ \sum_{i=1}^n \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1} \mathbf{\Omega}_i^{(2,1)} [\mathbf{\Omega}_i^{(1)}]^{-1} \mathbf{\Psi}_i^{(1)} & \sum_{i=1}^n \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1} \mathbf{\Delta}_i^{(2)}. \end{pmatrix} \qquad (12)
$$

and

$$
-\mathbf{H}_{\mathbf{g}^\star} = \begin{pmatrix} \sum_{i=1}^n \mathbf{\Psi}_i^{(1)\top} [\mathbf{\Omega}_i^{(1)}]^{-1} \mathbf{\Psi}_i^{(1)} & 0 \\ & \\ \sum_{i=1}^n \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1} \mathbf{\Psi}_i^{(2,1)} & \sum_{i=1}^n \mathbf{\Delta}_i^{(2)} [\mathbf{\Omega}_i^{(2)}]^{-1} \mathbf{\Delta}_i^{(2)} \end{pmatrix}. \qquad (13)
$$

Details on the calculation of $\mathbf{\Psi}_i^{(1)}, \mathbf{\Psi}_i^{(2,1)}, \mathbf{\Delta}_i^{(2)}$ and $\mathbf{\Omega}_i^{(1)}, \mathbf{\Omega}_i^{(1,2)}, \mathbf{\Omega}_i^{(2)}$ are given in the Appendix.

To summarise, the steps to obtain the WCL estimates and standard errors, along with the involved numerical methods, are as follows.

1. Obtain the CL estimates $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$ solving the CL estimating functions in (4) and (6), respectively, or equivalently by maximizing the sum of univariate log-likelihoods in (2) and then the sum of bivariate log-likelihoods in (5) with the univariate marginal parameters fixed from the first step.

2. Compute using standard matrix operations the weight matrices $\mathbf{W}_i^{(1)}$ with the CL estimates $\widetilde{\mathbf{a}}$ and $\widetilde{\mathbf{R}}$.

3. Obtain the WCL estimates $\widehat{\mathbf{a}}$ of the univariate marginal parameters solving the estimating function in (7) using a non-linear system solver such as `multiroot` in the R package **rootSolve** (Soetaert, 2021) or `BBsolve` in the R package **BB** (Varadhan and Gilbert, 2009). The CL estimator $\widetilde{\mathbf{a}}$ is a good starting point for the non-linear root solver.

4. Compute using standard matrix operations the weight matrices $\mathbf{W}_i^{(2)}$ with the WCL estimates $\widehat{\mathbf{a}}$ of the univariate marginal parameters and the CL estimates $\widetilde{\mathbf{R}}$ of the correlation parameters.

5. Obtain the WCL correlation parameter estimates $\widehat{\mathbf{R}}$ solving the estimating function in (8) using a non-linear system solver such as `multiroot` in the R package **rootSolve** (Soetaert, 2021) or `BBsolve` in the R package **BB** (Varadhan and Gilbert, 2009). The CL estimator $\widetilde{\mathbf{R}}$ is a good starting value for the non-linear system solver.

6. Finally, the WCL standard errors for $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{R}}$ are obtained by calculating the estimated covariance matrix $\mathbf{G}_{g^\star}^{-1}$ of $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{R}}$ in in (9) with $\mathbf{J}_{\mathbf{g}^\star}$ and $-\mathbf{H}_{\mathbf{g}^\star}$ as given in (12) and (13), respectively, using standard matrix operations.

3 Relative efficiency: comparison based on asymptotic variances

We consider the MVN copula with positive exchangeable dependence and univariate ordinal logistic regressions as the marginals. For positive exchangeable dependence, if one computes the rectangle MVN probabilities with the 1-dimensional integral method in Johnson and Kotz (1972), then one is using a numerically accurate maximum likelihood (ML) method that is valid for any dimension (Nikoloulopoulos, 2013b, 2016b).

For the covariates and univariate marginal parameters we chose $p = 1, \mathbf{x}_{ij} = x_{1ij}^\top$ where x_{1ij} are taken as uniform random variables in the interval $[-1, 1]$; $\beta_1 = 0.5$ and $\boldsymbol{\gamma} = (0.33, 0.67)$. For the above parameters we computed the inverse of the Fisher information matrix $\mathcal{I}$, viz.

$$\mathcal{I}=\frac{1}{n}\sum_{i=1}^{n}\sum_{\mathbf{y}}\frac{\partial f_d(\mathbf{y};\nu_{i1},\ldots,\nu_{id},\mathbf{R})}{\partial\boldsymbol{\theta}}\frac{\partial f_d(\mathbf{y};\nu_{i1},\ldots,\nu_{id},\mathbf{R})}{\partial\boldsymbol{\theta}^T}\Big/ f_d(\mathbf{y};\nu_{i1},\ldots,\nu_{id},\mathbf{R}),$$

where the inner sum is taken over all the possible vectors $\mathbf{y}$, and the inverse Godambe matrix $\mathbf{G}_{g^\star}^{-1}$ in (9) with $\mathbf{J}_{\mathbf{g}^\star}$ and $-\mathbf{H}_{\mathbf{g}^\star}$ as given in (12) and (13), respectively. The former is the asymptotic covariance matrix of the ML estimates, while the latter is the asymptotic covariance matrix of the WCL estimates of univariate marginal and correlation parameters that are the solutions of the weighted versions of the CL estimating equations in (7) and (8). We have also computed the asymptotic covariance matrix of the CL estimates, that is the matrix $\mathbf{G}_{g^\star}^{-1}$ in (9) with $\mathbf{J}_{\mathbf{g}^\star}$ and $-\mathbf{H}_{\mathbf{g}^\star}$ as given in (10) and (11), respectively, where the weight matrices are simply the identity matrices.

