Computational Statistics and Data Analysis 106 (2017) 1–11

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Generalized estimating equations with stabilized working correlation structure

Yongchan Kwon[a], Young-Geun Choi[a], Taesung Park[a], Andreas Ziegler[b,c,d], Myunghee Cho Paik[a,*]

[a] Seoul National University, Seoul, Republic of Korea

[b] Institute of Medical Biometry and Statistics, University of Lübeck, University Medical Center Schleswig-Holstein, Campus Lübeck, Germany

[c] Center for Clinical Trials, ZKS Lübeck, University of Lübeck, Germany

[d] School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 19 August 2016
Accepted 26 August 2016
Available online 1 September 2016

Keywords:
Generalized estimating equations
Working correlation
Unstructured correlation matrix
Positive definiteness
Linear shrinkage

ABSTRACT

Generalized estimating equations (GEE) proposed by Liang and Zeger (1986) yield a consistent estimator for the regression parameter without correctly specifying the correlation structure of the repeatedly measured outcomes. It is well known that the efficiency of regression coefficient estimator increases with correctly specified working correlation and thus unstructured correlation could be a good candidate. However, lack of positive-definiteness of the estimated correlation matrix in unbalanced case causes practitioners to choose independent, autoregressive or exchangeable matrices as working correlation structure. Our goal is to broaden practical choices of working correlation structure to unstructured correlation matrix or any other matrices by proposing a GEE with a stabilized working correlation matrix via linear shrinkage method in which the minimum eigenvalue is forced to be bounded below by a small positive number. We show that the resulting regression estimator of GEE is asymptotically equivalent to that of the original GEE. Simulation studies show that the proposed modification can stabilize the variance of the GEE regression estimator with unstructured working correlation, and improve efficiency over popular choices of working correlation. Two real data examples are presented where the standard error of the regression coefficient estimator can be reduced using the proposed method.

1. Introduction

Generalized estimating equations (GEE) proposed by Liang and Zeger (1986) have been a popular analytic tool for correlated data. A consistent estimator for the regression parameter can be achieved without correctly specifying the correlation structure of the repeatedly measured outcomes. However, the efficiency of regression coefficient estimator increases if the working correlation matrix is close to the true one (Albert and McShane, 1995). Structured working correlations such as independent, autoregressive and exchangeable are available from built-in functions from software. These choices give a manageable number of parameters in the correlation matrix, and can be helpful when the sample size is small and the number of time points is large. To select a working correlation matrix from various choices, criteria such

* Corresponding author.
E-mail address: myungheechopaik@snu.ac.kr (M.C. Paik).

http://dx.doi.org/10.1016/j.csda.2016.08.016
0167-9473/

as the 'quasi-likelihood under the independence model criterion' (Pan, 2001) and the 'correlation information criterion' (Hin and Wang, 2009) have been proposed among others (Carey and Wang, 2011; Gosho et al., 2011; Zhou et al., 2012; Westgate, 2013, 2014). The unstructured working correlation matrix can correctly model the correlation structure and is available from built-in functions from software, but the number of unknown parameters increases as the number of time points. When the sample size is small relative to the number of time points, variability of many nuisance parameters in the unstructured correlation matrix affects the variance of the regression parameter estimators, and Westgate (2013) proposed a method to address this problem. However, when the maximum of numbers of time points is fixed, the asymptotic variance of the regression coefficient estimator is unaffected by the variance of the correlation estimator, and reducing the number of parameters does not lead to gain in asymptotic efficiency of the regression coefficient estimator. Misspecification of working correlation could not only lead to loss of efficiency, but more seriously, could lead to infeasibility of the GEE solutions (Qu et al., 2008; Wang and Carey, 2004). Despite these shortcomings, choosing aforementioned structured working correlation matrix guarantees the correlation matrix to be positive definite. The estimated unstructured correlation matrix sometimes fails to be positive definite due to varying numbers of subunits, in which case the GEE estimates are not defined. Even when the estimated unstructured matrix is positive definite, if the minimum eigenvalue is small, the coefficient estimate can be unstable and the standard error of regression parameter estimates can be large (Vens and Ziegler, 2012). If lack of positive definiteness can be solved, the unstructured working correlation matrix can be an attractive choice since it improves the asymptotic variance of the regression coefficient estimator.

Many researchers have worked on solving lack of positive-definiteness of the sample covariance matrix mainly by replacing the eigenvalues of sample covariance matrix by their linear or nonlinear transforms (Stein, 1956; Haff, 1991; Daniels and Kass, 1999, 2001; Ledoit and Wolf, 2004; Schäfer and Strimmer, 2005; Ledoit and Wolf, 2012; Won et al., 2013; Lam, 2016). In a regression setting with longitudinal data, Daniels and Kass (2001) obtained stabilized regression coefficients estimators by placing a normally-distributed prior to the logarithm of the sample eigenvalues. This method requires that the eigenvalues of the sample covariance matrix are positive.

Our goal is to broaden practical choices of working correlation structure to unstructured correlation matrix by alleviating problems due to lack of positive definiteness. To achieve this goal we propose to modify working correlation matrix by linear shrinkage method proposed by Choi (2015). We show that the resulting regression estimator of GEE is asymptotically equivalent to that of the original GEE. Simulation studies show that the proposed modification has advantages in cases where the minimum eigenvalue of the estimated working correlation structure is small. Two real data examples are presented where the standard error of the regression coefficient estimator is reduced using the proposed method.

2. Basic notations

We denote the $n_i \times 1$ vector of the outcomes and the $n_i \times p$ matrix of covariates for the ith subject $(i = 1, \ldots, K)$ by $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ and $\mathbf{X}_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^T$, respectively. We assume that the first two moments of y_{ij} are given by

$$E(y_{ij} \mid x_{ij}) = \mu_{ij} = g(\eta_{ij}) = g(x_{ij}^T\boldsymbol{\beta}), \quad \text{and} \quad Var(y_{ij} \mid x_{ij}) = \phi a(\mu_{ij}),$$

where $\boldsymbol{\beta}$ is a $p \times 1$ regression parameter, and $g^{-1}(\cdot)$ is a link function. The true $n_i \times n_i$ covariance matrix of $\mathbf{y}_i$ given $\mathbf{X}_i$, $Var(\mathbf{y}_i \mid \mathbf{X}_i)$ is denoted by $\boldsymbol{\Omega}_i$. Let the maximum of n_i be q, and assume that q is bounded. The working correlation matrix for q repeated outcomes is denoted by $\mathbf{R}(\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is an $s \times 1$ vector fully characterizing $\mathbf{R}(\boldsymbol{\alpha})$. When the working correlation matrix is unstructured, $\boldsymbol{\alpha}$ can be $q^2 \times 1$ vectorized elements of $\mathbf{R}(\boldsymbol{\alpha})$. We denote by $\mathbf{R}_i(\boldsymbol{\alpha})$ the ith sub-matrix of $\mathbf{R}(\boldsymbol{\alpha})$ extracted according to the corresponding indices, and write $\boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{A}(\boldsymbol{\mu}_i)^{1/2}\mathbf{R}_i(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\mu}_i)^{1/2}$, where $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i})^T$, and $\mathbf{A}(\boldsymbol{\mu}_i)$ is a diagonal matrix with $a(\mu_{ij})$ as the jth diagonal element. Assume that we have $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}_0$ that satisfy $K^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) = O_p(1)$. The limit of $\hat{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha}_0$, is determined by the value that satisfies the expectation of the estimating function for $\boldsymbol{\alpha}$ being zero. When the true and specified correlation structures are different, $\boldsymbol{\alpha}_0$ could be different depending on the estimating function for $\boldsymbol{\alpha}$, which leads to different asymptotic relative efficiency (Wang and Carey, 2003). A lack of definition of $\boldsymbol{\alpha}_0$ when working correlation is different from the true correlation is discussed in Crowder (1995). The GEE estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is obtained by solving GEE,

