Published in final edited form as:

J Comput Graph Stat. 2015 April 1; 24(2): 455–476. doi:10.1080/10618600.2014.909733.

Estimating and Identifying Unspecified Correlation Structure for Longitudinal Data

Jianhua Hu [Associate Professor],

University of Texas MD Anderson Cancer Center, Houston, TX 77030 (jhu@mdanderson.org).

Peng Wang [Assistant Professor], and

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403 (wangp@bgsu.edu).

Annie Qu¹ [Professor]

Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820

Abstract

Identifying correlation structure is important to achieving estimation efficiency in analyzing longitudinal data, and is also crucial for drawing valid statistical inference for large size clustered data. In this paper, we propose a nonparametric method to estimate the correlation structure, which is applicable for discrete longitudinal data. We utilize eigenvector-based basis matrices to approximate the inverse of the empirical correlation matrix and determine the number of basis matrices via model selection. A penalized objective function based on the difference between the empirical and model approximation of the correlation matrices is adopted to select an informative structure for the correlation matrix. The eigenvector representation of the correlation estimation is capable of reducing the risk of model misspecification, and also provides useful information on the specific within-cluster correlation pattern of the data. We show that the proposed method possesses the oracle property and selects the true correlation structure consistently. The proposed method is illustrated through simulations and two data examples on air pollution and sonar signal studies.

Keywords

correlated data; eigenvector decomposition; oracle property; quadratic inference function; SCAD penalty

1 Introduction

Longitudinal data arise frequently for studies where long term treatment effect is one of the main scientific interests and identifying correlation structure is critical for appropriate statistical analysis. For longitudinal data analysis, it is important to incorporate the correlation among repeated measurements since utilizing true correlation structure can increase the efficiency of regression parameter estimation and reduce the bias of the

¹(anniequ@illinois.edu)..

estimator in nonparametric modeling (Wang, 2003). Estimating and identifying correlation structure for large-scale clustered data remains challenging since applying the empirical correlation estimator could be infeasible when the sample size is relatively small compared to the cluster size.

Liang and Zeger (1986) proposed the generalized estimating equation (GEE) approach assuming a common working correlation structure with a small number of nuisance parameters. However, it is often difficult to determine which working correlation structure can approximate the true structure sufficiently well. In fact, the three most commonly used working correlation structures of independence, AR-1 and exchangeable (displayed in Figure 1) most often cannot represent the true correlation structure well.

The existing literature mainly focuses on the estimation of covariance matrices for continuous responses including Cholesky decomposition (Huang et al., 2006; Huang et al., 2007), factor modeling (Fan et al., 2008), and the spectrum random matrix approach (El Karoui, 2008). For high-dimensional covariance matrix estimation, there is extensive work using nested LASSO (Levina et al., 2008; Bien and Tibshirani, 2011), the non-convex regularization method (Zhu et al. 2014) and thresholding approaches (Bickel and Levina, 2008; Rothman et al., 2009; Rothman, 2012). However, there are limited studies on estimation of correlation structures and identifying the correlation structure for discrete longitudinal data.

Moreover, the above approaches for high-dimensional data mainly focus on estimation of the sparse covariance matrix or its inverse. Although sparsity is a desired property in applications such as network analysis (Peng et al., 2009, James et al., 2010) and graphical models (Cheng et al., 2012; Guo et al., 2011; Yang et al., 2012; Yuan et al., 2012; Yuan et al., 2013; Zhu et al., 2014), we do not have such a requirement in identifying the correlation structure. This is because the proposed method obtains a small number of basis matrices to capture the structure of the correlation, and therefore the correlation or inverse correlation matrix do not need to be sparse. Zhou and Qu (2012) propose model selection for correlation structures for continuous and discrete responses, conditional on that the candidate basis matrices be known in advance. However, this assumption could be restrictive, as prior information of the correlation structure might not be available.

In this paper, we propose an alternative strategy based on eigenvalue decomposition through utilizing the first several eigenvectors of the empirical correlation matrix to approximate the inverse correlation matrix. We minimize the Euclidean norm of the difference between two estimating functions based on the empirical correlation information and the eigenvector approximation. In addition, we penalize the model if the approximation of the inverse correlation matrix involves a large number of eigenvectors. The penalization allows one to capture essential correlation information, yet not be overburdened by estimating the high dimension of the nuisance parameters. Computationally, we only need to solve one penalized objective function, in contrast to the existing inverse covariance matrix estimation approaches which solve numerous lasso-type problems iteratively.

One advantage of the proposed approach is that it does not require the specification of the likelihood function. Therefore it is applicable to discrete correlated data where the likelihood function might be difficult to specify. Another advantage is that we can avoid pre-specifying any basis matrices, as in Zhou and Qu (2012). The selection of the basis matrices in our approach is data-adaptive in that the basis matrices are obtained from the data without prior information. In addition, the basis matrices selected in the proposed approach are either identity or rank-one matrices, and this greatly reduces the computational cost in the optimization process when the cluster size is large.

We show in our numerical studies that our approach can capture the most important correlation information, since the inverse correlation matrix can be sufficiently approximated by a linear combination of the first several eigenvectors even with a large size of clustered data. The approach also works well when the correlation matrix is sparse. In addition, by decomposing the correlation structure into orthogonal spaces, we are able to investigate specific patterns of the within-cluster correlation matrix through the column space expanded by the selected eigenvectors. In theory, we show that the proposed estimators of the correlation parameters possess the oracle property, and the non-zero correlation estimators follow a normal distribution asymptotically.

The rest of the paper is organized as follows. In Section 2, we describe the proposed method of estimating and identifying correlation structures for longitudinal data in the general linear model framework. In Section 3, we illustrate the performance of the proposed method through simulation studies for Gaussian and binary responses. In Section 4, we demonstrate the proposed method through application to air pollution data and sonar signal data. We provide discussion in Section 5 and proofs in the Appendix.

