Statistica Sinica Preprint No: SS-2016-0435	
Title	Discrete Longitudinal Data Modeling with a Mean-Correlation Regression Approach
Manuscript ID	SS-2016-0435.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0435
Complete List of Authors	Cheng Yong Tang Weiping Zhang and Chenlei Leng
Corresponding Author	Weiping Zhang
E-mail	zwp@ustc.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

Statistica Sinica

DISCRETE LONGITUDINAL DATA MODELING WITH A MEAN-CORRELATION REGRESSION APPROACH

ChengYong Tang[1] and Weiping Zhang[2] and Chenlei Leng[3]

[1] *Temple Universityand* [2] *University of Science and Technology of China and* [3] *University of Warwick*

Abstract: Joint mean-covariance regression modeling with unconstrained parametrization for continuous longitudinal data has provided statisticians and practitioners with a powerful analytical device. How to develop a delineation of such a regression framework amongst discrete longitudinal responses remains an open and more challenging problem. This paper studies a novel mean-correlation regression for a family of generic discrete responses. Targeting the joint distributions of the discrete longitudinal responses, our regression approach is constructed by using a copula model whose correlation parameters are represented in hyperspherical coordinates with no constraint on their support. To overcome computational intractability in maximizing the full likelihood function of the model, we propose a computationally efficient pairwise likelihood approach. A pairwise likelihood ratio test is then constructed and validated for statistical inferences. We show that the resulting estimators of our approaches are consistent and asymptotically normal. We demonstrate the effectiveness, parsimoniousness and desirable performance of the proposed approach by analyzing three data sets and conducting

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

extensive simulations.

Key words and phrases: Joint Distribution; Discrete longitudinal data, Hyperspherical coordinates, Likelihood ratio test, Mean-correlation regression, Cholesky decomposition, Pairwise likelihood.

1. Introduction Longitudinal observations are characterized by repeated measurements from the same subjects, giving rise to their feature of rich, interesting, and practically meaningful covariance structures. In contrast to analyzing independent data, revealing, understanding, and explaining the correlation structures are fundamental and crucial not only for developing appropriate models but also for drawing and interpreting conclusions from the data sets on the trends, changes, and other aspects of interest in various studies (Diggle et al. (2002); Fitzmaurice et al. (2004)). With multiple subjects in a longitudinal study, a specific goal is to characterize the covariance matrices, one for each subject, for those repeated measurements using parsimonious regression techniques. While it is useful to employ conventional ARMA structures or random effects (Diggle et al., 2002) for modeling the covariance/correlation of the longitudinal responses, one often find that only limited choices of such devices are available (Pourahmadi (1999); Zhang et al. (2015)). One often resorts to developing regression models that utilize covariates for depicting various target associations of interest.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

For instance, for more comprehensive interpretations and predictions, one can explore correlation structures incorporating more explanatory variables additional to times of the observations; see, for example, Hoffman (2012) for modeling with multiple random effects. Indeed, as shown in a data example in Section 4.2, additional covariates to the time-lag of observations are found significant for the explaining the correlation structures of the longitudinal data.

A key challenge in dealing with a covariance matrix with regression techniques is the positive definite requirement. For continuous longitudinal responses, Pourahmadi (1999, 2000) pioneered joint modeling approaches. Pivotal to these approaches is a modified Cholesky decomposition of covariance that allows unconstrained parametrization of the entries in the decomposition. This permits the development of interpretable regression models akin to autoregressive models in a time series context (Pourahmadi (2011)). A new class of models motivated by moving average models were further developed by Zhang and Leng (2012). Zhang et al. (2015) proposed models to investigate marginal variances and correlations from a geometric perspective. Other important works on joint modeling for continuous longitudinal data include Pan and Mackenzie (2003); Ye and Pan (2006); Pourahmadi (2007); Daniels and Pourahmadi (2009); Leng et al. (2010);

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

Xu and Mackenzie (2012).

These developments have mainly focused on continuous longitudinal data with observations from social, economic, and medical studies often contain a substantial number of discrete variables. See, among others, some studies in Lynn (2009), that focus on discrete responses. Hence, it is as important for practitioners to parsimoniously model the dependence structure of the discrete longitudinal responses as in investigating continuous cases; see, among others, the monographs by Molenberghs and Verbeke (2005) and Bergsma et al. (2009).

Despite the ubiquity of discrete longitudinal responses, analyzing them is more challenging due to the lack of suitable multivariate joint distributions for discrete variables that broadly incorporate the correlations between measurements from the same subject. It is known that even for given marginal distributions of the discrete variables, such as Bernoulli or Poisson, specifying the joint distributions of multiple longitudinal measurements incorporating between measurements correlations remains difficult (Molenberghs and Verbeke (2005); Bergsma et al. (2009)). Moreover, although progress has been made in modeling the mean for longitudinal discrete responses (Diggle et al. (2002)), it is an open difficult problem to develop regression methods for simultaneously analyzing the mean and covariance

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

structure for discrete data. In particular, for identifiability issues, the covariance matrix is constrained as a correlation matrix (Chib and Greenberg (1998)). The need to parametrize a matrix to be positive definite and have unit diagonals immediately renders the inapplicability of the modified Cholesky approach in Pourahmadi (1999, 2000) and the moving average decomposition method in Zhang and Leng (2012). In the Bayesian context, Daniels and Pourahmadi (2009) made use of the partial autocorrelations (PACs). But, difficulties remain in explaining these PAC and in building more elaborate regression models. Wang and Daniels (2013) studied a Bayesian modeling approach for continuous longitudinal data via PACs and marginal variances, and Gaskins, et al. (2014) proposed models to obtain sparse PACs. Other existing approaches for modeling and incorporating correlations include the Markov model on the transitional probability matrix for binary data (Muenz and Rubinstein (1985)), the working model approach (Zeger et al. (1985)), the estimating equation approach (Zeger and Liang (1986)), and the double hierarchical modeling approach with random effects (Lee and Nelder (2006)). None of them discuss the problem of building general regression models using covariates for modeling correlations of discrete longitudinal data.

In this paper, we propose an approach for adaptively and flexibly mod-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

eling discrete longitudinal data, focusing on a mean-correlation regression analysis that solves both problems of generally specifying joint distributions and parsimoniously modeling correlations with no constraint. To our best knowledge, this is the first time such tools have been made available. To accommodate a broad class of dependent discrete longitudinal data that can be unbalanced and observed at irregular times, we advocate a unified framework for the joint distributions of the discrete responses from the same subject by using a copula, in conjunction with appropriate univariate marginal distributions.

The paper is organized as follows. Section 2 introduces the joint mean-correlation-dispersion modeling approach of the paper. Section 3 discusses the theoretical properties of the estimators and presents a new test based on pairwise likelihood ratio for hypothesis testing. Section 4 presents extensive numerical simulations and three data analyses. Conclusions and an outline of future study are found in Section 5. Technical details including sketch of proofs, additional data analysis example and simulations studies are given in the Supplementary Material of this paper.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

2. Main methodology

2.1 The joint modeling approach

An appealing approach for modeling correlated discrete longitudinal variables is the copula construction (Song, et al. (2009)). Sklar's theorem ensures that a multivariate distribution can be determined jointly by the univariate marginal distributions and a copula, a multivariate function of these marginals responsible for dependence. For our paper, we use the Gaussian copula. As a counterpart of the Gaussian distribution, the Gaussian copula has merits of being convenient and has been demonstrated useful in recent studies (e.g. Liu et al. (2009)). Formally, a set of random variables $\boldsymbol{U} = (U_1, \ldots, U_d)^{\mathrm{T}}$ follows a Gaussian copula model if their joint distribution is specified by

$$F(u_1, \ldots, u_d) = P(U_1 \leq u_1, \ldots, U_d \leq u_d) = \Phi_d(v_1, \ldots, v_d; \mathbf{R}).$$

Here Φ_d is the probability distribution function of the d-dimensional standardized normal distribution with zero mean, $\mathbf{R}$ is the correlation matrix, and $v_i = \Phi_1^{-1}(w_i)$ where $w_i = P(U_i \leq u_i)$ is the marginal distribution of U_i $(1 \leq i \leq d)$. The copula construction is attractive in that it decouples the marginal feature from the dependence structure, and can treat con-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

tinuous, categorical and mixed data in a unified fashion. Because of the decoupling, models developed for independent data can be seamlessly incorporated by appropriately manipulating the marginal distributions. In our study, we consider the Gaussian copula because of its merits in flexibility, interpretability, and parsimony in its parameters for capturing the data features, sharing those of the multivariate normal distribution. We remark that other copulas, for example, the t-copula (Fang et al. (2002)), can also be applied without compromising the essence of our mean-correlation modeling framework.

Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})^{\mathrm{T}}$ be the m_i longitudinal measurements for the ith subject, where the discrete response y_{ij} is observed at time t_{ij}. We consider without loss of generality that the discrete variable takes integer values, $y_{ij} \in \{0, 1, 2, \ldots\}$. Let $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_i})^{\mathrm{T}}$, and let $\mathbf{x}_{ij} \in \mathbb{R}^p$ be the covariate for the jth measurement of subject i. With this notation, we intend to develop models that can handle general unbalanced longitudinal data. Existing methods, for example, those in Song, et al. (2009) and Gaskins, et al. (2014) work on balanced and equally-spaced longitudinal data.