$$\mathbf{U}\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\} = \sum_{i=1}^{K} \mathbf{U}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\} = \sum_{i=1}^{K} \mathbf{D}_i^T \boldsymbol{\Sigma}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0},$$

where $\mathbf{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}^T$. Let $\mathbf{W}_0(\boldsymbol{\beta}, \boldsymbol{\alpha}) = E(-K^{-1}\partial\mathbf{U}/\partial\boldsymbol{\beta}^T)$, $Var\{K^{-\frac{1}{2}}\mathbf{U}(\boldsymbol{\beta}, \boldsymbol{\alpha})\}$ be $\mathbf{W}_1(\boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\mathbf{W}_1(\boldsymbol{\beta}_0, \boldsymbol{\alpha}) = E\left\{K^{-1}\sum_{i=1}^{K}\mathbf{D}_i^T\boldsymbol{\Sigma}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}^{-1}\mathbf{D}_i\right\}$. Notation $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ emphasizes that $\hat{\boldsymbol{\alpha}}$ is a function of $\boldsymbol{\beta}$. Under some conditions, $K^{\frac{1}{2}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ is shown to be asymptotically normal with mean $\mathbf{0}$ and variance $\mathbf{W}_0^{-1}\mathbf{W}_1\mathbf{W}_0^{-1}$ (Liang and Zeger, 1986).

3. Motivation

To motivate the proposed method, we first quantify the loss of the asymptotic relative efficiency (ARE) by limiting the choice of working correlation structure to exchangeable and autoregressive of order 1 (AR-1). This quantification

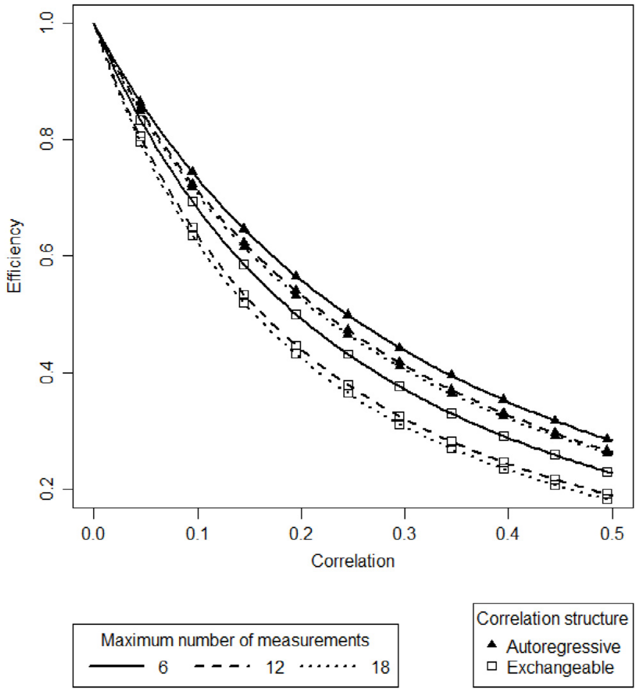

Fig. 1. Efficiency loss curves by number of measurements and correlation structures. Maximum number of measurements are from 6 to 18 and correlation structure is either autoregressive or exchangeable.

demonstrates a range of potential improvement in efficiency by using the unstructured working correlation structure. Second, we quantify prevalence of a negative minimum eigenvalue for the estimated unstructured correlation matrix.

First, Fig. 1 shows how much efficiency is lost for the intercept in the linear model by choosing exchangeable or AR-1 instead of unstructured correlation structure when the true correlation structure is 1-dependent and the number of measurements per independent unit is distributed as uniform from 2 to q. Specifically, we set the linear model with mean $\mu_{ij} = x_{ij}^T \boldsymbol{\beta}$ where $x_{ij}^T = (1, z_i - \bar{z})$, z_i is independently identically distributed, and $\bar{z} = K^{-1} \sum_{i=1}^{K} z_i$. In this case, the limit of the off-diagonal element of the information matrix is zero, and the covariate does not affect the ARE of the intercept estimator. The y-axis of the figure indicates the ARE of the estimated intercept, the first diagonal element of $\mathbf{W}_0^{*-1}(\mathbf{W}_0^{-1}\mathbf{W}_1\mathbf{W}_0^{-1})^{-1}$, where $\mathbf{W}_0^*$ is evaluated using the true correlation structure, and $\mathbf{W}_0$ and $\mathbf{W}_1$ are evaluated at exchangeable or AR-1 structure. The limit of $\hat{\boldsymbol{\alpha}}$ for exchangeable or AR-1 working correlation structure is obtained to satisfy the expectation of the general moment-based estimating equation being zero (Wang and Carey, 2003). The efficiency loss increases as the true correlation size and the number of measurements increase. Since the ARE is computed when the true correlation is 1-dependent, the efficiency loss is more severe in using exchangeable since AR-1 structure is closer to the true structure. Even when the correlation is as low as 0.2, efficiency loss can be as large as 0.5.

In Fig. 2, we show frequencies out of 1000 replications that the sample unstructured correlation matrix is not positive definite when multivariate normal outcomes with marginally standard normal are randomly generated with true correlation structure as 1-dependent, and the number of measurements per independent unit is uniformly distributed from 2 to q. Sample correlation matrix is obtained using available data in a pairwise fashion, and the resulting matrix is not guaranteed to be positive definite. Although not shown, when the number of time points is 5, and the correlation is between 0 and 0.5, the sample correlation was positive definite over 99%. When the number of time points is as low as 10 and the correlation is between 0 and 0.4, the sample correlation is positive definite with over 95%. As the correlation and the number of time points increase, the sample correlation matrix becomes non-positive definite with higher frequency. When the number of time points is as large as 18, even with correlation of 0.2, the sample correlation matrix has a negative minimum eigenvalue over 80% frequency. When the working correlation is not positive definite, the solution for GEE is not be defined, and popular packages do not provide estimates. In the next section, we provide a principled approach of handling working correlation with negative or small eigenvalues in GEE. The main idea of the method is to linearly shrink the estimated working correlation matrix so that the minimum eigenvalue is bounded below by a small positive number.