2 Methodology

2.1 Preambles

Let y_{ij} denote a response variable measured at time j ($j = 1, \dots, J$) for the i th ($i = 1, \dots, N$) subject. Note that the measurements within a subject y_{ij} and $y_{ij'}$ can be dependent on each other. We denote $\mathbf{X}_{ij}$ as the corresponding p -dimensional vector of the covariates, and assume that the expectations of the observed data satisfy

$$\mu_{ij} = E(y_{ij}) = \mu(\mathbf{x}_{ij}^T \boldsymbol{\beta}), \quad (1)$$

where $\mu(\cdot)$ is a known inverse link function $\boldsymbol{\beta}$ and is a p -dimensional parameter vector. The most well-known quasi-likelihood equation (e.g., Wedderburn, 1974; Liang and Zeger, 1986) for parameter estimation in a longitudinal data setting is of the form

$$\sum_{i=1}^N \dot{\boldsymbol{\mu}}_i^T \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})$, $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{iJ})$, $\dot{\boldsymbol{\mu}}_i = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$, and $\mathbf{V}_i = \text{var}(\mathbf{y}_i)$. Note that in practice $\mathbf{V}_i$ is usually unknown. A natural choice of an estimator is based on the sample variance of

the responses, which could be unstable in high-dimensional cases where the cluster size J is large relative to the sample size. The generalized estimating equations (GEE) proposed by Liang and Zeger (1986) assume

$$\mathbf{V}_i = \mathbf{A}_i^{1/2} \mathbf{R} \mathbf{A}_i^{1/2},$$

where $\mathbf{A}_i$ is a diagonal marginal variance matrix and $\mathbf{R}_{J \times J}$ is a common working correlation matrix for all subjects.

Liang and Zeger (1986) introduced several common working correlation structures for longitudinal data. The drawback of the GEE is that a possible misspecification of the correlation structure could lead to the loss of efficiency in parameter estimation. Qu et al. (2000) proposed approximating the inverse of $\mathbf{R}$ by a linear combination of basis matrices

$$\mathbf{R}^{-1} \approx b_0 \mathbf{I}_J + b_1 \mathbf{M}_1 + \cdots + b_K \mathbf{M}_K, \quad (2)$$

where $\mathbf{I}_J$ is the identity matrix and the $\mathbf{M}_k$'s are symmetric basis matrices. For a common working correlation structure R , the inverse of R could be decomposed by (2). For illustration, we present the following two examples:

1. The exchangeable structure: the correlation parameters $\{\rho_{ij}\}$ of $\mathbf{R}$ are the same, then

$$\mathbf{R}^{-1} = b_0 \mathbf{I}_J + b_1 \mathbf{M}_{e,1},$$

where $\mathbf{M}_{e,1}$ has 1 as the off-diagonal elements and 0 as the diagonal elements.

2. The AR(1) structure: $\rho_{ij} = \rho_{ij}^{|i-j|}$ in $\mathbf{R}$, then

$$\mathbf{R}^{-1} = b_0 \mathbf{I}_J + b_1 \mathbf{M}_{a,1} + b_2 \mathbf{M}_{a,2},$$

where $\mathbf{M}_{a,1}$ has 1 on the sub-diagonal entries, and 0 elsewhere; and $\mathbf{M}_{a,2}$ has 1 on the entries (1, 1) and (J, J), and 0 elsewhere.

Applying (2) to the GEE, we observe that the GEE is approximately a linear combination of elements in the estimating functions $\mathbf{g}_N = N^{-1} \sum \mathbf{g}_i$, with

$$\mathbf{g}_i = \left[\left\{ \dot{\boldsymbol{\mu}}_i^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right\}^T, \left\{ \dot{\boldsymbol{\mu}}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_1 \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right\}^T, \dots, \left\{ \dot{\boldsymbol{\mu}}_i^T \mathbf{A}_i^{-1/2} \mathbf{M}_K \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) \right\}^T \right]^T.$$

The parameter vector $\boldsymbol{\beta}$ can be estimated by setting $\mathbf{g}_N$ as close to zero as possible. Since there are more equations than parameters, it is impossible to solve all equations in $\mathbf{g}_N$ simultaneously. Hansen (1982) introduced the generalized method of moments when the moment conditions are over-identified, and estimated the parameter by minimizing the weighted quadratic distance function, with the weighting matrix as the covariance matrix of the moment conditions. Qu et al. (2000) utilized the generalized method of moments to

accommodate correlation information from clustered data, and defined the quadratic inference function (QIF) as

$$Q_N(\beta) = N \bar{\mathbf{g}}_N^T C_N^{-1} \bar{\mathbf{g}}_N, \quad (3)$$

where $C_N = N^{-1} \sum_{i=1}^N \mathbf{g}_i \mathbf{g}_i^T$ is a consistent estimate of $\Sigma = \text{var}(\mathbf{g}_i)$, and $\hat{\beta}$ obtained by minimizing $Q_N(\beta)$. The advantage of the QIF approach is that it can provide statistical inference of the regression parameters without requiring the estimation of the coefficients associated with basis matrices.

For both the GEE and the QIF approaches, it is important to specify the correlation structure correctly in order to improve the efficiency of the regression parameter estimator. Zhou and Qu (2012) proposed a model selection strategy to identify the relevant basis matrices which provide sufficient approximation of the correlation structure based on the representation in (2). The main advantage of their approach is that the required number of pre-specified basis matrices is relatively small compared to the cluster size. However, their approach could impose the risk of misspecification for the correlation structure if the pre-specified candidates for the basis matrices do not contain the true structure.

2.2 Estimation of nonparametric correlation structure

To address the misspecification problem of the working correlation structure, we develop a nonparametric and data-adaptive method for selecting the correlation structure. We propose constructing the basis matrices as

$$\mathbf{R}^{-1} \approx b_0 \mathbf{I}_J + \sum_{k=1}^K b_k \mathbf{e}_k \mathbf{e}_k^T, \quad (4)$$

where $\mathbf{e}_k$ is the k^{th} eigenvector associated with the k^{th} largest eigenvalue of the sample correlation matrix $\mathbf{S}_{J \times J}$ of $\mathbf{y}$. Qu and Lindsay (2003) proposed a conjugate gradient approach where the inverse of the sample correlation matrix can be approximated as a linear space spanned by the identity matrix, the sample correlation matrix and its power sequence. Here we propose an alternative strategy using a series of eigenvectors to approximate the sample correlation matrix, and the inverse of the correlation matrix without inversion.