Representative summaries of findings on the performance and the comparison of the competing methods are given in Table 1 for three-, six- and nine-dimensional MVN copula models with univariate ordinal logistic regressions. We took $n = 500$ to get a good approximation of the asymptotic efficiency. The comparisons are made on the scaled diagonal elements, corresponding to the asymptotic variances of the parameters, of the three matrices with different values of ρ.

Conclusions from the values in the table are the following.

- The WCL method for the univariate marginal parameters is nearly as efficient as ML.
- The CL method for the univariate marginal parameters is inefficient as the asymptotic variances are overestimated.
- The WCL method for the correlation parameters shares similar efficiency with the CL method. That is, their efficiency decreases for strong correlation as the dimension increases. This is not a worry though as for real discrete response data one does not expect correlations greater than 0.8.

Table 1: Asymptotic variances, multiplied by n, of the ML, WCL and CL estimates of the MVN copula model parameters. Efficiencies with respect to ML are shown in parenthesis.

d	True ρ	Method	β_1	γ_1	γ_2	ρ
3	0.1	ML	4.064 (1.000)	1.572 (1.000)	1.695 (1.000)	0.856 (1.000)
		WCL	4.064 (1.000)	1.572 (1.000)	1.695 (1.000)	0.856 (1.000)
		CL	4.097 (0.992)	1.572 (1.000)	1.695 (1.000)	0.856 (1.000)
	0.4	ML	3.610 (1.000)	2.107 (1.000)	2.263 (1.000)	0.851 (1.000)
		WCL	3.623 (0.996)	2.111 (0.998)	2.266 (0.999)	0.856 (0.994)
		CL	4.086 (0.884)	2.112 (0.998)	2.266 (0.999)	0.856 (0.994)
	0.7	ML	2.699 (1.000)	2.711 (1.000)	2.921 (1.000)	0.473 (1.000)
		WCL	2.744 (0.983)	2.734 (0.992)	2.939 (0.994)	0.481 (0.982)
		CL	4.104 (0.658)	2.735 (0.991)	2.940 (0.994)	0.482 (0.981)
	0.9	ML	1.735 (1.000)	3.241 (1.000)	3.508 (1.000)	0.113 (1.000)
		WCL	1.866 (0.930)	3.237 (1.001)	3.481 (1.008)	0.116 (0.977)
		CL	4.141 (0.419)	3.240 (1.000)	3.485 (1.007)	0.117 (0.964)
6	0.1	ML	2.010 (1.000)	0.916 (1.000)	0.983 (1.000)	0.234 (1.000)
		WCL	2.010 (1.000)	0.916 (1.000)	0.982 (1.000)	0.235 (0.999)
		CL	2.047 (0.982)	0.916 (1.000)	0.982 (1.000)	0.235 (0.999)
	0.4	ML	1.705 (1.000)	1.582 (1.000)	1.690 (1.000)	0.374 (1.000)
		WCL	1.715 (0.994)	1.591 (0.994)	1.696 (0.996)	0.384 (0.974)
		CL	2.102 (0.811)	1.591 (0.994)	1.696 (0.996)	0.384 (0.974)
	0.7	ML	1.239 (1.000)	2.319 (1.000)	2.494 (1.000)	0.258 (1.000)
		WCL	1.267 (0.978)	2.370 (0.979)	2.538 (0.983)	0.279 (0.926)
		CL	2.203 (0.562)	2.371 (0.978)	2.539 (0.982)	0.279 (0.926)
	0.9	ML	0.866 (1.000)	2.956 (1.000)	3.205 (1.000)	0.067 (1.000)
		WCL	0.894 (0.969)	2.999 (0.986)	3.218 (0.996)	0.075 (0.898)
		CL	2.312 (0.375)	3.003 (0.984)	3.221 (0.995)	0.076 (0.889)
9	0.1	ML	1.327 (1.000)	0.698 (1.000)	0.745 (1.000)	0.128 (1.000)
		WCL	1.327 (1.000)	0.697 (1.000)	0.745 (1.000)	0.128 (0.999)
		CL	1.359 (0.976)	0.697 (1.000)	0.745 (1.000)	0.128 (0.999)
	0.4	ML	1.096 (1.000)	1.405 (1.000)	1.497 (1.000)	0.272 (1.000)
		WCL	1.104 (0.993)	1.417 (0.992)	1.506 (0.994)	0.284 (0.957)
		CL	1.382 (0.793)	1.418 (0.991)	1.507 (0.994)	0.284 (0.957)
	0.7	ML	0.806 (1.000)	2.180 (1.000)	2.343 (1.000)	0.204 (1.000)
		WCL	0.827 (0.975)	2.248 (0.969)	2.404 (0.974)	0.231 (0.884)
		CL	1.448 (0.557)	2.249 (0.969)	2.405 (0.974)	0.231 (0.884)
	0.9	ML	0.621 (1.000)	2.843 (1.000)	3.085 (1.000)	0.055 (1.000)
		WCL	0.621 (1.001)	2.919 (0.974)	3.130 (0.986)	0.065 (0.844)
		CL	1.530 (0.406)	2.923 (0.973)	3.133 (0.985)	0.065 (0.838)