4. GEE using the working correlation matrix with linear shrinkage

Let ϵ be a positive number smaller than minimum of eigenvalues of true correlation matrix and $\gamma_{\min}(\mathbf{A})$ and $\gamma_{\max}(\mathbf{A})$ be minimum and maximum eigenvalues of the matrix $\mathbf{A}$, respectively. Choi (2015) proposed the linear shrinkage

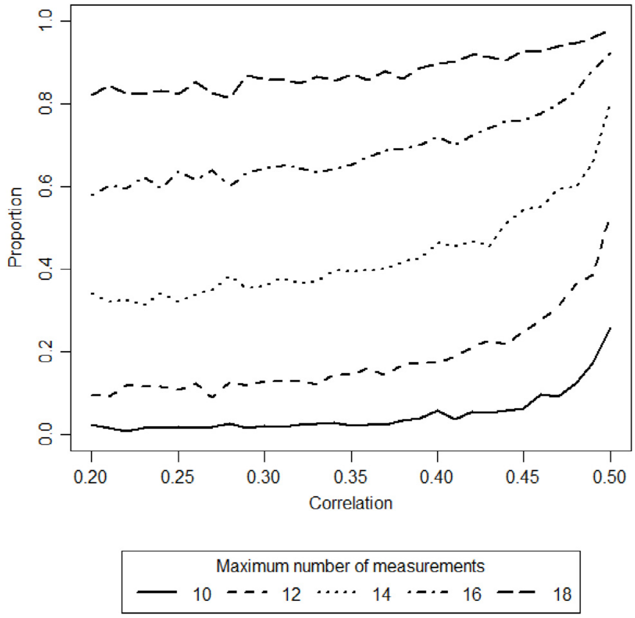

Fig. 2. Proportion of the sample unstructured correlation having negative minimum eigenvalue out of 1000 replications. Maximum number of measurements are from 10 to 18.

transformation of a symmetric matrix given by

$$\tilde{\mathbf{R}}(\boldsymbol{\alpha}, \epsilon) = t\mathbf{R}(\boldsymbol{\alpha}) + (1-t)\nu\mathbf{I}_q,$$

where

$$t = \begin{cases} 1 & \text{if } \gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\} \geq \epsilon \\ (\nu - \epsilon)/(\nu - \gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\}) & \text{if } \gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\} < \epsilon, \end{cases}$$

$$\nu = \max\left\{\frac{\gamma_{\max}\{\mathbf{R}(\boldsymbol{\alpha})\} + \gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\}}{2}, \ M + \frac{V}{M - \gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\}}\right\},$$

M and V denote the sample mean and sample variance of the eigenvalues of $\mathbf{R}(\boldsymbol{\alpha})$, respectively. Choi (2015) showed that $\tilde{\mathbf{R}}(\hat{\boldsymbol{\alpha}}, \epsilon)$, a linearly shrunk version of $\mathbf{R}(\hat{\boldsymbol{\alpha}})$, is guaranteed to be positive definite with its minimum eigenvalue being larger than or equal to ϵ. Furthermore, it is asymptotically equivalent to $\mathbf{R}(\hat{\boldsymbol{\alpha}})$ in a sense of

$$\left\|\tilde{\mathbf{R}}(\hat{\boldsymbol{\alpha}}, \epsilon) - \mathbf{R}(\boldsymbol{\alpha}_0)\right\| \leq C \left\|\mathbf{R}(\hat{\boldsymbol{\alpha}}) - \mathbf{R}(\boldsymbol{\alpha}_0)\right\| \tag{1}$$

in both spectral and Frobenius norms for a constant C. We propose the following GEE estimator $\tilde{\boldsymbol{\beta}}(\epsilon)$ of $\boldsymbol{\beta}$, the solution of

$$\tilde{U}\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \epsilon\} = \sum_{i=1}^{K} \tilde{U}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \epsilon\} = \sum_{i=1}^{K} \mathbf{D}_i^T \tilde{\boldsymbol{\Sigma}}_i\{\boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \epsilon\}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = 0 \tag{2}$$

where $\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}, \epsilon) = \mathbf{A}(\boldsymbol{\mu}_i)^{1/2}\tilde{\mathbf{R}}_i(\boldsymbol{\alpha}, \epsilon)\mathbf{A}(\boldsymbol{\mu}_i)^{1/2}$ and $\tilde{\mathbf{R}}_i(\boldsymbol{\alpha}, \epsilon)$ is defined by ith sub-matrix of $\tilde{\mathbf{R}}(\boldsymbol{\alpha}, \epsilon)$ according to corresponding indices. Computation of $\tilde{\boldsymbol{\beta}}(\epsilon)$ involves a one-step update from $\hat{\boldsymbol{\beta}}$ given ϵ.

We show that the modified GEE estimator $K^{1/2}\{\tilde{\boldsymbol{\beta}}(\epsilon) - \boldsymbol{\beta}_0\}$ obtained from Eq. (2) is asymptotically equivalent to $K^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$.

Theorem 1. *Under mild regularity conditions stated in Liang and Zeger (1986), $K^{1/2}\{\tilde{\boldsymbol{\beta}}(\epsilon) - \boldsymbol{\beta}_0\}$ is asymptotically multivariate Gaussian with zero mean and covariance matrix* $\mathbf{V}$ *given by*

$$\mathbf{V} = \lim_{K\to\infty} K\left(\sum_{i=1}^{K} \mathbf{D}_i^T \breve{\boldsymbol{\Sigma}}_i^{-1}\mathbf{D}_i\right)^{-1}\left\{\sum_{i=1}^{K} \mathbf{D}_i^T \breve{\boldsymbol{\Sigma}}_i^{-1}\operatorname{cov}(\mathbf{y}_i)\breve{\boldsymbol{\Sigma}}_i^{-1}\mathbf{D}_i\right\}\left(\sum_{i=1}^{K} \mathbf{D}_i^T \breve{\boldsymbol{\Sigma}}_i^{-1}\mathbf{D}_i\right)^{-1}$$

where $\breve{\boldsymbol{\Sigma}}_i = \tilde{\boldsymbol{\Sigma}}_i\{\tilde{\boldsymbol{\beta}}(\epsilon), \hat{\boldsymbol{\alpha}}, \epsilon\}$.

A proof is sketched in the Appendix. It is noteworthy that the diagonal elements of $\tilde{\mathbf{R}}(\hat{\boldsymbol{\alpha}}, \epsilon)$ may not be one, but nonetheless, $\tilde{\boldsymbol{\beta}}(\epsilon)$ has similar property with $\hat{\boldsymbol{\beta}}$ due to the asymptotic equivalence between $\tilde{\mathbf{R}}(\hat{\boldsymbol{\alpha}}, \epsilon)$ and $\mathbf{R}(\hat{\boldsymbol{\alpha}})$.

Table 1
Simulation results of Scenario 1. Comparing bias and standard error of estimates using five different working correlation matrices based on 1000 Monte Carlo samples. Standard errors are presented in parenthesis.

	IND	AR-1	EXC	UNS	PRO
β_1	-2.4×10^{-3} (8.8×10^{-2})	-2.5×10^{-3} (8.8×10^{-2})	-2.4×10^{-3} (8.8×10^{-2})	1.2×10^{-2} (6.0×10^{-1})	-2.4×10^{-3} (8.8×10^{-2})
β_2	1.1×10^{-3} (1.2×10^{-1})	1.3×10^{-3} (1.2×10^{-1})	1.1×10^{-3} (1.2×10^{-1})	-2.9×10^{-2} (1.0)	9.2×10^{-4} (1.2×10^{-1})
β_3	2.4×10^{-6} (1.2×10^{-3})	-2.5×10^{-6} (9.6×10^{-4})	2.4×10^{-6} (1.2×10^{-3})	-1.8×10^{-3} (5.0×10^{-2})	-1.3×10^{-5} (7.6×10^{-4})
β_4	-1.0×10^{-4} (1.7×10^{-3})	-6.8×10^{-5} (1.3×10^{-3})	-1.0×10^{-4} (1.7×10^{-3})	2.7×10^{-3} (8.9×10^{-2})	-4.6×10^{-5} (1.0×10^{-3})
TR	2.42×10^{-2}	2.43×10^{-2}	2.42×10^{-2}	1.52	2.41×10^{-2}

IND, Independent; AR-1, Autoregressive of order 1; EXC, Exchangeable; UNS, Unstructured; PRO, Proposed; TR, Trace of the estimated covariance.