The advantage of using the linear representation of $\mathbf{R}^{-1}$ is its capability of identifying the non-zero coefficients b_k associated with the relevant basis matrices for the correlation structure. Therefore, this can be transformed to a model selection problem. In addition, the eigenvector decomposition approach has the following advantages. The nonparametric approach minimizes the model assumption and the possibility of correlation structure misspecification, and therefore improves the efficiency of parameter estimation. In addition, most of the correlation matrix information can be captured by a small number of eigenvectors sufficiently. This is due to the fact that most of the remaining eigenvectors do not provide much information for the correlation structure even when the cluster size is large. Finally, the eigenvector decomposition allows one to identify certain correlation

structures. For example, the common exchangeable correlation structure can be represented equivalently by an eigenvector basis matrix.

The proposed method is motivated by the following representation using the sample correlation matrix $\mathbf{S}$ as a basis matrix:

$$\mathbf{R}^{-1} \approx b_0 \mathbf{I}_J = b_1 \mathbf{S}_{J \times J}. \quad (5)$$

However, using the sample correlation matrix might introduce additional variability, especially when the cluster size J is large relative to the sample size N . This is because, except for a few eigenvectors associated with large eigenvalues, most eigenvectors associated with small eigenvalues provide little information for correlation matrix estimation. Including information which is not very relevant for parameter estimation could affect the efficiency of the estimation when the sample size is small. Our numerical studies in Section 3 also confirm this observation. In addition, the sample correlation matrix does not provide additional information regarding the specific correlation pattern within the clusters.

To obtain an informative correlation matrix, we proceed to estimate the coefficients, b_k 's, by minimizing the discrepancy between the estimating functions, using the empirical correlation matrix estimator $\tilde{\mathbf{R}}^{-1}$ and the one based on the eigenvector basis matrix approximation. We can formulate the discrepancy for the i cluster as

$$D_i = \dot{\boldsymbol{\mu}}_i^T(\tilde{\boldsymbol{\beta}}) \mathbf{A}_i^{-1/2} \left(\tilde{\mathbf{R}}^{-1} - b_0 \mathbf{I}_J - \sum_{k=1}^K b_k \mathbf{e}_k \mathbf{e}_k^T \right) \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\tilde{\boldsymbol{\beta}})), \quad (6)$$

where $\tilde{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$ obtained from the standard GEE approach using the independent correlation structure, and $\tilde{\mathbf{R}}$ is the corresponding sample correlation matrix of the residuals.

Asymptotically it seems to be true that including more basis matrices should achieve higher model accuracy. However, including too many basis matrices may cause over-fitting of the model for the correlation structure in finite samples. This problem becomes more serious as the cluster size increases. We propose to estimate the b_k 's by minimizing the objective function

$$L = \sum_{i=1}^N D_i^T D_i + N_p \sum_{k=1}^K p_\lambda(|b_k|), \quad (7)$$

where the penalty function $p_\lambda(\cdot)$ is the SCAD penalty (Fan and Li, 2001) based on the local linear approximation (Zou and Li, 2008). Note that the first coefficient b_0 is not penalized here since the identity matrix $\mathbf{I}_J$ is an important component for any correlation structure representation, and should be kept in the model.

We propose a BIC-type of criterion to select the tuning parameter λ through minimizing the following objective function:

$$BIC(\lambda) = N \frac{\|\tilde{\mathbf{R}}^{-1} - \hat{\mathbf{R}}^{-1}\|}{\|\tilde{\mathbf{R}}^{-1} + \hat{\mathbf{R}}^{-1}\|} + \log(N) q(\lambda), \quad (8)$$

where $\|\cdot\|$ is the Euclidean distance, and $q(\lambda)$ is the number of nonzero estimators of b_k 's obtained with the corresponding tuning parameter value λ . The denominator of the first term of $BIC(\lambda)$ plays the role of standardizing the difference between the inverse of the sample correlation matrix $\tilde{\mathbf{R}}^{-1}$ and the inverse of the estimated correlation matrix $\hat{\mathbf{R}}^{-1}$. The adaptive scaling ensures that the magnitude of the first term is comparable to the second penalty term.

2.3 Implementation and Algorithm

To achieve high computational efficiency, we first define

$$\begin{aligned} U_i &= \dot{\boldsymbol{\mu}}_i^T(\tilde{\beta}) \mathbf{A}_i^{-1/2} \tilde{\mathbf{R}}^{-1} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\tilde{\beta})), \\ V_{i0} &= \dot{\boldsymbol{\mu}}_i^T(\tilde{\beta}) \mathbf{A}_i^{-1/2} \mathbf{I}_J \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\tilde{\beta})), \end{aligned}$$

and

$$V_{ik} = \dot{\boldsymbol{\mu}}_i^T(\tilde{\beta}) \mathbf{A}_i^{-1/2} \mathbf{e}_k \mathbf{e}_k^T \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\tilde{\beta})), \quad k=1, \dots, K.$$

We apply a one-step local linear approximation to the SCAD function (Zou and Li, 2008), and obtain an approximate solution to the optimization of L via minimizing

$$L^* = \sum_{i=1}^N \|U_i - \sum_{k=0}^K b_k V_{ik}\|^2 + N_p \sum_{k=1}^K p'_\lambda(|b_k^*|) |b_k|, \quad (9)$$

where b_k^* is the least-squares estimator without a penalty. This modified objective function is equivalent to an adaptive LASSO objective function (Zou, 2006) and can be directly minimized by implementing the efficient least angle regressions (LARS) algorithm of Efron et al. (2004).

We outline the algorithm for identifying the correlation structure as follows:

1. Obtain an initial estimate $\tilde{\beta}$ using standard GEE with independence structure.
2. Compute the sample correlation matrix of the residuals R using $\tilde{\beta}$ obtained in Step 1.
3. Identify non-zero b_k 's in (4) through minimizing the objective function (9), and selecting the tuning parameter via minimizing the BIC-type of criterion (8).
4. For the non-zero b_k 's, construct the basis matrices as $\mathbf{M}_k = \mathbf{e}_k \mathbf{e}_k^T$, and apply the QIF described in (3) to obtain $\hat{\beta}$ the final estimator of β .

2.4 Theoretical Results

The one-step SCAD penalty ensures the sparsity of the b_k estimation. This implies that the estimators of the true zero-parameters are shrunk to zero with probability approaching 1 as the sample size N goes to infinity.