4 Simulations

We study the small-sample efficiency of the weighted composite score functions in both a low- and high-dimensional case. Section 4.1 focuses on simulated longitudinal ordinal data with small size clusters. Section 4.2 contains simulations for ordinal time-series of dimension up to $d =$

1000.

4.1 Longitudinal ordinal

We randomly generate $B = 10^4$ samples of size $n = 100, 300, 500$ from the multivariate ordinal probit model with an unstructured dependence. We use the same dimension ($d = 4$) and latent correlation matrix as in Pitt et al. (2006), viz.,

$$\mathbf{R} = \begin{pmatrix} 1 & 0.6348 & 0.5821 & 0.6916 \\ 0.6348 & 1 & 0.3662 & 0.8059 \\ 0.5821 & 0.3662 & 1 & 0.0435 \\ 0.6916 & 0.8059 & 0.0435 & 1 \end{pmatrix}$$

and $K = 5$ ordinal categories (equally weighted). For the covariates and ordinal probit regression parameters, we chose $p = 4, \mathbf{x}_{ij} = (x_{1ij}, x_{2ij}, x_{3ij}, x_{4ij})^\top$ with x_{1ij} the time, $x_{2ij} \in \{0, 1\}$ a group variable, , $x_{3ij} = x_{1ij} \times x_{2ij}$, and x_{4ij} a uniform random variable in the interval $[-1, 1]$; $\beta_1 = -\beta_2 = -\beta_3 = -0.5, \beta_4 = 1$.

Table 2 contains the parameter values, the bias, standard deviations (SD) and root mean square errors (RMSE), along with average standard errors (ASEs), scaled by n, of the WCL and CL estimates. The WCL and CL standard errors are the square root of the diagonal of the inverse Godambe matrix $\mathbf{G}_{g^\star}^{-1}$ in (9) with $\mathbf{J}_{\mathbf{g}^\star}$ and $-\mathbf{H}_{\mathbf{g}^\star}$ as given in (12) and (13), respectively; for the CL standard errors the weight matrices are simply the identity matrices. It is clear from the table that both the WCL and CL methods provide unbiased estimates and the variances computed from the simulations are similar to the asymptotic variances for both the WCL and CL methods.

4.2 Ordinal time series

We randomly generate $B = 10^4$ samples of dimension $d = 100, 200, 500, 1000$ from the multivariate ordinal probit model with latent correlation matrix corresponding to that of an autore-

Table 2: Small sample of sizes $n = 100, 300, 500$ simulations (10^4 replications) and resulted biases, standard deviations (SD), average standard errors (ASE), and root mean square errors (RMSE), scaled by n, for the WCL and CL estimates of the correlation parameters for the quadrivariate MVN copula with an unstructured dependence and ordinal probit regressions. The true parameters are $\beta_1 = -\beta_2 = -\beta_3 = -0.5, \beta_4 = 1, \gamma_j = \Phi^{-1}(0.2j), j = 1, \ldots, 4, \rho_{12} = 0.6348, \rho_{13} = 0.5821, \rho_{14} = 0.6916, \rho_{23} = 0.6348, \rho_{24} = 0.8059, \rho_{34} = 0.0435$.