In implementing the method in practice, we should specify ϵ. We suggest finding the optimal ϵ empirically by minimizing the trace of estimated covariance matrix of the element(s) of interest in $\tilde{\boldsymbol{\beta}}$. The trace of empirical variance has been advocated to be used as a criterion for model selection in different contexts (Westgate, 2014; Song et al., 2009). If the study has a focused question, one can take a submatrix of the covariance matrix. For example, in clinical trial, ϵ can be chosen to minimize the variance of the estimate of primary interest as shown in Section 6.1.

5. Simulation studies

We conducted simulation studies to evaluate finite sample performance of the proposed estimator in three settings. The first scenario mimics the correlation structure of the data from the randomized controlled trial presented in Section 6.1, the second scenario mimics the correlation structure of the data from the study presented in Section 6.2, and the third considers the case in which the true correlation structure is 1-dependent. In all scenarios the estimated unstructured correlation is unstable due to small minimum eigenvalue of true correlation matrix. In the first scenario, the proposed estimator performs comparably to or slightly better than those from popular choices of structured working correlation structures, but in the second and third scenarios, outperforms. Five GEE estimators were compared using different working correlation matrices which are (i) independent, (ii) AR-1, (iii) exchangeable, (iv) unstructured, and (v) proposed working correlation with linear shrinkage. We present bias, simulation standard error of the estimates based on 1000 Monte Carlo simulation samples.

5.1. Scenario 1

Mimicking the setup of Vens and Ziegler (2012), we set $n_i = 5$, $K = 226$, and $p = 4$. Identity link is used, i.e., $\mu_{ij} = \eta_{ij}$, and $a(\mu_{ij}) = 1$ with $\phi = 1$. We set the mean model

$$E[y_{ij} \mid x_{ij}] = \beta_1 + \beta_2 x_{ij}^{Treat} + \beta_3 x_{ij}^{Time} + \beta_4 x_{ij}^{Treat} x_{ij}^{Time},$$

where $\boldsymbol{\beta} = (2, 3, 1, 1)$, $x_{ij}^{Treat} = 1$ for $i = 1, \ldots, 113, 0$, otherwise, $x_{i1}^{Time} = 4$, $x_{i2}^{Time} = 6$, $x_{i3}^{Time} = 8$, $x_{i4}^{Time} = 12$, and $x_{i5}^{Time} = 16$. The true covariance matrix $\boldsymbol{\Omega}_i$ of $\mathbf{y}_i$ given $\mathbf{X}_i$ is

$$\begin{bmatrix} 1 & 0.9594 & 0.9539 & 0.9593 & 0.9703 \\ 0.9594 & 1 & 0.9973 & 0.9973 & 0.9971 \\ 0.9539 & 0.9973 & 1 & 0.9973 & 0.9973 \\ 0.9593 & 0.9973 & 0.9973 & 1 & 0.9973 \\ 0.9703 & 0.9971 & 0.9973 & 0.9973 & 1 \end{bmatrix}.$$

Eigenvalues are (4.9310, 0.0627, 0.0027, 0.0026, 0.0007).

Table 1 shows that the standard error of the estimators obtained by using the unstructured correlation structure was over 65 fold of the proposed estimator for β_3 since the estimated unstructured covariance matrix was unstable with a small minimum eigenvalue as in the example of Vens and Ziegler (2012). When the proposed method was applied, the modified unstructured working correlation estimate produced estimators with stable variances. The standard error of the proposed estimator was similar to those from exchangeable or AR-1 for β_1 and β_2, but was reduced to 79% and 77% for β_3 and β_4, respectively. The last row shows that the trace of simulation covariance, which is the criterion for selection of working correlation by Westgate (2014), was the smallest for the proposed one.

5.2. Scenario 2

In this setting, we kept the mean structure the same as in Scenario 1, but assumed that the true covariance matrix is the same as the linearly shrinkage transformation of sample covariance obtained from the study presented in Section 6.2. A heat

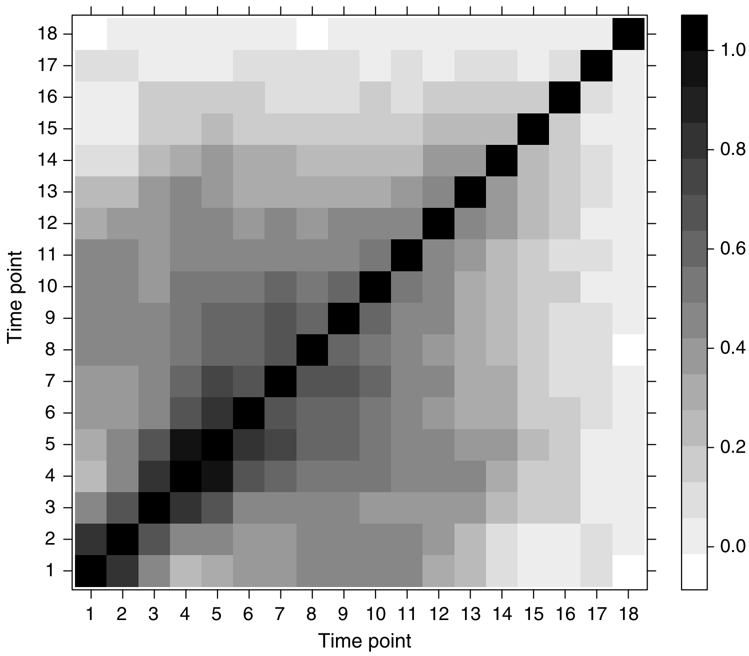

Fig. 3. Heatmap of 18 by 18 true covariance matrix $\boldsymbol{\Omega}_i$ of $\mathbf{y}_i$ given $\mathbf{X}_i$ used in Scenario 2.

Table 2
Simulation results of Scenario 2. Comparing bias and standard error of estimates using five different working correlation matrices based on 1000 Monte Carlo samples. Standard errors are presented in parenthesis.