Moreover, our estimation approach possesses the following oracle property. We assume that

the true parameter vector $\mathbf{b}^0 = (b_0^0, b_1^0, \dots, b_K^0)^T$ consists of two subvectors $\mathbf{b}_I^0$ and $\mathbf{b}_{II}^0$, where $\mathbf{b}_I^0 = 0$ are true-zero coefficients corresponding to the null basis matrices and $\mathbf{b}_{II}^0 \neq 0$ are nonzero coefficients corresponding to the relevant basis matrices. We denote the

corresponding estimator of $\mathbf{b}^0$ obtained by minimizing (9) as $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_I^T, \hat{\mathbf{b}}_{II}^T)^T$. We also denote the tuning parameter as λ_N to reflect its dependence on the sample size N . The following theorem states the asymptotic properties of the proposed estimators for the correlation coefficient.

Theorem 1 Assume that the correlation matrix has a structure which can be represented by basis matrices through eigenvector decomposition with coefficients $\mathbf{b}_{II}^0$. Suppose the following conditions are satisfied:

$$C_1 : E(\|y_i\|_2^4) < \infty;$$

$$C_2 : \lambda_N \rightarrow 0 \text{ and } N^{1/2}\lambda_N \rightarrow \infty.$$

The following asymptotic properties hold:

- a. The correlation structure of $\mathbf{R}$ is identified correctly with probability approaching 1. Specifically, $\hat{\mathbf{b}}_I = 0$.
- b. The estimators of the nonzero coefficients $\mathbf{b}_{II}^0$ are asymptotically normal. That is, $\sqrt{N}(\hat{\mathbf{b}}_{II} - \mathbf{b}_{II}^0) \xrightarrow{\mathcal{D}} N(0, \Omega)$, where the covariance matrix Ω is as defined in the Appendix.

3 Simulation studies

We conducted simulation studies to investigate normal and binary response data. We examined the proposed eigenvector-based correlation structure estimation procedure, denoted by *QIFE*. We investigated up to 10 eigenvector bases ($K = 10$), which is generally sufficient according to our empirical studies. For comparison, we examined other approaches with several specified basis matrices: (I) $\mathbf{I}_J$ -only (independent) basis matrix, denoted by *IND*; (II) exchangeable basis matrices, denoted by *EXC*; (III) AR(1) basis matrices, denoted by *ARI*; and (IV) sample correlation matrix $S_{J \times J}$ as in equation (5), denoted by *SMV*.

3.1 Continuous data

We first generated the data from the model

$$y_{ij} = x_{ij}\beta + \varepsilon_{ij},$$

where y_{ij} is the observed value of the i^{th} ($i = 1, \dots, N$) sample in the j^{th} cluster ($j = 1, \dots, J$), and the covariate $x_{ij} = \frac{j-1}{J-1}$. Let $\mathbf{R} = (\rho_{ij})$ denote the correlation matrix of $\mathbf{y}_i$. The residuals for the i^{th} sample ε_i are generated from a normal distribution with mean 0 and correlation matrix $\mathbf{R}$. The true parameter $\beta = 1$.

We investigated various correlation structures and present simulation results when the true correlation structures are block diagonal. We generated data with sample size $N = 100$ with two cluster sizes of $J = 20$ and 50. The first case is two block diagonals with an exchangeable correlation structure for each block. For the cluster size $J = 20$, the first block is 5×5 -dimensional, and the second block is 15×15 -dimensional. For the cluster size $J = 50$, the dimensions of the two blocks are 10×10 and 40×40 , respectively. The pairwise correlation coefficient is 0.5 for the exchangeable structure. In addition, we considered a three-block diagonal correlation matrix with cluster size $J = 20$ and the corresponding block dimensions of 5×5 , 10×10 and 5×5 ; and cluster size $J = 50$ with three corresponding block dimensions of 5×5 , 35×35 and 10×10 . The first two blocks have exchangeable correlation structures with 0.5 correlation and the third block has an AR(1) correlation structure with correlation $\rho_{ij} = 0.8^{|i-j|}$.

The means, standard errors and the square roots of the mean square errors of $\hat{\beta}$ from 100 simulations are presented in Table 1. In all cases, the *QIFE* is superior in terms of the standard error and mean square error. The performances of *QIFE* and *SMV* are more similar compared to those of the other three approaches, and the *QIFE* is more favorable than the *SMV*. In particular, the *QIFE* achieves a 21% reduction in the square root of the mean square error (MSE) compared to the *SMV* when the true correlation matrix has a three-block diagonal structure with $J = 50$. In addition, the *QIFE* has at least a 22% reduction in the square root of the MSE compared to the other three methods with specified working correlation structures, in most cases. We also notice that the efficiency improvement of *QIFE* over the other methods increases as the cluster size increases. In addition, the mean square errors of the estimators based on the *TRUE* correlation structure are very close to that of the *QIFE*.

Table 2 provides the frequencies of the number of eigenvector bases selected by the *QIFE*. In the case of two-block exchangeable structures, the *QIFE* selected either the first one or two eigenvector bases with the most frequency. In the case of three-block exchangeable and AR(1) correlation structures, the *QIFE* selected either the first two or three eigenvector bases most often. This indicates that the majority of eigenvectors do not provide much information for the correlation structure. The inclusion of all eigenvector bases, as in the *SMV* approach, likely introduces more noise, and therefore reduces the efficiency of the regression parameter estimation compared to that of the *QIFE*. We also note that a correlation matrix with larger dimension does not necessarily require many eigenvector bases to capture the correlation structure.

The decomposition of the orthogonal eigenvector bases matrices also allows one to examine the specific correlation structure of the data. For example, in a single simulated data set with $J = 50$ wherein the true structure is a two-block exchangeable structure, the *QIFE* selected the first eigenvector $\mathbf{b}_1$ bases. We display $\mathbf{b}_1$ in the right panel of Figure 2. We observe that there are two clusters of data between the observations at time points 1 – 10 and 11 – 50, and the entries of $\mathbf{b}_1$ fluctuate around a common value within each cluster. Therefore the first eigenvector is able to reflect the true correlation structure. We also provide the plot of the first eigenvector of the true correlation matrix in the left panel of Figure 2 for comparison. It is clear that the first eigenvector of the sample correlation matrix is very similar to that of the true correlation matrix.