With multiple subjects, we denote the observations as $\{y_{ij}, \mathbf{x}_{ij}, t_{ij}\}$ ($i = 1, \ldots, n; j = 1, \ldots, m_i$). For categorical responses, we assume that y_{ij} has

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

an exponential family distribution so that generalized linear models (GLMs) can be used for the discrete responses marginally (McCullagh and Nelder (1989)). We write the marginal probability mass function of Y as $f(y) = c(y;\varphi)\exp[\{y\theta - \psi(\theta)\}/\varphi]$ with canonical parameter θ and scale parameter φ. Since $\psi'(\theta) = E(Y) := \mu$, we denote the canonical link function by $(\psi')^{-1}(\mu) := g(\mu)$. For the mean, we postulate the usual GLM marginally for each y_{ij} as

$$g\left(E(y_{ij})\right) = g(\mu_{ij}) = \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}. \tag{2.1}$$

Here $\mathrm{var}(y) = \varphi\psi''(\theta)$ with dispersion parameter φ depending on the family of the discrete response variables. We then take the joint distribution of $\mathbf{y}_i$ as

$$F_{m_i}(\mathbf{y}_i) = P(Y_{i1} \le y_{i1}, \ldots, Y_{im_i} \le y_{im_i}) = \Phi_{m_i}(z_{i1}, \ldots, z_{im_i}; \mathbf{R}_i), \tag{2.2}$$

where $z_{ij} = \Phi_1^{-1}\{F(y_{ij})\}$ $(j = 1, \ldots, m_i)$, F is the marginal distribution function of Y specified by the GLM, and $\mathbf{R}_i = (\rho_{ijk})_{j,k=1}^{m_i}$ is the correlation matrix for the ith subject. This copula modeling device allows the marginal distributions and the correlations of the discrete longitudinal responses to be treated separately. Although the elements in $\mathbf{R}_i$ are not directly the correlations between the discrete observations, they are determining the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

dependence of the longitudinal observations via (2.2). When the responses are binary, the correlation between two observations is a monotone function of the corresponding element in $\mathbf{R}_i$; see also Fan et al. (2017). We also refer to the discussions in Song (2000) on the connection between the correlation coefficients in $\mathbf{R}_i$ and those of the observed variables.

Clearly, with so many parameters in $\{\mathbf{R}_i\}$ $(i = 1, \ldots, n)$ associated with un-balanced longitudinal data, existing conventional copula modeling approaches generally do not apply due to the problem of over-parametrization. In our approach, we decompose $\mathbf{R}_i$ as

$$\mathbf{R}_i = \mathbf{T}_i \mathbf{T}_i^{\mathrm{T}}, \tag{2.3}$$

where $\mathbf{T}_i$ is a lower triangular matrix given by

$$\mathbf{T}_i = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ c_{i21} & s_{i21} & 0 & \cdots & 0 \\ c_{i31} & c_{i32}s_{i31} & s_{i32}s_{i31} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{im_i1} & c_{im_i2}s_{im_i1} & c_{im_i3}s_{im_i2}s_{im_i1} & \cdots & \prod_{l=1}^{m_i-1} s_{im_il} \end{pmatrix}, \tag{2.4}$$

where $c_{ijk} = \cos(\omega_{ijk})$ and $s_{ijk} = \sin(\omega_{ijk})$ are trigonometric functions of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

angles $\omega_{ijk} \in [0, \pi)$ $(1 \leq k < j \leq m_i)$ that are the parameters under the new parametrization.

For any matrix $\mathbf{T}_i$, $\mathbf{R}_i = \mathbf{T}_i\mathbf{T}_i^{\mathrm{T}}$ is guaranteed to be nonnegative definite. The special form of $\mathbf{T}_i$ in (2.3) ensures further that the diagonals of $\mathbf{R}_i$ are 1. The order of the angles added into the lower triangular $\mathbf{T}_i$ respects the longitudinal nature of the data collected along the time dimension. Thus, the effect of the decomposition is to transform the unknown positive definite correlations $\{\mathbf{R}_i\}$ into unconstrained parameters in $\{\omega_{ijk}\}$ on $[0, \pi)$. This decomposition in (2.3) appeared in Creal et al. (2011) for analyzing time series and was studied by Zhang et al. (2015) for regression with continuous longitudinal responses where it was argued that the angles ω_{ijk} represent rotations of these coordinates and their magnitude reflects roughly the correlations amongst different components.

Since all angles in (2.3) are unconstrained on $[0, \pi)$, we propose to model these angles $\{\omega_{ijk}\}$ collectively via a regression model after a monotone transformation from $\mathbb{R}$:

$$\omega_{ijk} = \pi/2 - \mathrm{atan}(\mathbf{w}_{ijk}^{\mathrm{T}}\boldsymbol{\gamma}), \tag{2.5}$$

where $\mathbf{w}_{ijk} \in \mathbb{R}^q$ is a covariate and $\boldsymbol{\gamma}$ is the $q \times 1$ unknown parameters.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

We opt to use the arctan transformation to ensure that the parameter $\boldsymbol{\gamma}$ for covariate $\mathbf{w}_{ijk}$ in (2.5) is completely constraint free. A dimension reduction is achieved by (2.5) that uses only q parameters for modeling all n correlation matrices $\{\mathbf{R}_i\}$ $(i = 1, \ldots, n)$. While $\mathbf{w}_{ijk}$ depends on two indices j and k of the ith subject, we need to examine the covariates of the ith subject at the corresponding observations. We follow the convention of longitudinal data analysis by taking $\mathbf{w}_{ijk}$ as some function of the time lag $|t_{ij} - t_{ik}|$ between observations, which effectively ensures the correlation to be stationary; see also Pourahmadi (1999). Other time-dependent covariates may also be meaningfully exploited; an example is available in Section 4.2 for analyzing the Mayo PBC liver data. Thus our regression approach for the correlations can incorporate a broad class of covariates available for explaining the covariations between longitudinal measurements.

We refer to our proposed method for modeling discrete longitudinal data collectively using (2.1)-(2.5) as the mean-correlation regression approach. By combining all unknown parameters in this modeling framework, we write collectively the parameter vector of interest as $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\mathrm{T}}, \boldsymbol{\gamma}^{\mathrm{T}}, \varphi)^{\mathrm{T}}$. Using the GLM for the responses marginally in (2.1) and the model in (2.5) for the correlations, we are ready to develop the maximum likelihood estimators for $\boldsymbol{\theta}$. A difficulty is that applying copula to fit discrete data is known to

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

be computationally intensive. This arises in the fact that a d-dimensional Gaussian Copula has continuous support on $\mathbb{R}^d$ while discrete response variable are conceptually defined only on discrete grid points, only probabilities evaluated on the grid points are well defined. To see this, we write the full likelihood as

$$\begin{aligned} L(\boldsymbol{\theta}) &= \prod_{i=1}^{n} P(Y_{i1} = y_{i1}, \ldots, Y_{im_i} = y_{im_i}) \\ &= \prod_{i=1}^{n} P(y_{i1} - 1 < Y_{i1} \leq y_{i1}, \ldots, y_{im_i} - 1 < Y_{im_i} \leq y_{im_i}) \\ &= \prod_{i=1}^{n} \int \cdots \int_{\mathbf{z}_i^- < \mathbf{u} \leq \mathbf{z}_i} \phi_{m_i}(\mathbf{u}; \mathbf{R}_i) d\mathbf{u}, \end{aligned} \tag{2.6}$$

where $\mathbf{z}_i = (z_{i1}, \ldots, z_{im_i})^{\mathrm{T}}$ and $\mathbf{z}_i^- = (z_{i1}^-, \ldots, z_{im_i}^-)^{\mathrm{T}}$ with $z_{ij} = \Phi_1^{-1}\{F(y_{ij})\}$, $z_{ij}^- = \Phi_1^{-1}\{F(y_{ij} - 1)\}$, and $z_{ij}^- = -\infty$ when y_{ij} takes the smallest possible value on its support. The vector inequality $\mathbf{z}_i^- < \mathbf{u} \leq \mathbf{z}_i$ means $z_{i1}^- < u_1 \leq z_{i1}$, ..., $z_{im_1}^- < u_{m_i} \leq z_{im_i}$. Though integrals in the full likelihood can be approximated numerically, the computational cost is high and may not scale easily to even a moderate number of repeat measurements. Directly calculating the distribution function of each subject i specified by (2.2) requires 2^{m_i} summations of lower dimensional distribution functions as in the approach of Song, et al. (2009), and the computational cost grows exponentially with m_i.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

For computation, we apply the composite likelihood idea in Varin, et al. (2011) by using pairwise likelihood.