	IND	AR-1	EXC	UNS	PRO
β_1	-6.2×10^{-3} (1.0×10^{-1})	-5.7×10^{-3} (1.0×10^{-1})	-6.2×10^{-3} (1.0×10^{-1})	-1.1×10^{-4} (6.1×10^{-1})	-4.0×10^{-3} (7.4×10^{-2})
β_2	1.0×10^{-2} (1.4×10^{-1})	9.4×10^{-3} (1.3×10^{-1})	1.0×10^{-2} $(s1.4 \times 10^{-1})$	9.5×10^{-3} (6.2×10^{-1})	7.2×10^{-3} (1.0×10^{-1})
β_3	2.5×10^{-4} (6.1×10^{-3})	2.4×10^{-4} (6.2×10^{-3})	2.5×10^{-4} (6.1×10^{-3})	1.2×10^{-4} (3.5×10^{-2})	1.9×10^{-4} (4.6×10^{-3})
β_4	-5.1×10^{-4} (8.2×10^{-3})	-4.6×10^{-4} (8.4×10^{-3})	-5.1×10^{-4} (8.2×10^{-3})	-7.1×10^{-4} (3.9×10^{-2})	-3.7×10^{-4} (6.5×10^{-3})
TR	3.13×10^{-2}	2.98×10^{-2}	3.13×10^{-2}	7.72×10^{-1}	1.65×10^{-2}

IND, Independent; AR-1, Autoregressive of order 1; EXC, Exchangeable; UNS, Unstructured; PRO, Proposed; TR, Trace of the estimated covariance.

map of the true covariance matrix $\boldsymbol{\Omega}_i$ is given in Fig. 3. The main difference from the Scenario 1 is the number of repeated measurements, $n_i = 18$. The identity link is used, i.e., $\mu_{ij} = \eta_{ij}$, and $a(\mu_{ij}) = 1$ with $\phi = 1$, and $K = 226$ and $p = 4$ were kept the same. We set the mean model

$$E[y_{ij} \mid x_{ij}] = \beta_1 + \beta_2 x_{ij}^{Treat} + \beta_3 x_{ij}^{Time} + \beta_4 x_{ij}^{Treat} x_{ij}^{Time},$$

where $\boldsymbol{\beta} = (2, 3, 1, 1)$, $x_{ij}^{Treat} = 1$ for $i = 1, \ldots, 113$, 0, otherwise, $x_{ij}^{Time} = j + 1$. Eigenvalues of the covariance are (7.1517, 1.5877, 1.1438, 1.0614, 1.0053, 0.9819, 0.8875, 0.7881, 0.6191, 0.5375, 0.4908, 0.4446, 0.3928, 0.3513, 0.2745, 0.1609, 0.1201, 0.0009).

Table 2 shows that the standard error of the estimator obtained by using the unstructured correlation structure ranged between 6 and 8 folds of the proposed estimator. The standard error of the proposed estimator showed clear reduction from the best performing structured working correlation matrix, ranging from 74% to 79%. The trace of simulation covariance was the smallest for the proposed estimator by a larger margin than in Scenario 1.

5.3. Scenario 3

In this setting we feature the case where the true correlation matrix is 1-dependent so that the true structure was not represented well by any of the popular working correlation structures such as independent, exchangeable, and AR-1. In addition, the estimated unstructured working correlation matrix yielded unstable regression parameter estimates due to

Table 3
Simulation results of Scenario 3. Comparing bias and standard error of estimates using five different working correlation matrices based on 1000 Monte Carlo samples. Standard errors are presented in parenthesis.

	IND	AR-1	EXC	UNS	PRO
β_1	6.3×10^{-4} (2.01×10^{-2})	5.8×10^{-4} (2.00×10^{-2})	6.3×10^{-4} (2.01×10^{-2})	-1.1×10^{-2} (3.16×10^{-1})	5.0×10^{-4} (1.72×10^{-2})
β_2	-1.0×10^{-3} (2.80×10^{-2})	-8.5×10^{-4} (2.79×10^{-2})	-1.0×10^{-3} (2.80×10^{-2})	9.3×10^{-3} (2.82×10^{-1})	-9.5×10^{-4} (2.40×10^{-2})
β_3	-1.1×10^{-3} (3.16×10^{-2})	-1.1×10^{-3} (3.12×10^{-2})	-1.1×10^{-3} (3.16×10^{-2})	1.9×10^{-2} (5.55×10^{-1})	-8.5×10^{-4} (2.54×10^{-2})
β_4	2.0×10^{-3} (4.41×10^{-2})	1.8×10^{-3} (4.35×10^{-2})	2.0×10^{-3} (4.41×10^{-2})	-1.6×10^{-2} (4.96×10^{-1})	1.7×10^{-3} (3.56×10^{-2})
TR	4.13×10^{-3}	4.06×10^{-3}	4.13×10^{-3}	7.35×10^{-1}	2.79×10^{-3}

IND, Independent; AR-1, Autoregressive of order 1; EXC, Exchangeable; UNS, Unstructured; PRO, Proposed; TR, Trace of the estimated covariance.

a small eigenvalue. In this case, the proposed method produced estimates with smaller standard errors than those from popular choices of independent, exchangeable, and AR-1 structures.

We set $n_i = 10$, for all i, and $K = 4000$. We assumed the mean model as

$$E[y_{ij} \mid x_{ij}] = \beta_1 + \beta_2 x_{ij}^{Treat} + \beta_3 x_{ij}^{Time} + \beta_4 x_{ij}^{Treat} x_{ij}^{Time},$$

where $\beta = (2, 3, 1, 1)$, $x_{ij}^{Treat} = 1$ for $i = 1, \ldots, 2000$, 0, otherwise, and $x_{ij}^{Time} = j/10$. The covariance of $\mathbf{y}_i$ given $\mathbf{X}_i$ was 1-dependent structure with $\alpha = 0.521$. That is, for $1 \le i, j \le 10$

$$\mathbf{R}(\alpha)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.521 & \text{if } |i-j| = 1 \\ 0 & \text{if otherwise.} \end{cases}$$

Eigenvalues are (1.9997, 1.8765, 1.6823, 1.4328, 1.1482, 0.8517, 0.5671, 0.3176, 0.1234, 0.0002).

Table 3 shows that the standard errors of the estimates obtained by using the unstructured correlation structure was around 10–20 fold of those from the proposed method since the estimated unstructured covariance matrix was unstable. Among the structured correlation matrices, the AR-1 yielded the smallest standard error. The proposed method reduced the standard errors to around 81%–86% compared to the AR-1 working correlation structure. The trace of simulation covariance was the smallest for the proposed estimator as in Scenarios 1 and 2.

6. Real data analysis

In this section, we present two real data analyses with (i) SB–LOT data from a randomized clinical trial and (ii) Urinary data from an observational study. We compared five GEE estimators as in Section 5 including the estimates from the proposed method.

6.1. Re-analysis of the SB-LOT trial

The SB-LOT data were collected from a double-blind placebo-controlled randomized multicenter trial with repeated measurements (Vanscheidt et al., 2002). The primary goal of the study was to investigate the oedema-protective effect of a vasoactive drug in patients suffering from chronic venous insufficiency after decongestion of the legs. Patients were followed-up five times: 4, 6, 8, 12, and 16 weeks after initiation of the drug therapy. The primary outcome was the lower leg volume measured by water plethysmometry. At baseline, 226 patients were randomized to medical compression stockings plus SB-LOT (90 mg Coumarin and 540 mg Troxerutin per day) or medical compression stockings plus placebo for the first 4 weeks and SB-LOT or placebo for the following 12 weeks of the study. In the first four weeks all patients wore medical compression stockings, and then the stockings were discontinued in both treatment groups. Therefore, besides the primary hypothesis, the interaction between treatment and time was of secondary interest. The intention-to-treat analysis included 113 patients per treatment group. A secondary analysis was presented by Vens and Ziegler (2012) for detecting a difference in the slopes by making use of the repeated nature of the data. In this paper we revisited the secondary analysis, and we thus assumed the mean model to be

$$E[y_{ij} \mid x_{ij}] = \beta_1 + \beta_2 x_{ij}^{Treat} + \beta_3 x_{ij}^{Time} + \beta_4 x_{ij}^{Treat} x_{ij}^{Time},$$

where $x_{ij}^{Treat} = 1$ for $i = 1, \ldots, 113$, 0, otherwise, $x_{i1}^{Time} = 4$, $x_{i2}^{Time} = 6$, $x_{i3}^{Time} = 8$, $x_{i4}^{Time} = 12$, and $x_{i5}^{Time} = 16$.