	Method	n	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}
nBias	CL	100	0.65	-0.90	-0.65	1.34	-0.83	-0.09	0.56	1.36	-0.76	-1.19	-0.43	-1.37	-0.81	-1.05
	WCL		0.61	-0.95	-0.60	1.44	-0.81	-0.05	0.62	1.41	-1.10	-1.24	-0.66	-1.08	-0.55	0.13
	CL	300	0.66	-1.12	-0.60	1.49	-0.98	-0.19	0.40	1.14	-0.88	-1.24	-0.32	-1.68	-0.70	-1.48
	WCL		0.58	-1.10	-0.53	1.45	-0.97	-0.14	0.48	1.20	-1.08	-1.16	-0.50	-1.48	-0.45	-0.99
	CL	500	0.79	-0.10	-0.81	1.00	-1.64	-0.60	0.03	0.76	-1.10	-1.31	-0.59	-1.49	-0.84	-1.83
	WCL		0.72	-0.10	-0.70	1.25	-1.66	-0.64	0.01	0.76	-1.16	-1.09	-0.65	-1.17	-0.46	-1.60
nSD	CL	100	5.16	20.34	6.26	11.36	16.05	15.13	15.07	15.72	7.80	8.56	7.32	11.20	5.33	13.65
	WCL		4.76	20.10	5.71	8.88	15.89	14.98	14.91	15.53	8.82	9.59	9.47	12.05	7.77	14.82
	CL	300	8.83	35.47	10.80	19.43	27.48	26.03	25.85	27.07	13.45	14.71	12.52	19.34	8.81	23.77
	WCL		7.99	34.98	9.76	14.99	27.23	25.76	25.57	26.75	15.17	15.79	14.33	20.13	11.72	25.54
	CL	500	11.36	45.52	13.90	24.98	35.41	33.80	33.85	35.09	17.04	19.00	16.06	24.88	11.37	30.50
	WCL		10.33	44.89	12.60	19.13	35.01	33.41	33.37	34.51	18.34	19.62	16.89	25.44	12.51	32.77
nASE	CL	100	5.15	20.09	6.25	11.27	15.68	14.86	14.78	15.46	7.64	8.60	7.29	11.05	5.20	13.40
	WCL		4.63	19.73	5.57	8.63	15.46	14.63	14.53	15.18	9.78	10.22	8.97	13.58	6.71	17.07
	CL	300	8.78	34.76	10.69	19.33	27.07	25.67	25.52	26.67	13.16	14.76	12.57	19.14	8.80	23.40
	WCL		7.93	34.22	9.55	14.77	26.75	25.34	25.16	26.27	13.87	16.23	14.45	19.96	10.39	23.17
	CL	500	11.30	44.86	13.76	24.91	34.93	33.12	32.93	34.40	16.97	19.01	16.22	24.68	11.33	30.24
	WCL		10.21	44.19	12.31	19.02	34.53	32.71	32.49	33.91	17.81	20.79	18.40	25.62	13.16	29.85
nRMSE	CL	100	5.20	20.36	6.30	11.44	16.08	15.13	15.08	15.78	7.83	8.65	7.33	11.28	5.39	13.69
	WCL		4.80	20.12	5.75	9.00	15.91	14.98	14.92	15.59	8.89	9.67	9.49	12.10	7.79	14.82
	CL	300	8.86	35.48	10.82	19.49	27.50	26.03	25.85	27.09	13.48	14.77	12.53	19.42	8.84	23.82
	WCL		8.01	34.99	9.78	15.06	27.25	25.76	25.58	26.78	15.20	15.83	14.34	20.18	11.72	25.56
	CL	500	11.39	45.52	13.92	25.00	35.45	33.80	33.85	35.10	17.08	19.05	16.07	24.92	11.40	30.55
	WCL		10.36	44.89	12.61	19.17	35.05	33.42	33.37	34.52	18.38	19.65	16.90	25.47	12.52	32.81

gressive process of order one with first-order autocorrelation equals to 0.8 and $K = 4$ categories. The above problem differs from the simulations in the preceding section in that rather than being repeated or clustered measurements of a variable, the data $y_j,\ j = 1, \ldots, d$ are multivariate, with a single measurement on each of d different variables.

For the covariates and 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 $p = 3, \mathbf{x}_j = (x_{1j}, x_{2j}, x_{3j})^\top$ with $x_{1j} \in \{0, 1\}$ a binary variable with probability of success 0.4, x_{2j} a time-varying variable from a d-variate MVN copula with latent correlation matrix corresponding to that of an autoregressive process of order one with first-order autocorrelation equals to 0.5, $x_{3j} = x_{1j} \times x_{2j}$; $\beta_1 = -\beta_2 = \beta_3 = -0.5$.

For a structured dependence and high dimension d, the computation of the bivariate weight matrices involved is prohibitive as the calculation of $\mathbf{\Omega}^{(1,2)}$ and $\mathbf{\Omega}^{(2)}$ requires the computation of d-variate probabilities and hence, summations over all possible K^d vectors $\mathbf{y}$. Nevertheless this is not the case for the calculation of $\mathbf{\Omega}^{(1)}$ in the univariate weight matrices which only requires the computation of bivariate marginal probabilities and hence, summations over the K^2 possible vectors $\mathbf{y}$. Hence, we study the small-sample efficiency of the WCL estimating equations for the univariate regression or non-regression parameters, i.e., the first stage of the proposed WCL method.

Table 3 contains the parameter values, the biases, SDs and RMSEs, along with ASEs, scaled by d, of the WCL and CL estimates. It is clear from the table that both the WCL and CL methods provide unbiased estimates. The variances computed from the simulations are similar to the asymptotic variances for both the WCL and CL methods. The WCL estimates of the regression parameters are remarkably more efficient than the CL estimates as the variances in the CL method are overestimated.