2.2 The pairwise likelihood (PL) approach

To estimate the parameters in the model specified by (2.1)-(2.5), we construct all pairwise likelihoods via the bivariate copula as

$$pL(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{1\le j<k\le m_i} \int_{z_{ij}^-}^{z_{ij}} \int_{z_{ik}^-}^{z_{ik}} \phi_2(\mathbf{u}; \rho_{ijk}) d\mathbf{u}, \tag{2.7}$$

where $\phi_2(\cdot;\rho)$ is the probability density function of bivariate normal $N(0,0,1,1,\rho)$. The computational cost is noticeably lower than that of the full likelihood as (2.7) involves $m_i(m_i-1)/2$ summations for each subject in the longitudinal data, a polynomial order complexity as compared to the exponential order in computing the full likelihood. Furthermore, each summand can be obtained by approximating a bivariate normal distribution function that can be evaluated very quickly and accurately with existing computational routines developed for low-dimensional integration, for example, those in Tong (1990) and the ones implemented in R (e.g. function `biv.nt.prob` in package `mnormt`;and function `pmvnorm` in package `mvtnorm`). Calculating the pairwise likelihood is highly scalable as each pairwise likelihood can be

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

done separately, which is an ideal fit for modern computational facilities.

By using (2.7) in conjunction with mean-correlation regression models specified in (2.1)-(2.5), our method enhances the conventional pairwise likelihood methods for studying covariance and correlation matrices. An appealing feature of our approach is that ρ_{ijk} in (2.7) is specified by the hyperspherical decomposition in (2.3), (2.4) and (2.5); This is highly parsimonious and the resulting correlation matrix is automatically positive definite.

Denote the log pairwise likelihood function as

$$pl(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{1\le j<k\le m_i} log \int_{z_{ij}^-}^{z_{ij}} \int_{z_{ik}^-}^{z_{ik}} \phi_2(\mathbf{u}; \rho_{ijk}) d\mathbf{u} := \sum_{i=1}^{n} \sum_{1\le j<k\le m_i} l_{ijk}(\theta), \tag{2.8}$$

and the score function as

$$\mathbf{S}_n(\boldsymbol{\theta}) = \frac{\partial pl}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \sum_{1\le j<k\le m_i} \frac{\partial l_{ijk}}{\partial \boldsymbol{\theta}} := \sum_{i=1}^{n} \mathbf{S}_{ni}(\boldsymbol{\theta}). \tag{2.9}$$

We employ the modified Fisher scoring algorithm to maximize (2.8). The exact forms of the score function and the expected Hessian matrix for $pl(\boldsymbol{\theta})$ are provided in the Supplementary Material.

Denote $\boldsymbol{\theta}^{(t-1)}$ as the updated value of $\boldsymbol{\theta}$ at iteration $(t-1)$. We update

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

the estimates by the iterative equation $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)} + \mathbf{H}_n^{-1}(\boldsymbol{\theta}^{(t-1)})\mathbf{S}_n(\boldsymbol{\theta}^{(t-1)})$, where $\mathbf{H}_n$ is the expected Hessian matrix given later in (3.1).

The parameters $\boldsymbol{\eta} = (\boldsymbol{\beta}^{\mathrm{T}}, \psi)^{\mathrm{T}}$ can be initialized by fitting the marginal GLMs, assuming an independent correlation structure where $\rho_{ijk} = 0$ equivalent to $\boldsymbol{\gamma} = 0$. These initial estimators of $\boldsymbol{\beta}$ and ψ are known to be root-n consistent (Zeger and Liang (1986)). If data are balanced with $\mathbf{R}_i = \mathbf{R}$, it is not difficult to find an initial consistent estimator of $\boldsymbol{\gamma}$. One can obtain a sample estimator of $\mathbf{R}$ that is root-n consistent using the initial consistent estimators of $\boldsymbol{\beta}$ and ψ. By noticing $\omega_{1jk} = \cdots = \omega_{njk}$ for balanced data, we can use (2.5) to consistently estimate $\boldsymbol{\gamma}$. It is then straightforward to show that the one-step estimator is as efficient as the fully iterated estimators, a reminiscent of what is true for one-step estimators for the MLE. If data are unbalanced, obtaining the global optimal solution of the likelihood or the pairwise likelihood is more difficult. Our experience, is that the iterative procedure discussed converges to the optimal solution, and the numerical results reported in Section 4 are based on this simple iterative procedure.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

3. Main results

3.1 Asymptotic properties

The asymptotic property of the maximum likelihood estimation involves the limit of the expected Hessian matrix $\mathbf{H}(\boldsymbol{\theta}) = \lim_{n\to\infty} -\frac{1}{n}E(\partial^2 pl/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}})$, and the limit of variance $\mathbf{J}(\boldsymbol{\theta}) = \lim_{n\to\infty} Var_{\boldsymbol{\theta}}(\frac{1}{\sqrt{n}}\mathbf{S}_n(\boldsymbol{\theta}))$, where the expectation is conditioned on the covariates $\mathbf{x}_{ij}$ and $\mathbf{w}_{ijk}$. To formally establish the theoretical properties, we need some regularity conditions.

Condition A1: The dimensions p and q of covariates $\mathbf{x}_{ij}$ and $\mathbf{w}_{ijk}$ are fixed; $n \to \infty$ and $\max_i m_i$ is bounded from above.

Condition A2: The true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0^{\mathrm{T}}, \boldsymbol{\gamma}_0^{\mathrm{T}}, \varphi_0)^{\mathrm{T}}$ is in the interior of the parameter space Θ, a compact subset of $\mathbb{R}^{p+q+1}$.

Condition A3: $\mathbf{H}(\boldsymbol{\theta}_0)$ and $\mathbf{J}(\boldsymbol{\theta}_0)$ are positive definite matrices.

Condition A4: If the expected Hessian matrix for the full likelihood method is $\mathbf{I}(\boldsymbol{\theta}) = -E(\partial^2 \log L/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}})$, as $n \to \infty$, $\mathbf{I}(\boldsymbol{\theta}_0)/n$ converges to a positive definite matrix $\mathcal{I}(\boldsymbol{\theta}_0)$.

Theorem 1. *If the conditions A1, A2 and A4 hold and* $\check{\boldsymbol{\theta}} = (\check{\boldsymbol{\beta}}^{\mathrm{T}}, \check{\boldsymbol{\gamma}}^{\mathrm{T}}, \check{\varphi})^{\mathrm{T}}$ *is the maximum likelihood estimator, the maximizer of (2.6), then* $\sqrt{n}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0))$, *where* $\mathcal{I}(\theta)$ *is the Fisher information matrix of Condition A4.*

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

Theorem 2. *If the conditions A1, A2 and A3 hold and* $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\mathrm{T}}, \hat{\boldsymbol{\gamma}}^{\mathrm{T}}, \hat{\varphi})^{\mathrm{T}}$ *is the maximum pairwise likelihood estimator, the maximizer of (2.7), then* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N(0, \mathbf{G}^{-1}(\boldsymbol{\theta}_0))$, *where* $\mathbf{G}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\mathbf{J}^{-1}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})$, *the Godambe information matrix.*

Since $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$, $\mathbf{H}$ and $\mathbf{J}$ in the asymptotic covariance matrix are consistently estimated by

$$\mathbf{H}_n(\hat{\boldsymbol{\theta}}) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{1\le j<k\le m_i} \ddot{l}_{ijk}(\hat{\boldsymbol{\theta}}), \tag{3.1}$$

where $\ddot{l}_{ijk}(\boldsymbol{\theta}) = \partial^2 l_{ijk}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}$, and $\mathbf{J}_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_{ni}(\hat{\boldsymbol{\theta}})\mathbf{S}_{ni}^{\mathrm{T}}(\hat{\boldsymbol{\theta}})$. Therefore, $\mathbf{G}(\boldsymbol{\theta}_0)$ can be consistently estimated as

$$\mathbf{G}_n(\hat{\boldsymbol{\theta}}) = \mathbf{H}_n(\hat{\theta})\mathbf{J}_n(\hat{\boldsymbol{\theta}})^{-1}\mathbf{H}_n(\hat{\boldsymbol{\theta}}). \tag{3.2}$$

The difference between the efficiencies of the pairwise likelihood and the full likelihood depends on the difference between the Godambe information matrix in Theorem 2 and the Fisher information matrix in Theorem 1, where the latter determines the lower variance bound of unbiased estimators. Our method for estimating $\boldsymbol{\beta}$ and φ, is consistent even when (2.2) is not correctly specified. When the $\mathbf{R}_i$ in (2.2) are the identity matrix,

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

our method is equivalent to the approach ignoring all dependence between the longitudinal data that remains consistent for the parameters $\boldsymbol{\beta}$ and φ. When there is a departure from the model assumption on the correlations, one can follow the existing framework of statistical inference with mis-specified model, e.g. White (1982). The probability limit of the parameter estimate is the one in the support of the parameter space such that the corresponding model has the smallest Kullback-Leibler divergence from the truth.