Table 4 of Vens and Ziegler (2012) shows the estimates and their standard errors of the interaction term between treatment and time obtained from various working correlation structures. The estimates and their standard errors for independent, AR-1, exchangeable, and unstructured were duplicated in the last row of Table 4 of this paper. Table 4 also reports the estimates from the proposed method. Fig. 4 presents the variance of $\hat{\boldsymbol{\beta}}(\epsilon)$ for the interaction term as a function

Table 4
Comparing GEE estimation results with their standard errors using independent, AR-1, exchangeable, and unstructured working correlation matrices and proposed working correlation matrix using SB–LOT database. We present standard errors in parenthesis.

	IND	AR-1	EXC	UNS	PRO
INT	2.24×10^3 (3.56×10^1)	2.24×10^3 (3.56×10^1)	2.24×10^3 (3.56×10^1)	2.19×10^3 (4.62×10^1)	2.24×10^3 (3.56×10^1)
Treat	9.48 (4.85×10^1)	4.08 (4.86×10^1)	9.48 (4.85×10^1)	1.00×10^2 (6.56×10^1)	1.30×10^1 (4.87×10^1)
Time	4.82×10^{-1} (9.30×10^{-1})	-5.08×10^{-3} (8.95×10^{-1})	4.82×10^{-1} (9.30×10^{-1})	2.62 (2.69)	2.69×10^{-1} (9.31×10^{-1})
T × T	2.50 (1.26)	2.63 (1.20)	2.50 (1.26)	−4.26 (3.71)	2.02 (1.30)

IND, Independent; AR-1, Autoregressive of order 1; EXC, Exchangeable; UNS, Unstructured; PRO, Proposed; INT, Intercept; T × T, Treat × Time.

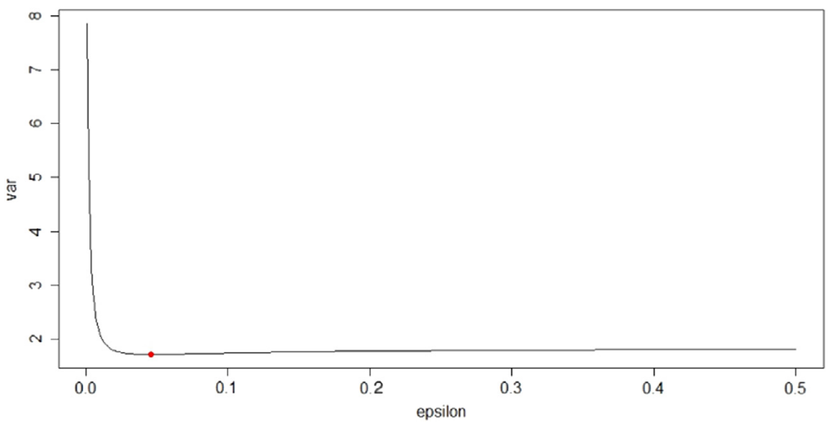

Fig. 4. Variance estimates of the interaction term between time and treatment as a function of ϵ in SB–LOT database.

of ϵ and how ϵ is determined. It is notable that the result from the unstructured working correlation structure was quite different from the rest, exhibiting flipped sign in the estimate and tripled inflation in standard error compared to others. This is due to unstable working correlation estimates as the eigenvalues of the estimated unstructured working correlation matrix were (4.8632, 0.0859, 0.0390, 0.0116, 0.0001). When the proposed method was applied, the coefficient estimate was similar to those from other working correlation structures and the standard error was comparable. In this case, AR-1 is the choice of the working correlation matrix according to the criterion of Westgate (2014) which we applied to the variance of the interaction only.

6.2. Urinary database

Park and Lee (2002) describe the study of ovarian steroid secretion in reproductive-age women from the family planning clinics of the King County Hospital Center and the State University Hospital in Brooklyn, New York. Blood and urine specimens were collected during all study menstrual cycles for 175 women recruited between 1985 and 1988. Among these, 118 women were seeking tubal ligation and 57 women were fertile women using a variety of contraceptive methods and not seeking tubal ligation. Urinary oestradiol levels and serum oestrogen levels were repeatedly observed for each subject over three or four menstrual periods, respectively. We focus on first-year urinary oestradiol levels in this paper. At each study menstrual period, daily morning urine specimens were collected 8–18 times. The number of observations varied from subject to subject, because subjects missed visits or failed to collect urine specimens.

A total of 2330 observations from 165 subjects recruited first year were included in the analysis. We assumed linear model with covariates, smoking, age, weight, age at menarche (mage), treat (tubal ligation), day at ovulation (ovday), ovday2 ($(\text{ovday})^2$), and ovday3 ($(\text{ovday})^3$). Fig. 5 presents the trace of covariance matrix of regression parameters except intercept, as a function of ϵ. Eigenvalues of the unstructured working correlation matrix were (7.674, 1.638, 1.156, 1.067, 1.006, 0.980, 0.878, 0.770, 0.587, 0.498, 0.448, 0.397, 0.341, 0.296, 0.213, 0.090, 0.045, −0.084). Due to negative minimum eigenvalue, the GEE estimates under the unstructured correlation structure cannot be defined. Table 5 shows that the treatment effect was not significant, and smoking significantly reduced urinary oestradiol levels. The variance of the proposed estimate for tubal ligation was around 92% of that of the estimate using AR-1 under which the smallest standard error was obtained among the

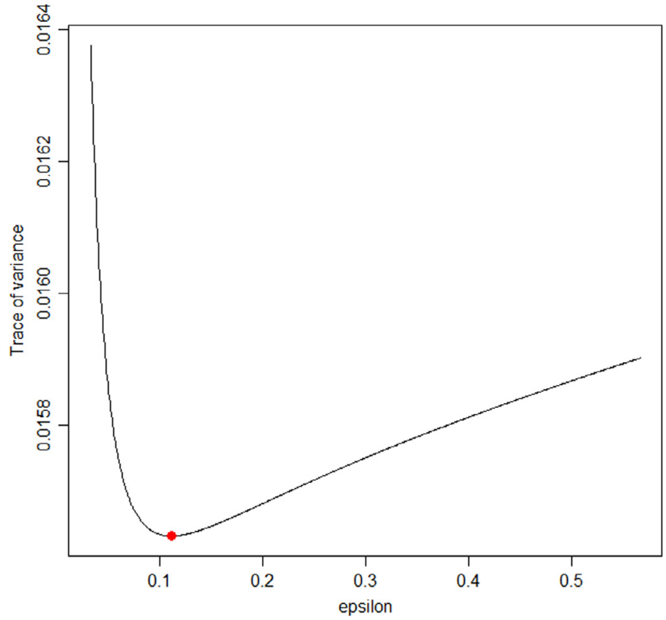

Fig. 5. Trace of variance matrix excluding intercept term as function of ϵ in urinary database.

Table 5
Comparing five GEE estimation results with their standard errors using independent, AR-1, and exchangeable working correlation matrices and proposed working correlation matrix using urinary database. We present standard errors in parenthesis.