Table 3: Small sample of time-series lengths $d = 100, 200, 500, 1000$ simulations (10^4 replications) and resulted biases, standard deviations (SD), and root mean square errors (RMSE), scaled by d, for the WCL and CL estimates of the univariate marginal parameters for the multivariate ordinal probit model with latent correlation matrix corresponding to that of an autoregressive process of order one with first-order autocorrelation equals to 0.8 and $K = 4$ categories.

	d		β_1	β_2	β_3	γ_1	γ_2	γ_3
dBias	100	CL	4.97	-4.37	4.79	-6.24	0.02	6.26
		WCL	5.04	-4.72	5.17	-6.34	0.20	6.58
	200	CL	5.10	-5.12	4.59	-5.92	0.07	5.76
		WCL	5.33	-4.94	3.94	-6.16	0.26	6.21
	500	CL	5.41	-6.01	5.28	-4.96	0.80	6.94
		WCL	4.69	-5.90	5.73	-5.13	1.09	7.52
	1000	CL	5.65	-6.97	3.94	-5.41	0.97	7.11
		WCL	4.48	-6.62	7.07	-5.87	1.02	7.31
dSD	100	CL	49.46	68.20	86.34	49.89	47.99	49.57
		WCL	32.33	41.54	55.39	41.02	38.72	40.77
	200	CL	68.01	93.52	117.17	68.08	66.27	68.47
		WCL	42.68	53.91	72.96	54.90	52.53	55.24
	500	CL	103.72	145.21	180.47	106.17	103.41	106.56
		WCL	63.16	81.27	108.30	85.81	82.04	86.03
	1000	CL	144.65	202.00	250.58	148.67	144.94	148.27
		WCL	87.57	110.95	150.69	118.98	113.92	118.54
dASE	100	CL	46.49	61.65	81.00	43.57	42.28	43.56
		WCL	32.45	40.61	56.22	36.12	34.52	36.11
	200	CL	64.98	88.27	112.58	63.81	62.00	63.69
		WCL	42.31	53.43	72.61	52.15	49.80	52.04
	500	CL	102.43	141.11	176.85	103.22	100.27	103.04
		WCL	63.81	80.95	108.88	83.76	79.86	83.60
	1000	CL	144.93	200.68	249.90	147.33	142.78	147.01
		WCL	89.10	113.14	151.54	119.42	113.41	119.11
dRMSE	100	CL	49.71	68.34	86.47	50.28	47.99	49.96
		WCL	32.72	41.81	55.63	41.50	38.72	41.29
	200	CL	68.21	93.66	117.26	68.34	66.27	68.71
		WCL	43.01	54.14	73.06	55.25	52.53	55.58
	500	CL	103.86	145.33	180.55	106.29	103.41	106.78
		WCL	63.34	81.48	108.45	85.96	82.04	86.36
	1000	CL	144.76	202.12	250.61	148.77	144.94	148.44
		WCL	87.69	111.14	150.86	119.12	113.92	118.77

5 Applications

In this section we illustrate the proposed estimation method through two examples with longitudinal (low-dimensional) and time (high-dimensional) series ordinal response data. We include comparisons with other possible fitting methods, such as ML (low dimension) and CL (high dimension).

5.1 Arthritis data

We illustrate the weighted composite scores equations by analysing the rheumatoid arthritis dataset (Bombardier et al., 1986). 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 and 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). Further, both logit and probit links are used for the ordinal regressions.

Table 4 gives the estimates and standard errors of the model parameters obtained using ML and the WCL estimating equations. Ordinal logistic regression is slightly better than ordinal probit regression based on the composite and full likelihoods. Our analysis shows that the WCL estimates of all the parameters and their corresponding standard errors are nearly the same as the ML estimates.

Table 4: Maximized log-likelihoods, WCL and ML estimates (Est.) along with their standard errors (SE) for the arthritis data.

	Logit link				Probit link			
Covariates & cutpoints	WCL		ML		WCL		ML	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
$I(\text{time}=2)$	-0.007	0.124	-0.006	0.125	-0.005	0.071	-0.007	0.072
$I(\text{time}=3)$	-0.377	0.116	-0.377	0.115	-0.218	0.066	-0.220	0.066
trt	-0.500	0.168	-0.487	0.165	-0.337	0.097	-0.336	0.097
$I(\text{baseline}=2)$	-0.659	0.345	-0.607	0.357	-0.336	0.200	-0.341	0.200
$I(\text{baseline}=3)$	-1.208	0.329	-1.161	0.337	-0.580	0.190	-0.576	0.190
$I(\text{baseline}=4)$	-2.569	0.370	-2.487	0.382	-1.319	0.211	-1.315	0.211
$I(\text{baseline}=5)$	-4.040	0.555	-3.975	0.549	-2.264	0.324	-2.262	0.320
age	0.013	0.008	0.014	0.007	0.008	0.004	0.008	0.004
sex	-0.167	0.187	-0.179	0.179	-0.062	0.109	-0.062	0.108
α_1	-1.768	0.673	-1.831	0.625	-1.029	0.385	-1.016	0.383
α_2	0.351	0.656	0.230	0.608	0.071	0.381	0.059	0.382
α_3	2.324	0.662	2.222	0.613	1.249	0.383	1.250	0.385
α_4	4.641	0.682	4.526	0.625	2.544	0.390	2.545	0.390
ρ_{12}	0.393	0.057	0.376	0.061	0.393	0.057	0.373	0.061
ρ_{13}	0.505	0.051	0.503	0.052	0.509	0.051	0.505	0.052
ρ_{23}	0.530	0.050	0.536	0.046	0.523	0.050	0.528	0.046
Log-likelihood	-2114.855		-1041.477		-2117.755		-1043.083	

5.2 Infant sleep status data

The sleep data (e.g., Fokianos and Kedem 2003) consist of sleep state measurements of a newborn infant together with his heart rate and temperature sampled every 30 seconds. The sleep states are classified as: (1) quiet sleep, (2) indeterminate sleep, (3) active sleep, (4) awake. The total number of observations is equal to 1024 and the objective is to predict the sleep state based on covariate information.