3.2 Pairwise likelihood ratio and hypothesis testing

We discuss a procedure based on pairwise likelihood ratio for testing hypotheses. This test is useful when the interest is to assess the statistical evidence for single or multiple components in the parameter $\boldsymbol{\theta}$. Specifically, subject to a permutation of the entries of $\boldsymbol{\theta}$, write $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}})^{\mathrm{T}}$ where $\boldsymbol{\theta}_1$ is an $r \times 1$ parameter of interest, $\boldsymbol{\theta}_2$ is a nuisance parameter. We want to test $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$ against $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1,0}$. Let $\hat{\boldsymbol{\theta}}$ be the unrestricted maximum pairwise likelihood estimate and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{1,0}^{\mathrm{T}}, \tilde{\boldsymbol{\theta}}_2^{\mathrm{T}})^{\mathrm{T}}$ be the (profile) maximum pairwise likelihood estimate under the null hypothesis. We partition the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

total score statistic $\mathbf{S}_n(\boldsymbol{\theta})$ at (2.9) correspondingly as

$$\mathbf{S}_n(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{S}_{n,1}(\boldsymbol{\theta}) \\ \mathbf{S}_{n,2}(\boldsymbol{\theta}) \end{pmatrix}.$$

The maximum pairwise likelihood estimates $\hat{\boldsymbol{\theta}}$ under the alternative hypothesis and $\tilde{\boldsymbol{\theta}}$ under the null hypothesis satisfy, respectively, $\mathbf{S}_n(\hat{\boldsymbol{\theta}}) = 0, \quad \mathbf{S}_{n,2}(\boldsymbol{\theta}_{1,0}, \tilde{\boldsymbol{\theta}}_2) = 0$. We partition the Hessian matrix $\mathbf{H}$ and its inverse as

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}, \quad \mathbf{H}^{-1} = \begin{pmatrix} \mathbf{H}^{11} & \mathbf{H}^{12} \\ \mathbf{H}^{21} & \mathbf{H}^{22} \end{pmatrix},$$

and where $\mathbf{H}_{11\cdot 2} = (\mathbf{H}^{11})^{-1} = \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21}$. The same partitions are applied to $\mathbf{G}$ and $\mathbf{G}^{-1}$. Then the pairwise likelihood ratio statistic is

$$LRT = 2\{pl(\hat{\boldsymbol{\theta}}) - pl(\tilde{\boldsymbol{\theta}})\},$$

where $pl(\boldsymbol{\theta})$ is the log pairwise likelihood function given by (2.8).

Theorem 3. *Under conditions A1, A2 and A3, for testing the hypothesis* $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$ *versus* $H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_{1,0}$, *asymptotically as* $n \to \infty$, *the pairwise likelihood ratio statistic* $LRT = 2\{pl(\hat{\boldsymbol{\theta}}) - pl(\tilde{\boldsymbol{\theta}})\} \xrightarrow{d} \sum_{j=1}^{r} \lambda_j V_j$, *where* $V_1, \ldots, V_r$ *are independent* χ_1^2 *random variables and* $\lambda_1 \geq \cdots \geq \lambda_r$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

are the eigenvalues of $(\mathbf{H}^{11})^{-1}\mathbf{G}^{11}$.

Since $\mathbf{H}_n$ and $\mathbf{G}_n$ given by (3.1) and (3.2) are consistent estimators of $\mathbf{H}$ and $\mathbf{G}$, the eigenvalues $\lambda_1, \cdots, \lambda_r$ can be estimated consistently by the corresponding eigenvalues of $(\mathbf{H}_n^{11})^{-1}\mathbf{G}_n^{11}$. Then the critical value of the pairwise likelihood ratio test statistic can be obtained straightforwardly by simulations. We have applied the testing procedure in detecting significant features in both the mean and correlation parts of the regression model; see Section 4.2. Simulations given in Section 4.4 show that the testing procedure works satisfactorily.

4. Examples: data analyses and simulations

4.1 Mayo PBC liver data

We applied the proposed method to the primary biliary cirrhosis (PBC) of the liver data set as in Appendix D of Fleming and Harrington (1991). The PBC data set was collected in a study conducted by the Mayo Clinic from 1974 to 1984 and is available in many R packages (Eg. `mixAK` and `JM`). The major goal of this double-blinded randomised placebo-controlled trial is to assess the efficacy of a new drug, the D-penicillamine. This data set contains survival time and other information on 312 PBC patients participating in the trial. The original clinical protocol for these patients

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

specified visits at six months, one year, and annually thereafter, leading to unequally spaced observations times. Due to death and censoring, patients on average made 6.2 visits with a standard deviation 3.8, resulting in a highly un-balanced repeated measurement data set. Since earlier studies have shown that there were no therapeutic differences between control and D-penicillamine-treated patients, we examined the relationship between a patient's hepatomegaly status and other covariates.

We found clear evidence that the hepatomegaly status is highly correlated with other covariates. For example, Pearson chi-square tests gave highly significant statistical evidence for the existence of correlation between hepatomegaly and a variable named spiders. Let Y_{ij} denote the hepatomegaly indicator at visit j for patient i where $Y_{ij} = 1$ if hepatomegaly developed and 0 otherwise. We considered the following covariates: *Age = Age in years;* t_{ij} *= Number of years between enrollment and this visit date; drug = 0 for placebo and 1 for D-penicillmain treatment; ascites = presence of ascites, 0 for No and 1 for Yes; spiders = blood vessel malformations in the skin, 0 for No and 1 for Yes; Bili = Serum bilirubin, in mg/dl; Alb = Albumin in gm/dl; Plat = Platelet count; Protime = Prothrombin time, in second.*

Observations with incomplete covariates were ignored. The remaining

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

235 patients with 116 cases of developed hepatomegaly were analyzed using the logistic regression model:

$$\begin{aligned}\text{logit}(Y_{ij}) = \beta_0 + \beta_1 Age_i + \beta_2 Drug_i + \beta_3 Ascites_{ij} + \beta_4 Spiders + \beta_5 \log(Bili_{ij}) \\ + \beta_6 \log(Alb_{ij}) + \beta_7 \log(Plat_{ij}) + \beta_8 \log(Protime_{ij}),\end{aligned}$$

and the angles ω_{jk} at (2.3) for the correlations matrix are modeled by

$$\tan(\pi/2 - \omega_{ijk}) = f(t_{ij} - t_{ik}) + \gamma_3 \Big| \log\Big(\frac{Protime_{ij}}{Protime_{ik}}\Big)\Big|,$$

where $f(t_{ij}-t_{ik}) = \gamma_0+\gamma_1(t_{ij}-t_{ik})+\gamma_2(t_{ij}-t_{ik})^2$ is a quadratic polynomial of the time lag chosen by the composite likelihood versions of BIC criterion (Gao and Song (2010)). Here the difference in Prothrombin time (after log-transform) is a time dependent covariate additional to functions in time lag that we included in the regression analysis for correlations.

The estimated parameters with standard deviations for the mean were $\hat{\beta}_0 = 5.749_{2.155}$, $\hat{\beta}_1 = 0.002_{0.012}$, $\hat{\beta}_2 = -0.416_{0.239}$, $\hat{\beta}_3 = 0.470_{0.246}$, $\hat{\beta}_4 = 0.645_{0.154}$, $\hat{\beta}_5 = 0.541_{0.108}$, $\hat{\beta}_6 = -2.780_{0.346}$,$\hat{\beta}_7 = -0.337_{0.698}$, and $\hat{\beta}_8 = -0.403_{0.189}$. As a comparison, a GEE approach with unstructured working correlation was also implemented and we found $\tilde{\beta}_0 = 4.5296_{2.2296}$, $\tilde{\beta}_1 =$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

$0.0016_{0.0104}$, $\tilde{\beta}_2 = -0.4212_{0.2126}$, $\tilde{\beta}_3 = 0.3205_{0.2732}$, $\tilde{\beta}_4 = 0.5724_{0.1633}$, $\tilde{\beta}_5 = 0.5700_{0.0892}$, $\tilde{\beta}_6 = -1.9099_{0.5313}$, $\tilde{\beta}_7 = -0.3084_{0.7080}$ and $\tilde{\beta}_8 = -0.3593_{0.1770}$. Using the hypothesis testing approach in Theorem 3, the p-value of 0.734 for testing $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, suggested that a smaller model could be adequate for modeling the conditional mean function. The estimated correlation parameters were $\hat{\gamma}_0 = 0.633_{0.082}$, $\hat{\gamma}_1 = -0.140_{0.034}$, $\hat{\gamma}_2 = 0.007_{0.003}$ and $\hat{\gamma}_3 = 1.092_{0.488}$. By using the pairwise likelihood ratio test of Theorem 3, we tested $H_0 : \gamma_1 = \gamma_2 = 0$, $H_0 : \gamma_1 = 0$ or $H_0 : \gamma_2 = 0$. All the p-values turned out to be close to zero, indicating that the quadratic polynomial in time lag for the angles is highly significant. The p-value was 0.009 for $H_0 : \gamma_3 = 0$, showing that the difference in Prothrombin time (after log-transform) is highly significant in the correlation modeling. This indicated that, additional to the time, other more general variable can play a statistically significant role in explaining the correlation structures. The left plot of Figure 1 gives the plot of fitted $\tan(\pi/2 - \hat{\omega}_{ijk})$ versus time lags, and the right plot in Figure 1 shows the fitted correlations versus time lag. The correlations generally decrease with time lag, indicating that the hepatomegaly status could be highly correlated with the disease status at the most recent measuring times.