We examine another simulated data set generated from a three-block correlation matrix with two exchangeable and one AR1 correlation structures for $J = 50$, where the *QIFE* selects the first two eigenvectors, $\mathbf{b}_1$ and $\mathbf{b}_2$. The first two eigenvectors of the true correlation matrix are presented in the left panels of Figure 3. For comparison, $\mathbf{b}_1$ and $\mathbf{b}_2$ are plotted in the right panels. These two eigenvectors together can accurately differentiate three clusters. The plot of $\mathbf{b}_1$ in the upper right panel also confirms that the clusters of observations at time points 1 – 5 and 6 – 40 have common correlations (with slight fluctuation) within each cluster; and $\mathbf{b}_2$ in the right bottom panel captures the autoregressive correlation pattern within the third cluster. Since observations close to the time point 45 display the highest pairwise correlation for this data set, the convex curve centering around this time point captures the information of the AR(1) correlation. That is, the correlation decreases as the distance between observations at other time points and at the time point 45 increases. Again, the sample correlations based on the first two eigenvector bases are able to capture the information of the true correlation matrix.

We also compare our approach with two existing approaches proposed by Bien and Tibshirani (2011) and Rothman (2012), which can be implemented using R packages SPCOV and PDSCE, respectively. Figure 4 illustrates the heat map of the inverse of the true correlation matrix and the inverse of the correlation matrix estimators based on these two approaches and our approach. The two diagonal block structures presented in the true correlation matrix are captured by the proposed *QIFE* approach. In contrast, the estimators based on Bien and Tibshirani (2011) and Rothman's (2012) approaches cannot distinguish diagonal block patterns sufficiently well, as indicated by the weak contrasts in the heat map.

3.2 Binary data

We also provide a simulation study for binary data where the binary response was generated from the model

$$y_{ij} \sim \text{Bernoulli}(p_{ij}), \quad \text{where} \quad \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = x_{ij}\beta, \quad (10)$$

and the covariate value $x_{ij} = \frac{j-1}{J-1}$, which is the same as in the normal response data. We let $\beta = 0.5$ and $(N, J) = (100, 40)$ for the first simulation setting, and $(N, J) = (100, 30)$ for the second simulation setting. The correlated binary data were generated using the R package

“mvtBinaryEP” with two correlation structures: (1) a block-wise correlation structure, where the first two blocks are (10×10) dimensional AR-1 correlation matrices with the correlation parameter $\rho = 0.8$, and the third block is a (20×20) dimensional exchangeable correlation matrix with the correlation parameter $\rho = 0.5$; and, (2) a block-wise correlation structure, where the first two blocks are (5×5) and (20×20) dimensional exchangeable correlation matrices with the correlation parameter $\rho = 0.5$, and the third block is a (5×5) dimensional AR-1 correlation matrix with the correlation parameter $\rho = 0.8$.

The simulation results are provided in Table 3. For both simulation settings, the *QIFE* approach outperforms the *SMV* approach and the *QIF* approach using other working structures. The square roots of the MSEs of the *QIFE* are reduced by 5% and 18% relative to those of the *SMV* for the first and second simulation settings, respectively. The efficiency advantage of the *QIFE* over the other approaches is even more significant. We also compare our approach to the *QIF* based on the true correlations structure. In the first simulation setting, the square root of the MSE of *QIFE* is about 6% lower than that of the *QIF* estimator based on the *TRUE* structure, while in the second setting it is about 7% higher. This might be due to the fact that the *QIF* has to utilize at least 6 basis matrices for the *TRUE* structure, while the *QIFE* procedure uses only 2-4 basis matrices, in most cases.

The frequencies of the number of eigenvector bases selected by the *QIFE* for these two cases are provided in Table 2. For the first simulation setting, the proposed *QIFE* procedure selects the first two to four eigenvector bases 80% of the time. For the second simulation setting, the *QIFE* selects the first two or three eigenvector bases most of the time.

3.3 Unbalanced binary data

For missing data or unbalanced data arising in longitudinal studies, we define y_i and μ_i as the length- J_i observed response and its mean vector, respectively, for the subject i . To obtain the common cluster size J for all the subjects, we perform a data transformation using $y_i^* = \Lambda_i y_i$ and $\mu_i^* = \Lambda_i \mu_i$. Here Λ_i is a $J \times J_i$ matrix obtained by removing the columns corresponding to the missing observations in the $J \times J$ identity matrix. As a result, the missing observations are substituted by 0. Afterwards, we can directly apply the method proposed in Section 2. We obtain the initial empirical correlation matrix R and the eigenvector bases using the completely observed subjects.

We evaluate the performance of the proposed approach for unbalanced data or missing data in the following simulation study with binary outcomes. The data were generated from equation (10) with the same covariate value $x_{ij} = \frac{j-1}{J-1}$, and its coefficient = 0.5. We generated $N = 100$ subjects, among which 60 are fully observed and 40 are partially observed with a 50% chance of being unobserved at each time point. For subjects with complete observations, the cluster size $J = 30$, and three blocks with equal cluster size of 10 were constructed. Specifically, the first block is exchangeable with $\rho_{ij} = 0.6$, the second block has the correlation structure of $\rho_{ij} = 0.75 - 0.05|i - j|$ between any two observations, and the third block contains the independent observations. This slightly more complex correlation structure is considered here to investigate whether the proposed approach can work

effectively when the true structure is not a common working correlation type such as AR-1 or exchangeable.

Based on 100 simulation runs, the MSE of the QIFE estimate of β is computed to be 0.0094, which is a 10% efficiency improvement compared to the empirical correlation approach. The improvement of the QIFE is between 46-84% over the GEE approach, using independent, exchangeable and AR-1 working correlation structures whose MSE's of $\hat{\beta}$ are 0.00165, 0.00137, and 0.00173, respectively. Moreover, we evaluate performance in estimating the inverse correlation matrix. Figure 5 displays the heat maps of three correlation matrices corresponding to the true inverse correlation matrix, the QIFE estimate, and the empirical estimate from one simulated data set. It is clear that the proposed approach is able to capture the nontrivial information of the inverse of the true correlation structure much better than the empirical estimate. Note that the QIFE estimator shows clear diagonal three-block patterns, while the empirical estimate merely reflects some data fluctuation.