The response, sleep state, is an ordinal time series in the sense that the response increases from awake to active sleep, i.e., "(4)" $<$ "(1)" $<$ "(2)" $<$ "(3)". We use the standardized heart rate and temperature as covariates to avoid large estimates for the univariate cutpoints. Fokianos and Kedem (2003) have previously adopted regression models for this ordinal time series and confirmed an autoregressive model of order 1 to adequately capture the serial dependence among the ordinal observations.

Table 5: Maximized bivariate composite log-likelihoods L_2 in (5), WCL and CL estimates (Est.) along with their standard errors (SE) for the sleep data.

Logit link								
Covariates &	CL				WCL			
cutpoints	Est.	SE	Z	p-value	Est.	SE	Z	p-value
heart rate	0.074	0.274	0.271	0.786	0.097	0.044	2.232	0.026
temperature	0.284	0.339	0.837	0.402	0.254	0.190	1.336	0.182
α_1	-0.934	0.385	-2.425	0.015	-0.778	0.372	-2.091	0.037
α_2	0.771	0.376	2.053	0.040	0.865	0.376	2.300	0.021
α_3	1.233	0.406	3.040	0.002	1.322	0.408	3.240	0.001
L_2	-1321423							
Probit link								
Covariates &	CL				WCL			
cutpoints	Est.	SE	Z	p-value	Est.	SE	Z	p-value
heart rate	0.040	0.164	0.245	0.807	0.057	0.026	2.221	0.026
temperature	0.167	0.202	0.827	0.408	0.169	0.114	1.479	0.139
α_1	-0.582	0.231	-2.520	0.012	-0.515	0.227	-2.268	0.023
α_2	0.469	0.228	2.058	0.040	0.510	0.227	2.250	0.024
α_3	0.744	0.236	3.154	0.002	0.781	0.235	3.317	0.001
L_2	-1321157							

Table 5 summarizes the WCL and CL estimates of the regression and non-regression parameters. The first-order autocorrelation parameter of the MVN copula with latent correlation matrix corresponding to that of an autoregressive process of order one is estimated as 0.96 for both logistic and probit ordinal regression. The latter is slightly better than ordinal logit regression based on the composite likelihood values. It is obvious from the table that ignoring the actual serial dependence in the data on the CL estimation of the regression parameters leads to invalid conclusions resulting to no effect of the time-dependent covariates at sleep state. The WCL analyses reveal that the heart rate effect is a statistically significant predictor of sleep state.

6 Discussion

We have studied a weighted composite likelihood estimating equations approach, namely the WCL, based on weighting the univariate and bivariate scores of the univariate and bivariate margins of a MVN copula model with discrete margins to estimate the model parameters. The WCL method leads to efficient estimating equations for both the univariate marginal (regression and non-regression) and correlation parameters.

For high-dimensional discrete data such as discrete time or spatial series, the first stage of the method, i.e., the estimation of the univariate marginal parameters, is computationally feasible as the weight matrices of the WCL estimating equations in (7) depend on covariances of the univariate scores that only require the computation of bivariate marginal probabilities. Nevertheless, as such data require a structured correlation, such as an autoregressive moving average or the Matérn isotropic structure, the second stage of the method, i.e., the estimation of the correlation parameters, becomes cumbersome. However, as the WCL correlation parameter estimates share the same efficiency as that achieved by the traditional CL correlation parameter estimates, the problem can be circumvented by using the latter. That is, the correlation and univariate marginal parameters can be estimated by the CL method and the WCL method, respectively. Hence, estimation of MVN copula regression models is efficient and feasible for high-dimensional discrete response data with covariates.

Nikoloulopoulos et al. (2011) and Nikoloulopoulos (2016a, 2020) have considered estimating equations of this form when dependence is considered a nuisance; this leads to an extension of generalized estimating equations of Liang and Zeger (1986) in that a wider class of univariate regression models can be called. A sandwich-type estimator is used to obtain estimates of the covariance matrix of model parameters that are robust to misspecification. Nevertheless, such methods of inference about regression and non-regression parameters are not available in the setting where is a single measurement on each of d different variables.

Future research will focus on the model selection (e.g., Alexopoulos and Bottolo 2021) for high-dimensional normal copula regression models. The WCL is a likelihood method, and thus, analogues of the AIC and BIC for correlation structure and variable selection can be derived in the framework of the composite likelihood.

Software

A contributed R package **weightedCL** (Nikoloulopoulos, 2022) has R functions to estimate high-dimensional MVN copula regression models with the WCL estimating equations. It provides ARMA(p,q) correlation structures and binary, ordinal, Poisson, and negative binomial (both NB1 and NB2 parametrizations in Cameron and Trivedi 1998) regressions.