The difference between patterns in Figures 1 and 5 (in the online supple-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

mentary materials) is interesting, though both are decreasing. Importantly, our development in Theorems 1–3 provides an effective device for collecting data evidence for more effective model building in taking both the mean and correlation into consideration for unbalanced and unequally spaced discrete longitudinal data.

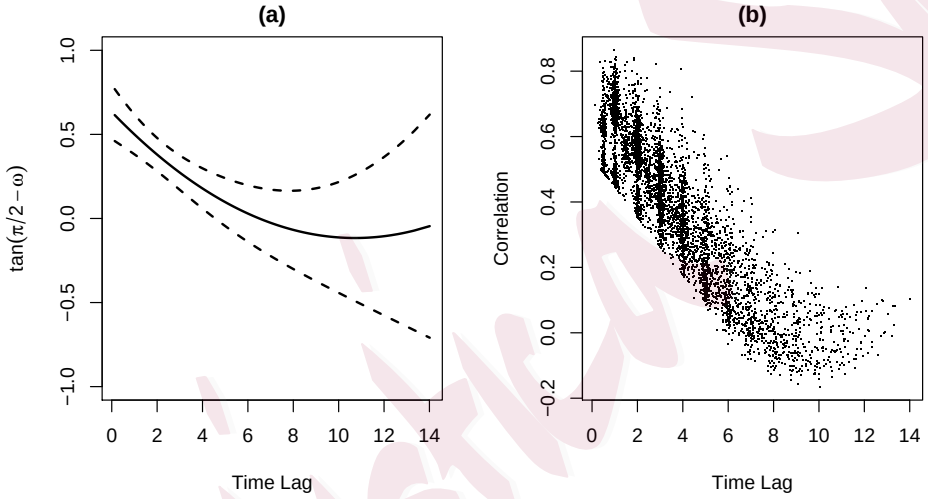

Figure 1: Mayo PBC liver data: (a) plot of fitted angles $\tan(\pi/2 - \hat{\omega}_{jk})$ versus time lag, (b) plot of fitted correlations versus time lag. In panel (a), the solid red line is the fitted line by the proposed model, and the dashed curves represent asymptotic 95% confidence intervals.

4.2 The Epileptic seizure data

The Epileptic seizure Data (Thall and Vail (1990)) concerns a randomised clinical trial of 59 epileptic patients who were randomly assigned to a new drug(trt=1) or a placebo(trt=0) as an adjuvant to the standard chemother-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

apy. This data set has been analyzed by, for example, Diggle et al. (2002) and Molenberghs and Verbeke (2005). Baseline data are available at the time when patients entered the trial, including the number of epileptic seizure recorded in the preceding 8-week period (*expind*=0) and age in years. The patients were then randomly assigned to the treatment by the drug Progabide (31 patients) or to the placebo group (28 patients). They were then followed for four 2-week periods (*expind*=1) and the number of seizures recorded. To account for the over-dispersion, we used a parametric negative binomial regression model for the mean (Diggle et al. (2002))

$$Y_{ij} \sim \text{Negbin}(\delta, \mu_{ij}),$$
$$\log(\mu_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 expind_i + \beta_2 trt_i + \beta_3 expind_i * trt_i,$$

where δ is the overdispersion parameter, $t_{ij} = 8$ if $j = 0$ and $t_{ij} = 2$ for $j = 1, 2, 3, 4$. The $\log(t_{ij})$ is needed to account for different observation periods.

We analyzed this data set via the proposed approach using a polynomial of the time lag for modelling the correlations, and started with a common correlation $\mathbf{R}_i = \mathbf{R}$ for all i. To model the angles ω_{jk} in the correlation, the angles were first directly estimated by maximizing the proposed composite

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

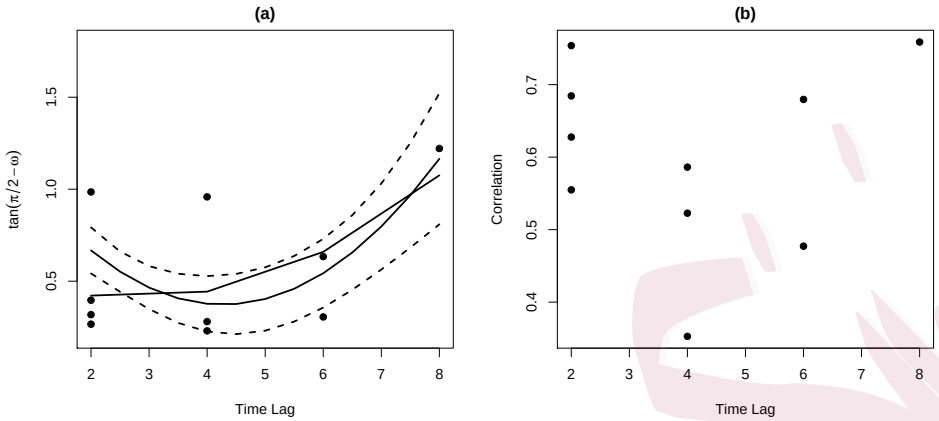

Figure 2: The Epileptic seizure Data: (a) plot of fitted angles $\tan(\pi/2 - \hat{\omega}_{jk})$ versus time lag, (b) plot of fitted correlations versus time lag. In panel (a), solid dots are fitted angles with a common correlation matrix for all subjects with parametrization (2.4), the solid black line is from fitting a LOWESS curve to the solid dots; the solid red line is the fitted line by the proposed model, and the dashed curves represent asymptotic 95% confidence intervals.

likelihood with respect to the full model, then a model including quadratic terms of the time lags for the angles were fitted based on the composite likelihood versions of BIC criterion. As discussed in Diggle et al. (2002), patient number 207 was deleted since he had unusual pre- and post-randomisation seizure counts. The estimated parameters in the mean model were $\hat{\beta}_0 = 1.346_{0.178}, \hat{\beta}_1 = 0.112_{0.144}, \hat{\beta}_2 = -0.107_{0.245}, \hat{\beta}_3 = -0.302_{0.208}$. Overall, there is little difference between the treatment and placebo groups in affecting seizure counts. A similar finding by using GEE was given by Diggle et al. (2002). The over-dispersion parameter $\hat{\delta} = 1.330_{0.221}$ is significant, suggesting that the counts are over-dispersed. For the parameters in the correlation

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

model, we obtained $\hat{\gamma}_0 = 1.413_{0.210}, \hat{\gamma}_1 = -0.487_{0.098}, \hat{\gamma}_2 = 0.057_{0.011}$. Figure 2 (a) shows the plots of the fitted angles in form of $\tan(\pi/2 - \omega_{jk})$ versus the time lag, suggesting that a polynomial model for correlations is reasonable. The curved pattern between the correlation and time in Figure 2 (b) is interesting, which may be due to the fact that the seizure counts may at first be more highly correlated with the most recent measurements, and then become more correlated with the baseline counts. This is coincident with the conclusion that there is little difference between the treatment and placebo groups in affecting seizure counts. The maximum time lag here is 8 such that the number of observations for estimating correlations between larger time lag is far fewer. Thus one needs to take caution because the associated level of uncertainty may be higher for inferring correlations at large time lag.

To assess the adequacy of the model fitting, we conducted some visual model diagnostics. Upon fitting the proposed model, we can get estimated probabilities denoted by $\widehat{F}_{1,i}(\mathbf{y}_i)$ $(i = 1, \ldots, n)$. On the other hand, we can calculate empirical distribution as $\widehat{F}_{2,i}(\mathbf{y}_i) = n^{-1} \sum_{j=1}^{n} I(y_{1j} \leq y_{1i}, \ldots, y_{mj} \leq y_{mi})$. A plot of $\widehat{F}_{1,i}$ vs $\widehat{F}_{2,i}$ is an overall diagnostic of goodness of fit; it is given in (a) of Figure 3, showing an overall reasonable fitting of the distribution. As a second diagnostic, we focused on the fitting of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

the correlation structure. In particular, we computed the empirical correlations between the z-scores, $z_{ij} = \Phi^{-1}(F(y_{ij}))$, and then plotted it against the fitted correlation with the proposed method; it is given in (b) of Figure 3, indicating a reasonable fit of the correlation matrix $\mathbf{R}_i$.

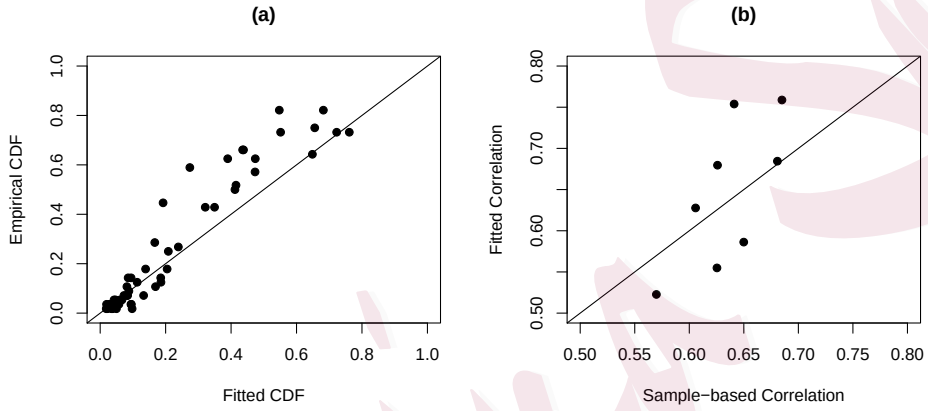

Figure 3: Plots of model diagnostics: (a) the empirical distribution function vs the fitted distribution function; (b) the empirical correlations of the z-scores vs the fitted correlations.