4 Real studies

4.1 Air pollution data example

We studied an air pollution data set containing various air pollution measurements and their impact on 39 asthmatic patients resident in Windsor, Ontario, Canada in 1992 (Fu, 2003). The data were collected over 21 consecutive days. The response variable of interest is the daily asthmatic status of each patient, which takes the value of 1 if a patient has difficulty breathing, and 0 otherwise. The covariates related to air pollution measurements include total reduced sulphur (TRS), daily mean temperature (MTEMP), nitrogen oxide (NO), nitrogen dioxide (NO₂), mixture of nitrogen oxide and nitrogen dioxide (NOX), ozone level (OZ), carbon monoxide (CO), and coefficient of haze (COH). The primary goal is to examine the health impact of these pollution indices on asthmatic patients. In the analysis, we apply a logistic model to fit the binary outcome variable. The intercept was not included in the model because each covariate was centered and standardized.

We implemented the proposed *QIFE*, and compared its performance to several methods with specified correlation basis matrices, that is, the *IND*, *EXC*, *ARI*, and *SMV* based on the sample correlation matrix. For the *QIFE* approach, we performed a grid search between 0 and 5 to select the optimal tuning parameter based on the BIC-type criterion in equation (8). This resulted in $\lambda = 0.64$ and the selection of the first eigenvector basis. The parameter estimators obtained by each method are shown in Table 4. It is clear that the *QIFE* estimators are most similar to the *EXC* estimators. We observe that the estimators from the *SMV* are also similar to those of the *EXC*, except for the estimators for NO₂. In contrast, both the *IND* and *ARI* approaches obtained estimators which are rather different from the others. This example demonstrates that the standard GEE approach of assuming a fixed working correlation structure could be incorrect and consequently lead to inefficient estimations for regression parameters. The possible misspecification problem can be solved effectively by using the proposed nonparametric method.

We also investigated whether the proposed method based on eigenvector bases contains sufficient information about the correlation structure. We display the plots of the

eigenvalues and the selected first eigenvector from the empirical correlation matrix in Figure 6. The upper panel indicates that the first eigenvalue is much larger than the other eigenvalues, indicating the dominating effect of the first eigenvector. The first eigenvector, displayed in the lower panel, also shows a pattern of fluctuation around a common value. This again confirms the adequacy of the exchangeable correlation structure for this longitudinal study.

4.2 Sonar signal data example

Next, we investigate a sonar signal data set (Gorman and Sejnowski, 1988) containing 111 and 97 sonar signals bounced off a metal cylinder and a rock, respectively. Each signal contains a longitudinal sequence of 60 energy measurements obtained ranging from 0 to 1 within a particular frequency band. Higher frequencies usually occur at the later time period. It is of scientific interest to investigate whether bouncing off different sources show difference in the signals. Previous studies (Rothman et al., 2009; Wang and Daniels, 2012) indicate that the distribution of these signals are approximately independent and identical multivariate normal. Therefore we fit a linear model, where the response variable is the signal measurement, and the covariate is an indicator of whether the signals are bounced off a rock or not. We report the parameter estimators and their standard errors in Table 5, and observe that the QIFE estimator of the slope is very different from the estimators obtained using independent and empirical correlation structures. The standard errors of the QIFE estimators are the smallest among the three estimators, especially for the slope parameter.

We compare the QIFE estimator of the precision matrix with the inverse of the empirical correlation matrix in Figure 7. The heatmap in Figure 7(b) for the QIFE estimators, which use the first two eigenvectors, indicates a two-block structure in the inverse correlation matrix, while Figure 7(c) based on the empirical inverse correlation estimators only reflects a higher correlation along the diagonal. Note that the sample correlation matrix displayed in Figure 7(a) shows some indications of block structure, and the finding of the QIFE estimators is concordant with that of the sample correlation matrix. In contrast, the inverse of the empirical correlation is unable to recover this block structure because the ill-conditioned empirical correlation matrix yields an unstable inverse matrix. This result is consistent with our simulation studies showing that QIFE estimation can capture the important information of the correlation structure.

5 Discussion

We propose a nonparametric approach to estimate correlation structures in longitudinal data for generalized linear models. The proposed approach is able to avoid model misspecification for the correlation structure. The proposed approach decomposes the nontrivial information of the correlation structure into orthogonal spaces using eigenvector-based basis matrices, and therefore allows one to examine the specific correlation patterns within the clusters more carefully. The nature of eigenvector decomposition provides another appealing benefit, that is, only a small number of basis matrices is required for a rather complicated correlation structure of large clustered data.

In addition, our numerical investigations show that the advantage of estimation efficiency for the proposed method compared to existing methods is more significant as the cluster size increases. Our approach also possesses the oracle property in selecting the true correlation structure and consistently estimating the correlation parameters. Furthermore, it guarantees the positive-definiteness of the resulting correlation matrix estimator if the sample size is sufficiently large. However, as a side note, the estimation procedure for the regression parameters based on the quadratic inference function does not require the positive-definiteness of the estimated correlation matrix.

When the cluster size is large compared to the sample size, the sample covariance matrix could be ill-conditioned. The proposed *QIFE* approach can still perform reasonably well as long as we make the following adjustments. First, we regularize the covariance matrix through the hard thresholding approach (e.g., Bickel and Levina, 2008). Next, we estimate the inverse of the covariance matrix using the regularization approaches (e.g., Friedman et al., 2008; Shen et al., 2012; Liu and Wang, 2012; Zhu et al., 2014). In addition, to ensure that a sufficient number of basis matrices are selected to recover the correlation structure, we assign a larger weight for the first term in (8). In our unreported simulation studies, we confirm that the adjusted method works effectively when the empirical covariance matrix breaks down.

Acknowledgements

The authors thank the Editor, an associate editor and a referee for their constructive comments that have substantially improved an earlier version of this paper.

Hu's research is supported by the National Institute of Health Grants R21CA129671, R01GM080503, R01CA158113, NCI CA97007 and CCSG P30 CA01667 and the National Science Foundation Award DMS-0706818. Qu's research is supported by the National Science Foundation DMS-0906660 and DMS-1308227.