Appendix

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

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

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

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

The elements of these matrices are calculated as below:

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

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

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

where

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

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

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

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

At the above formulas

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

where $F_{1ij}(y;\boldsymbol{\gamma}_{ij}) = \mathcal{F}(\gamma_{ijy})$, while the derivatives $\frac{\partial f_2(y_j,y_k;\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \rho_{jk}}$ and $\frac{\partial f_{2ijk}(y_j,y_k;\boldsymbol{\gamma}_{ij},\boldsymbol{\gamma}_{ik},\rho_{jk})}{\partial \Phi^{-1}\big(F_1(y_j;\boldsymbol{\gamma}_{ij})\big)}$ 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).

For $d=4$ the matrices involved in the calculation of the covariance matrix $\mathbf{J_g}$ of the composite score functions $\mathbf{g}$ take the form:

$$\boldsymbol{\Omega}_i^{(1)}=\begin{pmatrix}\mathrm{Var}\Big(\mathbf{s}_{i1}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i1}^{(1)}(\mathbf{a}),\mathbf{s}_{ij}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i1}^{(1)}(\mathbf{a}),\mathbf{s}_{id}^{(1)}(\mathbf{a})\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{ij}^{(1)}(\mathbf{a}),\mathbf{s}_{i1}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Var}\Big(\mathbf{s}_{ij}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{ij}^{(1)}(\mathbf{a}),\mathbf{s}_{id}^{(1)}(\mathbf{a})\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{id}^{(1)}(\mathbf{a}),\mathbf{s}_{i1}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{id}^{(1)}(\mathbf{a}),\mathbf{s}_{ij}^{(1)}(\mathbf{a})\Big) & \dots & \mathrm{Var}\Big(\mathbf{s}_{id}^{(1)}(\mathbf{a})\Big)\end{pmatrix},$$

$$\boldsymbol{\Omega}_i^{(1,2)}=\begin{pmatrix}\mathrm{Cov}\Big(\mathbf{s}_{i1}^{(1)},\mathbf{s}_{i,12}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i1}^{(1)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i1}^{(1)},\mathbf{s}_{i,d-1d}^{(2)}\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{ij}^{(1)},\mathbf{s}_{i,12}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{ij}^{(1)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{ij}^{(1)},\mathbf{s}_{i,d-1d}^{(2)}\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{id}^{(1)},\mathbf{s}_{i,12}^{(2)}\Big) & & \dots\mathrm{Cov}\Big(\mathbf{s}_{id}^{(1)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{id}^{(1)},\mathbf{s}_{i,d-1d}^{(2)}\Big)\end{pmatrix},$$

where

$$\begin{aligned}\mathrm{Cov}\Big(\mathbf{s}_{ij_1}^{(1)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) &= \sum_{\mathbf{y}}\mathbf{s}_{ij_1}^{(1)}\mathbf{s}_{i,j_1j_2}^{(2)}f_2(y_{j_1},y_{j_2};\nu_{ij_1},\nu_{ij_2},\rho_{j_1j_2}),\\ \mathrm{Cov}\Big(\mathbf{s}_{ij_1}^{(1)},\mathbf{s}_{i,j_2j_3}^{(2)}\Big) &= \sum_{\mathbf{y}}\mathbf{s}_{ij_1}^{(1)}\mathbf{s}_{i,j_2j_3}^{(2)}f_3(y_{j_1},y_{j_2},y_{j_3};\nu_{ij_1},\nu_{ij_2},\nu_{ij_3},\rho_{j_1j_2},\rho_{j_1j_3},\rho_{j_2j_3}),\end{aligned}$$

and

$$\boldsymbol{\Omega}_i^{(2)}=\begin{pmatrix}\mathrm{Var}\Big(\mathbf{s}_{i,12}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i,12}^{(2)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i,12}^{(2)},\mathbf{s}_{i,d-1d}^{(2)}\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{i,j_1j_2}^{(2)},\mathbf{s}_{i,12}^{(2)}\Big) & \dots & \mathrm{Var}\Big(\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i,j_1j_2}^{(2)},\mathbf{s}_{i,d-1d}^{(2)}\Big)\\ \vdots & \ddots & \vdots & \ddots & \vdots\\ \mathrm{Cov}\Big(\mathbf{s}_{i,d-1d}^{(2)},\mathbf{s}_{i,12}^{(2)}\Big) & \dots & \mathrm{Cov}\Big(\mathbf{s}_{i,d-1d}^{(2)},\mathbf{s}_{i,j_1j_2}^{(2)}\Big) & \dots & \mathrm{Var}\Big(\mathbf{s}_{i,d-1d}^{(2)}\Big)\end{pmatrix},$$