4.3 Simulations

We conducted extensive simulations to assess the performance of the mean-correlation modeling methodology with R. We also compared the pairwise likelihood estimates (PLEs) with the MLEs in terms of their biases and variances, and evaluated the accuracy of the inferential procedure for estimating the standard errors of the estimators. As a benchmark, we compared

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

our method to the GEE method in Liang and Zeger (1986) for estimating the parameters in the mean model and the dispersion, assuming unstructured correlations. In each of the studies, we generated 500 data sets and took sample sizes $n = 50, 100$ and 200. All simulations were conducted in R. Here are the differences in time for obtaining the PLEs and MLEs for Study 1 when $n = 50$. On the average, it takes twice as much time to obtain the MLEs when $m_i = 4$, twenty times as such time when $m_i = 6$. When $m_i = 8$, the computational time is intractable for the full likelihood approach, while for the pairwise likelihood approach, the computational time is manageable even for larger m_i.

Study 1. The data sets were generated from the model

$$y_{ij} \sim \text{Poisson}(\lambda_{ij}), \qquad \log(\lambda_{ij}) = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2,$$

$$\omega_{ijk} = \pi/2 - \text{atan}(\gamma_0 + w_{ijk1}\gamma_1 + w_{wjk2}\gamma_2), \ (i = 1, \ldots, n; j = 1, \ldots, m_i),$$

where the measurement times t_{ij} were uniform. We considered two cases: I $m_i \equiv 6$ and II $m_i - 1 \sim \mathit{Binomial}(6, 0.8)$. The latter case gives different numbers of repeated measurements m_i for different subjects. The covariate $x_{ij} = (x_{ij1}, x_{ij2})^{\mathrm{T}}$ was generated from a standard bivariate normal distribution with zero correlation. We took the covariates for the corre-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

lations as $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2\}^{\mathrm{T}}$. The parameters were set as $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) = (1.0, -0.5, 0.5)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2) = (0.5, -0.3, 0.5)$. There was no dispersion parameter for this study.

Table 1 shows the accuracy of the estimated parameters in terms of their mean biases (MB) and standard deviations. For PLEs, all the biases are small especially when n is large. Additionally, to evaluate the inference procedure, we compared the sample standard deviation (SD) of 500 parameter estimates to the sample average of 500 standard errors (SE) using formula (3.2). The standard deviation (Std) of 500 standard errors in Table 1 show that the SD and SE are quite close, especially for large n. This indicates that the standard error formula works well and demonstrates the validity of Theorem 1. Although estimators based on the pairwise likelihood function are slightly less efficient than the maximum likelihood estimates, they have smaller biases. In particular, the MLEs for estimating the parameters in the correlation matrices are highly biased, likely due to the computational difficulty of evaluating multidimensional integrals when a full likelihood is used. Compared to the GEE estimates with unstructured correlations for estimating the parameters in the mean model, the PLEs have very competitive performances. Though our method is not designed with specific consideration for enhancing the mean model estimation incorporating cor-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

relations from the longitudinal data, their performance is very close to those of the full likelihood and GEE methods. When the sample size is smaller, the PLEs even outperform the GEE with unstructured correlations, showing the advantage of using parsimonious correlation models.

Table 1: Simulation results for Study 1.Mean bias (MB) and standard deviation (SD) of each parameter us reported. SE is the average standard error calculated using the formula in Theorem 2. PL: Partial Likelihood; FL: Full Likelihood; GEE: Generalized Estimating Equation.

	Pairwise Likelihood			Full Likelihood			GEE		
n	50	100	200	50	100	200	50	100	200
	Case I								
MB_{β_0}	-0.007	-0.003	0.001	-0.006	-0.005	0.001	-0.014	-0.007	-0.001
SD	(0.073)	(0.046)	(0.034)	(0.071)	(0.046)	(0.033)	(0.076)	(0.051)	(0.034)
SE	0.069	0.049	0.034	-	-	-	-	-	-
Std	(0.005)	(0.002)	(0.001)	-	-	-	-	-	-
MB_{β_1}	-0.002	-0.001	0.000	-0.002	-0.001	0.000	-0.005	-0.001	0.001
SD	(0.033)	(0.022)	(0.015)	(0.031)	(0.021)	(0.014)	(0.037)	(0.021)	(0.016)
SE	0.032	0.023	0.016	-	-	-	-	-	-
Std	(0.004)	(0.002)	(0.001)	-	-	-	-	-	-
MB_{β_2}	0.002	0.001	0.000	0.002	0.001	0.000	0.003	0.002	0.000
SD	(0.034)	(0.022)	(0.016)	(0.032)	(0.020)	(0.015)	(0.038)	(0.021)	(0.015)
SE	0.032	0.023	0.016	-	-	-	-	-	-
Std	(0.004)	(0.002)	(0.001)	-	-	-	-	-	-
MB_{γ_0}	0.001	-0.001	-0.004	-0.039	-0.046	-0.047	-	-	-
SD	(0.119)	(0.078)	(0.056)	(0.069)	(0.050)	(0.036)	-	-	-
SE	0.090	0.063	0.044	-	-	-	-	-	-
Std	(0.013)	(0.007)	(0.003)	-	-	-	-	-	-
MB_{γ_1}	-0.023	-0.011	0.031	0.304	0.328	0.350	-	-	-
SD	(0.688)	(0.462)	(0.330)	(0.301)	(0.241)	(0.181)	-	-	-
SE	0.477	0.332	0.232	-	-	-	-	-	-
Std	(0.088)	(0.048)	(0.023)	-	-	-	-	-	-
MB_{γ_2}	0.058	0.035	-0.024	-0.359	-0.378	-0.407	-	-	-
SD	(0.814)	(0.555)	(0.391)	(0.340)	(0.279)	(0.212)	-	-	-
SE	0.558	0.385	0.268	-	-	-	-	-	-
Std	(0.116)	(0.063)	(0.032)	-	-	-	-	-	-
	Case II								
MB_{β_0}	-0.002	-0.002	-0.003	-0.004	-0.001	-0.002	-0.006	-0.004	-0.005
SD	(0.071)	(0.053)	(0.0360)	(0.067)	(0.050)	(0.034)	(0.087)	(0.052)	(0.034)
SE	0.074	0.052	0.036	-	-	-	-	-	-
Std	(0.006)	(0.003)	(0.001)	-	-	-	-	-	-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

MB_{β_1}	0.001	-0.001	-0.001	0.001	-0.000	-0.000	-0.003	-0.002	-0.002
SD	(0.034)	(0.026)	(0.018)	(0.033)	(0.025)	(0.017)	(0.065)	(0.025)	(0.019)
SE	0.036	0.026	0.018	-	-	-	-	-	-
Std	(0.005)	(0.002)	(0.001)	-	-	-	-	-	-
MB_{β_2}	-0.001	0.001	0.001	-0.001	0.001	0.000	-0.000	-0.001	0.001
SD	(0.035)	(0.025)	(0.018)	(0.033)	(0.024)	(0.017)	(0.054)	(0.026)	(0.019)
SE	0.036	0.023	0.018	-	-	-	-	-	-
Std	(0.005)	(0.002)	(0.001)	-	-	-	-	-	-
MB_{γ_0}	0.015	-0.001	-0.003	-0.037	-0.049	-0.048	-	-	-
SD	(0.132)	(0.099)	(0.065)	(0.077)	(0.053)	(0.041)	-	-	-
SE	0.110	0.076	0.054	-	-	-	-	-	-
Std	(0.017)	(0.009)	(0.004)	-	-	-	-	-	-
MB_{γ_1}	-0.084	-0.011	0.009	0.326	0.372	0.362	-	-	-
SD	(0.795)	(0.580)	(0.388)	(0.298)	(0.195)	(0.173)	-	-	-
SE	0.588	0.406	0.288	-	-	-	-	-	-
Std	(0.117)	(0.060)	(0.030)	-	-	-	-	-	-
MB_{γ_2}	0.132	0.034	-0.005	-0.386	-0.441	-0.442	-	-	-
SD	(0.963)	(0.689)	(0.464)	(0.347)	(0.209)	(0.189)	-	-	-
SE	0.700	0.479	0.338	-	-	-	-	-	-
Std	(0.162)	(0.080)	(0.043)	-	-	-	-	-	-

We assessed the finite sample performance of the approximation results in Theorem 3 by testing $H_0 : \beta_2 = 0$ and $H_0 : \gamma_0 = 0$ under simulation setup case I. Figure 4 (a) and (b) display the power functions by the proposed pairwise likelihood ratio testing procedure with a nominal level 0.05. Here the size of the test is well maintained at the nominal level and the power of the test increases when the true parameter value deviates from that in the null hypothesis. To examine the finite sample distribution under the null provided by Theorem 3, Figure 4 (c) shows the Q-Q plot of $LRT = 2\{pl(\hat{\boldsymbol{\theta}}) - pl(\tilde{\boldsymbol{\theta}})\}$ based on 500 simulated data sets with sample size $n = 50$, for testing $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_{1,0}$ with $\boldsymbol{\theta}_1 = (\beta_2, \gamma_0)^{\mathrm{T}}$ and $\boldsymbol{\theta}_{1,0} = (0, 0)^{\mathrm{T}}$. The estimated null distribution is found to be $4.81\chi^2_1 + 0.94\chi^2_1$, where each

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

eigenvalue is the average of 500 eigenvalues, one from each simulation. We treated this distribution as the null distribution and obtained its quantile via simulation as the theoretical quantiles. We plotted them against the observed quantiles from the 500 pairwise likelihood ratio statistics. There is a close agreement between these two sets of quantiles, even though the sample size $n = 50$ is fairly small.