Appendix

Proof of Theorem 1: Let $\mathbf{R}$ and $\tilde{\mathbf{R}}$ denote the population correlation matrix and sample correlation matrix, respectively. We also denote the eigenvalues of $\mathbf{R}$ as $0 < d_{(m)} < d_{(m-1)} < \dots < d_{(1)} < \infty$. We can obtain that the inverse sample correlation matrix $\tilde{\mathbf{R}}^{-1}$ is a n -consistent estimator of the inverse population correlation matrix $\mathbf{R}^{-1}$ based on the result in Wang (2011), assuming the regularity conditions stated in Wang (2011).

Furthermore, we denote the eigenvectors of $\tilde{\mathbf{R}}$ as $\mathbf{e}_k$ and the eigenvectors of $\mathbf{R}$ as $\mathbf{h}_k$, $k = 1, \dots, m$. Based on Theorem 4.1 of Konishi (1979), we directly obtain that $(\mathbf{e}_k - \mathbf{h}_k)$ follows a mean-zero normal distribution.

We denote the k^{th} basis matrix $M_k = \mathbf{e}_k \mathbf{e}_k^T$ based on the sample correlation matrix and $M_k^* = \mathbf{h}_k \mathbf{h}_k^T$. It is clear that the correlation structure can be represented well by the eigenvector basis matrices. We can derive the following result.

Lemma 1 *Given that the condition C_1 is satisfied, we have*

$$\text{i. } N^{-1} \sum_{i=1}^N \mathbf{V}_i^T \mathbf{V}_i \xrightarrow{P} \Sigma, \text{ for some covariance matrix } \Sigma.$$

- ii. $N^{-\frac{1}{2}} \sum_{i=1}^N (\mathbf{U}_i - \mathbf{V}_i \mathbf{b}^0)^T \mathbf{V}_i \xrightarrow{\mathcal{D}} N(\mathbf{0}, \Sigma^*)$, where the variance matrix Σ^* is a positive definite matrix.

Proof: We define

$$\mathbf{V}_{i,k} = \dot{\boldsymbol{\mu}}_i^T(\hat{\boldsymbol{\beta}}) \mathbf{A}_i^{-1/2} \mathbf{M}_k \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})),$$

and

$$\mathbf{V}_{i,k}^* = \dot{\boldsymbol{\mu}}_i^T(\hat{\boldsymbol{\beta}}) \mathbf{A}_i^{-1/2} \mathbf{M}_k^* \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})).$$

Because $\mathbf{e}_k = \mathbf{h}_k + O_p(N^{-\frac{1}{2}})$, we can obtain

$$\mathbf{M}_k = \mathbf{M}_k^* + O_p(N^{-\frac{1}{2}}).$$

Based on this, we have

$$\mathbf{V}_{i,k} = \dot{\boldsymbol{\mu}}_i^T(\hat{\boldsymbol{\beta}}) \mathbf{A}_i^{-1/2} (\mathbf{M}_k^* + O_p(N^{-\frac{1}{2}})) \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})).$$

This leads to

$$\mathbf{V}_{i,k} = \mathbf{V}_{i,k}^* + O_p(N^{-1}).$$

So we have

$$\mathbf{V}_i^T \mathbf{V}_i = \mathbf{V}_i^{*T} \mathbf{V}_i^* + O_p(N^{-1}).$$

Therefore, Lemma 1(i) holds given that $N^{-1} \sum_{i=1}^N \mathbf{V}_i^{*T} \mathbf{V}_i^* \rightarrow \Sigma$ as shown by Zhou and Qu (2012).

To prove Lemma 1(ii), we write

$$\begin{aligned} N^{-\frac{1}{2}} \sum_{i=1}^N (\mathcal{U}_i - \mathbf{V}_i \mathbf{b}^0)^T \mathbf{V}_i &= N^{-\frac{1}{2}} \sum_{i=1}^N (\mathcal{U}_i - (\mathbf{V}_i^* + O_p(N^{-1})) \mathbf{b}^0)^T (\mathbf{V}_i^* + O_p(N^{-1})) \\ &= N^{-\frac{1}{2}} \sum_{i=1}^N (\mathcal{U}_i - \mathbf{V}_i^* \mathbf{b}^0 - O_p(N^{-1}) \mathbf{b}^0)^T (\mathbf{V}_i^* + O_p(N^{-1})) \\ &= N^{-\frac{1}{2}} \sum_{i=1}^N [(\mathcal{U}_i - \mathbf{V}_i^* \mathbf{b}^0)^T \mathbf{V}_i^* + O_p(N^{-1})] \\ &= N^{-\frac{1}{2}} \sum_{i=1}^N (\mathcal{U}_i - \mathbf{V}_i^* \mathbf{b}^0)^T \mathbf{V}_i^* + O_p(N^{-\frac{1}{2}}). \end{aligned}$$

Zhou and Qu (2012) derived that

$$N^{-\frac{1}{2}} \sum_{i=1}^N \left(\mathcal{U}_i - \mathbf{V}_i^* \mathbf{b}^0 \right)^T \mathbf{V}_i^* \xrightarrow{\mathcal{D}} N(0, \Sigma^*).$$

The form of Σ^* is specified in the Appendix of Zhou and Qu (2012). Therefore, Lemma 1(ii) holds.

Based on Lemma 1, we can derive Theorem 1 in a similar way as is shown for Theorem 2 in Zhou and Qu (2012). Therefore, we can skip the technical details and only provide the outline of the proof. When N is large enough and the condition C_2 is satisfied, we have

$p'_{\lambda_N}(|b_k^{*0}|) = 0$ for $b_k^0 \neq 0$ and $p'_{\lambda_N}(|b_k^{*0}|) = \lambda_N$ for $b_k^0 = 0$. Based on this fact and Lemma 1, we can establish the convergence rate of $\mathbf{b}^{*0}$ to $\mathbf{b}^0$ as

$$\|\mathbf{b}^{*0} - \mathbf{b}^0\|_2 = O_p\left(N^{-\frac{1}{2}}\right).$$

This result and Lemma 1 lead to

$$\sum_{i=1}^N 2(\mathbf{b}^* - \mathbf{b}^0)^T \mathbf{V}_i^T \mathbf{V}_{ij} = O_p\left(N^{-\frac{1}{2}}\right),$$

where $\mathbf{V}_{ij}$ is the column of $\mathbf{V}_i$ corresponding to b_j .