	IND	AR-1	EXC	UNS	PRO
INT	4.49 (4.74×10^{-1})	4.59 (4.70×10^{-1})	4.58 (4.78×10^{-1})	4.21 (4.91×10^{-1})	4.39 (4.89×10^{-1})
Smoking	-2.81×10^{-1} (8.92×10^{-2})	-3.05×10^{-1} (8.45×10^{-2})	-2.98×10^{-1} (8.94×10^{-2})	-2.93×10^{-1} (8.64×10^{-2})	-3.08×10^{-1} (8.69×10^{-2})
Age	-1.43×10^{-2} (1.03×10^{-2})	-1.12×10^{-2} (9.86×10^{-3})	-1.61×10^{-2} (1.03×10^{-2})	-5.39×10^{-3} (9.95×10^{-3})	-9.55×10^{-3} (1.02×10^{-2})
Weight	1.01×10^{-3} (9.74×10^{-4})	6.19×10^{-4} (9.86×10^{-4})	1.08×10^{-3} (9.98×10^{-4})	6.01×10^{-4} (9.62×10^{-4})	5.76×10^{-4} (9.46×10^{-4})
Mage	-1.31×10^{-2} (2.70×10^{-2})	-1.88×10^{-2} (2.71×10^{-2})	-1.67×10^{-2} (2.68×10^{-2})	-5.42×10^{-3} (2.72×10^{-2})	-1.35×10^{-2} (2.71×10^{-2})
Treat	-2.90×10^{-2} (8.95×10^{-2})	-4.40×10^{-2} (8.74×10^{-2})	-4.24×10^{-2} (8.86×10^{-2})	-5.08×10^{-2} (8.61×10^{-2})	-4.22×10^{-2} (8.38×10^{-2})
Ovday	-5.59×10^{-2} (1.74×10^{-2})	-4.23×10^{-2} (1.49×10^{-2})	-5.05×10^{-2} (1.62×10^{-2})	-5.37×10^{-2} (1.58×10^{-2})	-4.26×10^{-2} (1.33×10^{-2})
Ovday2	6.32×10^{-3} (3.25×10^{-3})	3.62×10^{-3} (2.69×10^{-3})	5.04×10^{-3} (2.91×10^{-3})	5.81×10^{-3} (2.70×10^{-3})	5.42×10^{-3} (2.52×10^{-3})
Ovday3	-3.88×10^{-4} (1.68×10^{-4})	-3.29×10^{-4} (1.34×10^{-4})	-2.94×10^{-4} (1.43×10^{-4})	-3.51×10^{-4} (1.34×10^{-4})	-3.66×10^{-4} (1.31×10^{-4})
TR	1.71×10^{-2}	1.58×10^{-2}	1.69×10^{-2}	1.59×10^{-2}	1.56×10^{-2}

IND, Independent; AR-1, Autoregressive of order 1; EXC, Exchangeable; UNS, Unstructured; PRO, Proposed; INT, Intercept; Mage, Age at menarche; TR, Trace of the estimated covariance excluding intercept part.

structured working correlations. When we applied the criterion of Westgate (2014) to the empirical covariance excluding intercept, the proposed method showed the smallest trace among all choices of working correlation structures.

7. Discussion and conclusion

We proposed a GEE using the modified working correlation matrix with a linear shrinkage method. The proposed method could broaden choices of working correlation structures to unstructured when a small or negative minimum eigenvalue of the estimated working correlation causes problems. Asymptotically we showed that the proposed estimator has the same distributional property with the estimator without any adjustment, but in finite samples the linear shrinkage can help stabilizing the regression coefficient estimates and reducing the standard error. Adding the linear-shrinkage working correlation structure to a routine GEE analysis may broaden the choices among working correlation structures and eventually lead to more efficient estimators than those obtained by restricting to popular working correlation structures.

Acknowledgements

The authors are grateful to Schaper and Brümmer (contact: Dr. Hans-Heinrich Henneicke-von Zepelin) for making the SB-LOT data available. The authors also thank Nicole Heßler for the SB-LOT data re-analysis.

Appendix

As in Liang and Zeger (1986), under some regularity conditions and fixed ϵ, we can show that $K^{1/2}\{\tilde{\boldsymbol{\beta}}(\epsilon)-\boldsymbol{\beta}\}$ can be approximated by

$$\left[K^{-1}\sum_{i=1}^{K}-\frac{d}{d\boldsymbol{\beta}}\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}\right]^{-1}\left[K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}\right], \tag{A.1}$$

where

$$d\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}/d\boldsymbol{\beta}=\partial\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}/\partial\boldsymbol{\beta}+[\partial\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}/\partial\hat{\boldsymbol{\alpha}}]\{\partial\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\}.$$

The main idea is to claim that the first term of Eq. (A.1) converges to $(\lim_{K\to\infty}K^{-1}\sum_{i=1}^{K}\mathbf{D}_i^T\breve{\boldsymbol{\Sigma}}_i^{-1}\mathbf{D}_i)^{-1}$, and

$$K^{-1/2}\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}-K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}=o_p(1). \tag{A.2}$$

These two can be shown using the following lemma.

Lemma. *Suppose that $\mathbf{R}(\boldsymbol{\alpha})$ is well defined in the sense of Crowder (1995) and assume that there exists $\hat{\boldsymbol{\alpha}}$ that satisfies $K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)=O_p(1)$ as in Liang and Zeger (1986). If ϵ is smaller than the minimum eigenvalue, $\gamma_{\min}\{\mathbf{R}(\boldsymbol{\alpha})\}$, then, for any q-dimensional vector $\mathbf{a}$ such that $\|\mathbf{a}\|_2^2=\mathbf{a}^T\mathbf{a}=1$,*

$$\{\mathbf{a}^T(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\}^2\le(1+q)^2\|(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\|_2^2,$$

where $\tilde{\boldsymbol{\alpha}}$ is a vectorized version of the matrix $\tilde{\mathbf{R}}(\hat{\alpha},\epsilon)=\tilde{\mathbf{R}}(\hat{\alpha})$.

Proof. Denote by $\|\cdot\|_S$ the spectral norm for a symmetric matrix. By Theorem 3.4 in page 32 of Choi (2015), we have $q^{-1/2}\|\tilde{\mathbf{R}}(\hat{\alpha})-\mathbf{R}(\hat{\alpha})\|_F\le(\epsilon-\gamma_{\min}\{\mathbf{R}(\hat{\alpha})\})_+$ which is bounded by $\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_S$ provided that $\epsilon<\gamma_{\min}\{\mathbf{R}(\alpha_0)\}$. This yields that $q^{-1/2}\|\tilde{\mathbf{R}}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F\le q^{-1/2}\|\tilde{\mathbf{R}}(\hat{\alpha})-\mathbf{R}(\hat{\alpha})\|_F+q^{-1/2}\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F\le\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_S+q^{-1/2}\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F$. Since $\|\cdot\|_S\le q^{1/2}\|\cdot\|_F$ in general, it holds that $q^{-1/2}\|\tilde{\mathbf{R}}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F\le(q^{-1/2}+q^{1/2})\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F$ and

$$\begin{aligned}\|\tilde{\mathbf{R}}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F\le(1+q)\|\mathbf{R}(\hat{\alpha})-\mathbf{R}(\alpha_0)\|_F\ &\Leftrightarrow\ \sqrt{\frac{\|\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_2^2}{q}}\le(1+q)\sqrt{\frac{\|\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_2^2}{q}}\\ &\Leftrightarrow\ \|\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_2^2\le(1+q)^2\|(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\|_2^2.\end{aligned}$$