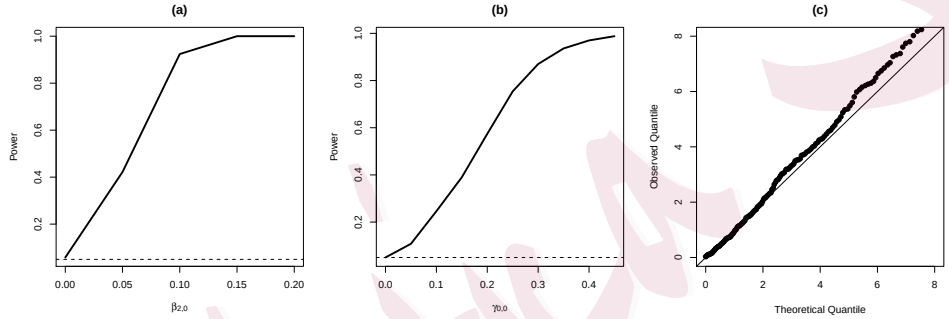

Figure 4: (a) The power function for testing $H_0 : \beta_2 = 0$; (b) The power function for testing $H_0 : \gamma_0 = 0$; (c) Quantile-Quantile plot of the pairwise likelihood ratio statistics relative to the mixture of χ^2_1 distributions as in Theorem 3. The dashed horizontal lines are at the 0.05 nominal level.

Study 2. The data sets were generated from the model

$$y_{ij} \sim Bernoulli(p_{ij}), \qquad \text{logit}(p_{ij}) = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2,$$

$$\omega_{ijk} = \pi/2 - \text{atan}(\gamma_0 + w_{ijk1}\gamma_1 + w_{ijk2}\gamma_2), \ (i = 1, \ldots, n; j = 1, \ldots, m_i),$$

where again $m_i \equiv 6$ for case I and $m_i - 1 \sim Binomial(6, 0.8)$ for case II. The measurement times t_{ij} were uniform. We set $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

$(1.0, -0.5, 0.5)$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2) = (0.5, -0.3, 0.5)$. The covariate x_{ij} was generated again from a standard normal distribution and we took $\mathbf{w}_{ijk} = \{1, t_{ij} - t_{ik}, (t_{ij} - t_{ik})^2\}^{\mathrm{T}}$. Table 2 shows the results qualitatively similar to those in Study 1.

Table 2: Simulation results for Study 2. Mean bias (MB) and standard deviation (SD) of each parameter us reported. SE is the average standard error calculated using the formula in Theorem 2. PL: Partial Likelihood; FL: Full Likelihood; GEE: Generalized Estimating Equation.

	Pairwise Likelihood			Full Likelihood			GEE		
n	50	100	200	50	100	200	50	100	200
	Case I								
MB_{β_0}	0.009	0.016	0.005	0.029	0.033	0.023	0.0311	0.033	0.014
SD	(0.234)	(0.153)	(0.111)	(0.227)	(0.147)	(0.105)	(0.280)	(0.160)	(0.112)
SE	0.220	0.156	0.110	-	-	-	-	-	-
Std	(0.016)	(0.008)	(0.004)	-	-	-	-	-	-
MB_{β_1}	-0.014	-0.006	-0.002	-0.017	-0.011	-0.005	0.021	-0.001	0.003
SD	(0.152)	(0.111)	(0.076)	(0.144)	(0.107)	(0.072)	(0.168)	(0.112)	(0.072)
SE	0.147	0.104	0.073	-	-	-	-	-	-
Std	(0.018)	(0.009)	(0.004)	-	-	-	-	-	-
MB_{β_2}	0.021	0.004	0.006	0.025	0.008	0.010	-0.013	-0.004	0.001
SD	(0.153)	(0.114)	(0.077)	(0.146)	(0.107)	(0.072)	(0.167)	(0.112)	(0.073)
SE	0.148	0.104	0.073	-	-	-	-	-	-
Std	(0.017)	(0.009)	(0.004)	-	-	-	-	-	-
MB_{γ_0}	-0.005	-0.004	0.004	-0.056	-0.048	-0.048	-	-	-
SD	(0.266)	(0.179)	(0.119)	(0.141)	(0.095)	(0.065)	-	-	-
SE_{Std}	0.203	0.143	0.100	-	-	-	-	-	-
Std	(0.039)	(0.019)	(0.008)	-	-	-	-	-	-
MB_{γ_1}	0.003	0.046	-0.013	0.343	0.329	0.324	-	-	-
SD	(1.562)	(1.031)	(0.728)	(0.495)	(0.270)	(0.199)	-	-	-
SE	1.042	0.721	0.505	-	-	-	-	-	-
Std	(0.205)	(0.106)	(0.051)	-	-	-	-	-	-
MB_{γ_2}	0.139	-0.006	0.037	-0.338	-0.368	-0.365	-	-	-
SD	(1.919)	(1.251)	(0.871)	(0.504)	(0.272)	(0.196)	-	-	-
SE	1.232	0.837	0.583	-	-	-	-	-	-
Std	(0.276)	(0.137)	(0.068)	-	-	-	-	-	-
	Case II								
MB_{β_0}	0.013	0.014	-0.002	0.024	0.031	0.017	0.044	0.030	0.007
SD	(0.240)	(0.166)	(0.117)	(0.224)	(0.157)	(0.106)	(0.244)	(0.169)	(0.115)
SE	0.233	0.166	0.118	-	-	-	-	-	-
Std	(0.020)	(0.010)	(0.005)	-	-	-	-	-	-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

MB_{β_1}	-0.014	-0.006	-0.002	-0.017	-0.006	-0.005	-0.005	0.002	-0.001
SD	(0.168)	(0.116)	(0.084)	(0.166)	(0.114)	(0.0768)	(0.177)	(0.116)	(0.080)
SE	0.165	0.117	0.082	-	-	-	-	-	-
Std	(0.024)	(0.011)	(0.005)	-	-	-	-	-	-
MB_{β_2}	0.005	0.010	0.004	0.011	0.013	0.007	-0.005	0.004	0.002
SD	(0.174)	(0.120)	(0.084)	(0.166)	(0.115)	(0.080)	(0.175)	(0.119)	(0.081)
SE	0.166	0.117	0.082	-	-	-	-	-	-
Std	(0.022)	(0.011)	(0.005)	-	-	-	-	-	-
MB_{γ_0}	0.009	-0.009	-0.008	-0.043	-0.058	-0.054	-	-	-
SD	(0.329)	(0.207)	(0.140)	(0.172)	(0.109)	(0.073)	-	-	-
SE	0.240	0.166	0.117	-	-	-	-	-	-
Std	(0.052)	(0.023)	(0.011)	-	-	-	-	-	-
MB_{γ_1}	-0.032	0.004	0.054	0.315	0.035	0.354	-	-	-
SD	(2.001)	(1.207)	(0.833)	(0.553)	(0.109)	(0.194)	-	-	-
SE	1.249	0.869	0.604	-	-	-	-	-	-
Std	(0.260)	(0.126)	(0.064)	-	-	-	-	-	-
MB_{γ_2}	0.164	0.095	-0.022	-0.334	-0.363	-0.392	-	-	-
SD	(2.531)	(1.558)	(1.011)	(0.587)	(0.3522)	(0.167)	-	-	-
SE	1.497	1.024	0.709	-	-	-	-	-	-
Std	(0.346)	(0.173)	(0.085)	-	-	-	-	-	-

Study 3. This is a study designed to investigate the impact on the mean model estimation from a misspecified correlation model. For such a purpose, we generated data from a random effect Poisson regression model

$$y_{ij} \sim Pois(\lambda_{ij}), \quad log(\lambda_{ij}) = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + z_{ij} b_i$$

where $b_i \sim N(0, \sigma_b^2)$ is a random effect accounting for the correlations. We took $\beta = (1, 0.5, -0.5)'$ and $\sigma_b = 0.8$. The covariates x_{ij1} and x_{ij2} were standard normal, $z_{ij} \sim Uniform(0, 1)$. The number of repeated measurements was 6. We applied the cubic polynomial of time lag for our approach when modeling the correlations, and we compared our approach with the

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

GEE method with different specifications of the working correlation structures. For this setting, the model is mis-specified for both our method and the GEE method. The simulation results are summarized in Table 3. From the results, we can see that our method performs competitively, even when the correlation structure is not correctly specified. Specifically, when sample sizes are small, our method consistently performs the best with the smallest MSE. When sample size is larger, the GEE with unstructured covariance specification works very well. When the sample size is smaller, at $n = 50$, the GEE with unstructured covariance specification has a high level of variation due to unstable covariance estimations. Overall, our method performs promisingly, indicating the potential benefit for estimating the mean model incorporating the correlations between the longitudinal data from using a parsimonious correlation model.