where

$$\begin{aligned}\mathrm{Var}\Big(\mathbf{s}_{i,j_1j_2}^{(2)}\Big) &= \sum_{\mathbf{y}}\mathbf{s}_{i,j_1j_2}^{(2)}\mathbf{s}_{i,j_1j_2}^{(2)}f_2(y_{j_1},y_{j_2}),\\ \mathrm{Cov}\Big(\mathbf{s}_{i,j_1j_2}^{(2)},\mathbf{s}_{i,j_1j_3}^{(2)}\Big) &= \sum_{\mathbf{y}}\mathbf{s}_{i,j_1j_2}^{(2)}\mathbf{s}_{i,j_1j_3}^{(2)}f_3(y_{j_1},y_{j_2},y_{j_3}),\\ \mathrm{Cov}\Big(\mathbf{s}_{i,j_1j_2}^{(2)},\mathbf{s}_{i,j_3j_4}^{(2)}\Big) &= \sum_{\mathbf{y}}\mathbf{s}_{i,j_1j_2}^{(2)}\mathbf{s}_{i,j_3j_4}^{(2)}f_4(y_{j_1},y_{j_2},y_{j_3},y_{j_4});\end{aligned}$$

the inner sum is taken over all possible vectors $\mathbf{y}$. Hence, for an unstructured dependence the calculation of $\boldsymbol{\Omega}_i^{(1)}$ requires summations over the K^2 possible vectors $\mathbf{y}$, while the calculation of $\boldsymbol{\Omega}_i^{(1,2)}$ and $\boldsymbol{\Omega}_i^{(2)}$ requires summations up to K^3 and K^4 possible vectors $\mathbf{y}$, respectively.

Acknowledgements

We thank the Associate Editor and the referees for comments leading to an improved presentation. The simulations presented in this paper were carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

References

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

Alexopoulos, A. and Bottolo, L. (2021). Bayesian variable selection for gaussian copula regression models. *Journal of Computational and Graphical Statistics*, 30(3):578–593.

Ashford, J. R. and Sowden, R. R. (1970). Multivariate probit analysis. *Biometrics*, 26:535–546.

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. (1997). Loss in efficiency due to misspecification of the correlation structure in GEE. In *Proc. 51st Sess. International Statistical Institute*, volume 2, pages 127–128, Istanbul. Voorburg: International Statistical Institute.

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.

Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science*, 18(3):357–376.

Godambe, V. P. (1991). *Estimating Functions*. Oxford University Press, Oxford.

Hausman, J. and Wise, D. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 45:319–339.

He, Y., Zhang, X., and Zhang, L. (2018). Variable selection for high dimensional gaussian copula regression model: An adaptive hypothesis testing procedure. *Computational Statistics & Data Analysis*, 124:132–150.

Henn, L. L. (2022). Limitations and performance of three approaches to bayesian inference for gaussian copula regression models of discrete data. *Computational Statistics*, (2):909–946.

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

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.

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

Li, J., Nott, D., Fan, Y., and Sisson, S. (2017). Extending approximate bayesian computation methods to high dimensions via a gaussian copula model. *Computational Statistics & Data Analysis*, 106:77–89.

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

Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics*, 6:1517–1549.

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*, 108(502):656–665.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43:551–560.

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*, volume vol 213, pages 231–249, Berlin, Heidelberg. 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. (2016a). Correlation structure and variable selection in generalized estimating equations via composite likelihood information criteria. *Statistics in Medicine*, 35:2377–2390.

Nikoloulopoulos, A. K. (2016b). Efficient estimation of high-dimensional multivariate normal copula models with discrete spatial responses. *Stochastic Environmental Research and Risk Assessment*, 30(2):493–505.

Nikoloulopoulos, A. K. (2020). Weighted scores estimating equations and CL1 information criteria for longitudinal ordinal response. *Journal of Statistical Computation and Simulation*, 90(11):2002–2022.

Nikoloulopoulos, A. K. (2022). *weightedCL: Efficient and feasible inference for high-dimensional normal copula regression models*. R package version 0.5. URL:http://CRAN.R-project.org/package=weightedCL.

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. (2009). Finite normal mixture copulas for multivariate discrete data modeling. *Journal of Statistical Planning and Inference*, 139:3878–3890.

Nikoloulopoulos, A. K. and Moffatt, P. G. (2019). Coupling couples with copulas: analysis of assortative matching on risk attitude. *Economic Inquiry*, 57(1):654–666.

Panagiotelis, A., Czado, C., and Joe, H. (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107:1063–1072.

Pedeli, X. and Varin, C. (2020). Pairwise likelihood estimation of latent autoregressive count models. *Statistical Methods in Medical Research*, 29(11):3278–3293.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, 93:537–554.

Smith, M. S. (in press). Implicit Copulas: An Overview. *Econometrics and Statistics*, DOI: j.ecosta.2021.12.002.

Soetaert, K. (2021). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*. R package version 1.8.2.3.

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

Sun, L.-H., Huang, X.-W., Alqawba, M. S., Kim, J.-M., and Emura, T. (2020). Copula Markov Models for Count Series with Excess Zeros. In *Copula-Based Markov Models for Time Series*, pages 101–127, Springer, Singapore.

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

Varadhan, R. and Gilbert, P. (2009). `BB`: An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *Journal of Statistical Software*, 32:1–26.

Varin, C. (2008). On composite marginal likelihoods. *Advances in Statistical Analysis*, 92:1–28.

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.

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