Thus, for any vector $\mathbf{a}$ such that $\|\mathbf{a}\|_2^2=1$, by Cauchy–Schwarz inequality

$$\begin{aligned}\{\mathbf{a}^T(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\}^2&\le\|\mathbf{a}\|_2^2\|\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0\|_2^2\\&\le(1+q)^2\|(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\|_2^2.\quad\square\end{aligned}$$

We now prove (A.2). Since

$$K^{-1/2}\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\}=K^{-1/2}\sum_{i=1}^{K}\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0)+K^{-1}\sum_{i=1}^{K}\frac{\partial\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+o_p(1)$$

and

$$K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}=K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\epsilon)+K^{-1}\sum_{i=1}^{K}\frac{\partial\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\epsilon)}{\partial\boldsymbol{\alpha}}K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)+o_p(1),$$

we have

$$
\begin{aligned}
&K^{-1/2}\sum_{i=1}^{K}\mathbf{U}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})\} - K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\} \\
&\quad = K^{-1/2}\sum_{i=1}^{K}\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0) + K^{-1}\sum_{i=1}^{K}\frac{\partial\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) \\
&\qquad - K^{-1/2}\sum_{i=1}^{K}\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha}_0,\epsilon) - K^{-1}\sum_{i=1}^{K}\frac{\partial\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\epsilon)}{\partial\boldsymbol{\alpha}}K^{1/2}(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) + o_p(1) \\
&\quad = K^{-1}\sum_{i=1}^{K}\frac{\partial\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) - K^{-1}\sum_{i=1}^{K}\frac{\partial\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\epsilon)}{\partial\boldsymbol{\alpha}}K^{1/2}(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) + o_p(1). \qquad (A.3)
\end{aligned}
$$

The last equality holds because $\tilde{\mathbf{R}}(\boldsymbol{\alpha}_0,\epsilon) = \mathbf{R}(\boldsymbol{\alpha}_0)$ by the definition of $\tilde{\mathbf{R}}(\boldsymbol{\alpha},\epsilon)$. The first term of Eq. (A.3) is $o_p(1)$, since $\partial\mathbf{U}_i(\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ is a mean zero linear function and $K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}) = O_p(1)$. For the second term, by the lemma

$$
\left\| K^{-1}\sum_{i=1}^{K}\frac{\partial\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\epsilon)}{\partial\boldsymbol{\alpha}}K^{1/2}(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\right\|_2 \le \left\| K^{-1}\sum_{i=1}^{K}\frac{\partial\tilde{\mathbf{U}}_i(\boldsymbol{\beta},\boldsymbol{\alpha},\epsilon)}{\partial\boldsymbol{\alpha}}\right\|_2 (1+q)\|K^{1/2}(\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0)\|_2,
$$

and right hand side is bounded by $o_p(1)$. This proves (A.2). Similarly, under the same assumption of the lemma, we can apply the lemma column-wise to show that $\left[K^{-1}\sum_{i=1}^{K} -d\tilde{\mathbf{U}}_i\{\boldsymbol{\beta},\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}),\epsilon\}/d\boldsymbol{\beta}\right]$ converges to $(\lim_{K\to\infty} K^{-1}\sum_{i=1}^{K}\mathbf{D}_i^T\breve{\boldsymbol{\Sigma}}_i^{-1}\mathbf{D}_i)^{-1}$.

References

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

Carey, V., Wang, Y.-G., 2011. Working covariance model selection for generalized estimating equations. Stat. Med. 30 (26), 3117–3124.

Choi, Y.-G., 2015. Positive-definite correction of covariance matrix estimators via linear shrinkage (Doctoral Dissertation), Retrieved from http://dcollection.snu.ac.kr/jsp/common/DcLoOrgPer.jsp?sItemId=00000006725.

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

Daniels, M.J., Kass, R.E., 1999. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. J. Amer. Statist. Assoc. 94 (448), 1254–1263.

Daniels, M.J., Kass, R.E., 2001. Shrinkage estimators for covariance matrices. Biometrics 57 (4), 1173–1184.

Gosho, M., Hamada, C., Yoshimura, I., 2011. Criterion for the selection of a working correlation structure in the generalized estimating equation approach for longitudinal balanced data. Comm. Statist. - Theory Methods 40 (21), 3839–3856.

Haff, L.R., 1991. The variational form of certain Bayes estimators. Ann. Statist. 19 (3), 1163–1190.

Hin, L.-Y., Wang, Y.-G., 2009. Working-correlation-structure identification in generalized estimating equations. Stat. Med. 28 (4), 642–658.

Lam, C., 2016. Nonparametric eigenvalue-regularized precision or covariance matrix estimator. Ann. Statist. 44 (3), 928–953.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivariate Anal. 88 (2), 365–411.

Ledoit, O., Wolf, M., 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. Ann. Statist. 40 (2), 1024–1060.

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

Pan, W., 2001. Akaike's information criterion in generalized estimating equations. Biometrics 57 (1), 120–125.

Park, T., Lee, Y.J., 2002. Covariance models for nested repeated measures data: Analysis of ovarian steroid secretion data. Stat. Med. 21 (1), 143–164.

Qu, A., Lee, J.J., Lindsay, B.G., 2008. Model diagnostic tests for selecting informative correlation structure in correlated data. Biometrika 95 (4), 891–905.

Schäfer, J., Strimmer, K., 2005. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat. Appl. Genet. Mol. Biol. 4 (1).

Song, P.X.-K., Jiang, Z., Park, E., Qu, A., 2009. Quadratic inference functions in marginal models for longitudinal data. Stat. Med. 28 (29), 3683–3696.

Stein, C., 1956. Inadmissibility of the usual estimator for the mean of a multi- variate normal distribution. In: Proc. Third Berkeley Symp. on Math. Statist. and Prob., Volume 1. University of California Press, Berkeley, CA, pp. 197–206.

Vanscheidt, W., Rabe, E., Naser-Hijazi, B., Ramelet, A., Partsch, H., Diehm, C., Schultz-Ehrenburg, U., Spengel, F., Wirsching, M., Götz, V., Schnitker, J., von Zepelin, H., 2002. The efficacy and safety of a Coumarin-/Troxerutin-combination (SB-LOT) in patients with chronic venous insufficiency: a double blind placebo-controlled randomised study. Vasa 31 (3), 185–190.

Vens, M., Ziegler, A., 2012. Generalized estimating equations and regression diagnostics for longitudinal controlled clinical trials: A case study. Comput. Statist. Data Anal. 56 (5), 1232–1242.

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

Wang, Y.-G., Carey, V., 2004. Unbiased estimating equations from working correlation models for irregularly timed repeated measures. J. Amer. Statist. Assoc. 99 (467), 845–853.

Westgate, P.M., 2013. A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. Stat. Med. 32 (16), 2850–2858.

Westgate, P.M., 2014. Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. Stat. Med. 33 (13), 2222–2237.

Won, J.-H., Lim, J., Kim, S.-J., Rajaratnam, B., 2013. Condition number regularized covariance estimation. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (3), 427–450.

Zhou, Q.M., Song, P.X.-K., Thompson, M.E., 2012. Information ratio test for model misspecification in quasi-likelihood inference. J. Amer. Statist. Assoc. 107 (497), 205–213.