In simulation results not reported here, we found substantial improvement of our method compared with the GEE with working independence. We also found that inferences including estimations and hypothesis testing to be highly effective using the pairwise likelihood instead of using the computationally intractable full likelihood. Using our mean-correlation regression approach with pairwise likelihood-based inferences could provide a powerful and convenient device for analyzing generic discrete longitudinal

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

Table 3: Simulation results. Mean bias (MB) and Mean square error (MSE) of each parameter is reported under different sample sizes and models. PL: pairwise likelihood approach; GEE: generalized estimating equations; Ind: Independent working correlation; AR: AR(1) working correlation; Unstr: Unstructured working correlation. All results are multiplied by 100.

		n	MB_{β_0}	MSE	MB_{β_1}	MSE	MB_{β_2}	MSE
		50	10.09	2.04	-0.26	0.45	-0.12	1.57
PL		100	11.28	1.66	-0.21	0.21	0.15	0.93
		150	9.72	1.27	-0.54	0.12	0.12	0.48
		50	10.27	2.17	0.26	0.54	-0.3	1.88
	Ind	100	11.27	1.68	0.08	0.21	-0.75	1.06
		150	10.68	1.51	-0.42	0.15	-0.41	0.59
		50	10.50	2.16	-0.13	0.45	-0.31	1.67
GEE	AR	100	11.29	1.65	-0.08	0.17	-1.06	0.85
		150	10.67	1.48	-0.49	0.13	-0.08	0.42
		50	8.43	6.12	-1.30	3.65	0.37	6.24
	Unstr	100	10.78	1.54	-0.02	0.17	-1.02	0.91
		150	10.23	1.37	-0.42	0.13	-0.17	0.45

data in practice.

5. Conclusion

The problem of developing regression models for correlation structures is an open problem when longitudinal responses are discrete. We propose the first model of this kind to address the problem. Equipped with the new parametrization of a correlation matrix in a copula model that enables unconstrained model building and a computationally efficient estimation method based on pairwise likelihood, we have developed a tool for investigating correlated responses.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

This paper focuses mainly on univariate discrete responses. It would be interesting to generalize the univariate models to situations where multiple mixed outcomes are available at each time point (Xu and Mackenzie (2012)). One way to simplify the multiple response time-dependent covariance is to factorize the covariance matrices via a Kronecker product decomposition that greatly reduces the dimensionality. This will be studied in a future paper. Another interesting problem is to develop model diagnostic tools for assessing model adequacy, especially for unbalanced data. For balanced data, as illustrated in the paper, graphical tools to compare the empirical estimates and the model estimates, such as those used for analyzing the toenail data and the epileptic seizure data, are useful. However, counterparts of those are not currently available when data are unbalanced. Another future line of research is to develop data-driven models for covariations.

Supplementary Materials

The online supplementary materials contain the proofs of Theorem 1, 2 and 3 in the main paper, additional data analysis and simulations studies.

Acknowledgements

We thank the Editor, an Associate editor, and two referees for their constructive comments and suggestions that have greatly improved the pa-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

per. Zhang acknowledges support from the National Key Research and Development Plan (No. 2016YFC0800100) and the NSF of China (No. 11671374,71631006). Tang acknowledges support from NSF Grants SES-1533956 and IIS-1546087. Leng was supported by the Alan Turing Institute under the EPSRC grant EP/N510129/1.

References

Bergsma, W., Croon, M., and Hagenaars, J. A. (2009). *Marginal Models for Dependent, Clustered, and Longitudinal Categorical Data*. Springer.

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

Creal, D., Koopman, S. J., and Lucas, A. (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *J. Bus. Econ. Stat.* **29**, 552–563.

Daniels, M. J. and Pourahmadi, M. (2009). Modeling covariance matrices via partial autocorrelations. *J. Multivariate Anal.* **100**, 2352–2363.

De Backer, M., De Keyser, P., De Vroey, C., and Lesaffre, E. (1996). A 12-week treatment for dermatophyte toe onychomycosis: terbinafine 250mg/day vs. itraconazole 200mg/day a double-blind comparative

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

trial. *Br. J. Dermatol.* **134**, 16–7.

Dickson, E. R., Grambsch, P. M., Fleming, T. R., Fisher, L. D. and Langworthy, A. (1989). Prognosis in primary biliary cirrhosis: Model for decision making. *Hepatology* **10**, 1–7.

Diggle, P. J., Heagerty, P., Liang, K. Y., and Zeger, S. L. (2002). Analysis of Longitudinal Data.2nd edition. Oxford University Press, Oxford.

Fan, J., Liu, H., Ning, Y., and Zou, H. (2017), High dimensional semiparametric latent graphical model for mixed data. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **79**, 405 - 421.

Fang, H., Fang, K., and Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* **82**, 1-16.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004). *Applied longitudinal analysis*, Wiley, New York.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

Gao, X. and Song, P.X.K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *J. Amer. Statist. Assoc.* **105**, 1531-1540.

Gaskins, J., Daniels, M.J., and Marcus, B. (2014). Sparsity inducing prior

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

distributions for correlation matrices of longitudinal data. *J. Comput. Graph. Stat.* **23**, 966–984.

Hoffman, L. (2012). Considering alternative metrics of time: Dos anybody really know what "time" is? In G. Hancock and J. R. Harring (Ed.) Advances in longitudinal methods in the social and behavioral sciences. Charlotte, Information Age Publishing, NC.

Liu, H., Lafferty, J. D., and Wasserman, L. A.(2009). The nonparanormal: semiparametric estimation of high dimensional undirected graphs. *J. Mach. Learn. Res.* **10**, 2295-2328.

Leng, C., Zhang, W., and Pan, J. (2010). Semiparametric mean-covariance regression analysis for longitudinal data. *J. Amer. Statist. Assoc.* **105**, 181–193.

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

Lee, Y. and Nelder, J. A. (2006). Double hierarchical generalized linear models (with discussion). *J. R. Stat. Soc. Ser. C-Appl. Stat.* **55**, 139–185.

Lynn, P. (2009). *Methodology of Longitudinal Surveys*, Wiley, New York.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chap-

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

man and Hall/CRC, New York.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer-Verlag.

Muenz, L. R. and Rubinstein, L. V. (1985). Markov models for covariate dependence of binary sequences. *Biometrics* **41**, 91–101.

Pan, J. and Mackenzie, G. (2003). Model selection for joint mean-covariance structures in longitudinal studies. *Biometrika* **90**, 239–244.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika* **86**, 677–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87**, 425–35.

Pourahmadi, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika* **94**, 1006–1013.

Pourahmadi, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Stat. Sci.* **26**, 369–387.

Song, P. X.-K., Li, M., and Yuan, Y. (2009). Joint regression analysis of

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

correlated data using Gaussian copulas. *Biometrics* **65**, 60–68.

Song, P. X. K. (2000). Multivariate Dispersion Models Generated From Gaussian Copula. *Scand. J. Stat.* **27**, 305 - 320.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with over-dispersion. *Biometrics* **46**, 657–671.

Tong, Y. L. (1990). *The Multivariate Normal Distribution*, Springer.

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

Wang, Y. and Daniels, M. J. (2013). Bayesian modeling of the dependence in longitudinal data via partial autocorrelations and marginal variances. *J. Multivar. Anal.* **116**, 130–140.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

Xu, J. and Mackenzie, G. (2012). Modelling covariance structure in bivariate marginal models for longitudinal data. *Biometrika* **99**, 649–662.

Ye, H. and Pan, J. (2006). Modelling covariance structures in generalized estimating equations for longitudinal data. *Biometrika* **93**, 927–941.

Zeger, S. L., Liang, K. Y., and Self, S. G. (1985). The analysis of binary longitudinal data with timeindependent covariates. *Biometrika* **72**, 31–

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0435

38.

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

Zhang, W. and Leng, C. (2012). A moving average cholesky factor model in covariance modeling for longitudinal data. *Biometrika* **99**, 141–150.

Zhang, W., Leng, C., and Tang, C. Y. (2015). A joint modeling approach for longitudinal studies. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* **77**, 219–238.

Department of Statistical Science, Temple University, Philadelphia, PA 19122, USA.

E-mail: yongtang@temple.edu

Department of Statistics and Finance, University of Science and Technology of China, Hefei 230026, P.R.China.

E-mail: zwp@ustc.edu.cn(Corresponding author)

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK.

E-mail: C.Leng@warwick.ac.uk