Because $b_j^* \neq 0$ implies

$$Np\text{Sign}(b_j^*) = \sum_{i=1}^N 2(\mathcal{U}_i - \mathbf{V}_i \mathbf{b}^0)^T \mathbf{V}_{ij} - \sum_{i=1}^N 2(\mathbf{b}^* - \mathbf{b}^0)^T \mathbf{V}_i^T \mathbf{V}_{ij},$$

and the condition C_2 ensures that

$$\frac{\sum_{i=1}^N 2(\mathcal{U}_i - \mathbf{V}_i \mathbf{b}^0)^T \mathbf{V}_{ij} - \sum_{i=1}^N 2(\mathbf{b}^* - \mathbf{b}^0)^T \mathbf{V}_i^T \mathbf{V}_{ij}}{N(p+1)\lambda_N \text{Sign}(b_j^*)} = o_p(1),$$

we have

$$P\{b_j^* \neq 0\} \leq P\left\{Np\text{Sign}(b_j^*) = \sum_{i=1}^N 2(\mathcal{U}_i - \mathbf{V}_i \mathbf{b}^0)^T \mathbf{V}_{ij} - \sum_{i=1}^N 2(\mathbf{b}^* - \mathbf{b}^0)^T \mathbf{V}_i^T \mathbf{V}_{ij}\right\} \rightarrow 0.$$

Therefore, $b_j^* = 0$ for $b_j^0 \in \mathbf{b}_I^0$ with probability approaching 1. So Theorem 1(i) holds.

We define

$$R_N(\mathbf{b}_{II}) = \sum_{i=1}^N \|\mathcal{U}_i - \mathbf{V}_{i,II} \mathbf{b}_{II}\|_2^2,$$

where $\mathbf{V}_{i,II}$ contains the columns of $\mathbf{V}_i$ corresponding to the parameters in $\mathbf{b}_{II}$. Based on Theorem 1(i) and the Taylor expansion of the first derivative of $R_N(\mathbf{b}_{II}^*)$, Theorem 1(ii) can consequently be proved.

References

- Bickel P, Levina E. Covariance regularization by thresholding. *The Annals of Statistics*. 2008; 36:2577–2604.
- Bien J, Tibshirani R. Sparse estimation of a covariance matrix. *Biometrika*. 2011; 98:807–820. [PubMed: 23049130]
- Cheng J, Levina E, Wang P, Zhu J. Sparse ising models with covariates. 2012 ArXiv preprint arXiv: 1209.6342.
- Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression (with discussion). *The Annals of Statistics*. 2004; 32:407–499.
- El Karoui N. Operator norm consistent estimation of large-dimensional sparse co variance matrices. *The Annals of Statistics*. 2008; 36:2717–2756.
- Fan J, Fan Y, Lv J. High-dimensional covariance matrix estimation using a factor model. *Econometrics*. 2008; 147:186–197.
- Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001; 96:1348–1360.
- Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance matrix estimation with graphical lasso. *Biostatistics*. 2008; 9:432–441. [PubMed: 18079126]
- Fu WJ. Penalized estimating equations. *Biometrics*. 2003; 59:126–132. [PubMed: 12762449]
- Gorman RP, Sejnowski TJ. Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*. 1988; 1:75–89.
- Guo J, Levina L, Michailidis G, Zhu J. Joint estimation of multiple graphical models. *Biometrika*. 2011; 98:1–15. [PubMed: 23049124]
- Hansen L. Large sample properties of generalized method of moments estimators. *Econometrica*. 1982; 50:1029–1054.
- Huang JZ, Liu L, Liu N. Estimation of large covariance matrices of longitudinal data with basis function approximations. *J. Comput. Graph. Statist*. 2007; 16:189–209.
- Huang JZ, Liu N, Pourahmadi M, Liu L. Covariance selection and estimation via penalised normal likelihood. *Biometrika*. 2006; 93:85–98.
- James G, Sabatti C, Zhou N, Zhu J. Sparse regulation networks. *Annals of Applied Statistics*. 2010; 4:663–686. [PubMed: 21625366]
- Konishi S. Asymptotic expansions for the distributions of statistics based on the sample correlation matrix in principal component analysis. *Hiroshima Mathematical Journal*. 1979; 9:647–700.
- Levina E, Rothman AJ, Zhu J. Sparse estimation of large covariance matrices via nested LASSO penalty. *Annals of Applied Statistics*. 2008; 2:245–263.
- Liang KY, Zeger SL. Longitudinal data analysis using generalised linear models. *Biometrika*. 1986; 73:13–22.
- Liu H, Wang L. TIGER: A Tuning-Insensitive Approach for Optimally Estimating Gaussian Graphical Models. 2012 arXiv:1209.2437.
- Peng J, Wang P, Zhou N, Zhu J. Partial correlation estimation by joint sparse regression models. *Journal of the American Statistical Association*. 2009; 104:735–746. [PubMed: 19881892]

- Qu A, Lindsay BG. Building adaptive estimating equations when inverse of covariance estimation is difficult. *Journal of the Royal Statistical Society, Series B*. 2003; 65:127–142.
- Qu A, Lindsay BG, Li B. Improving generalised estimating equations using quadratic inference functions. *Biometrika*. 2000; 87:823–836.
- Rothman AJ. Positive definite estimators of large covariance matrices. *Biometrika*. 2012; 99:733–740.
- Rothman AJ, Levina E, Zhu J. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*. 2009; 104:177–186.
- Shen X, Pan W, Zhu Y. Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*. 2012; 107:223–232. [PubMed: 22736876]
- Wang N. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*. 2003; 90:43–52.
- Wang, P. Ph.D. dissertation. Department of Statistics, University of Illinois at Urbana-Champaign; 2011. Mixed-effects modeling and correlation structure selection for high-dimensional correlated data..
- Wang Y, Daniels M. Estimating large correlation matrices by banding the partial autocorrelation matrix. 2012 Manuscript.
- Wedderburn RWM. Quasi-likelihood functions, generalised linear models and the Gauss-Newton method. *Biometrika*. 1974; 61:439–488.
- Yang S, Pan Z, Shen X, Wonka P, Ye J. Fused multiple graphical Lasso. 2012 Manuscript.
- Yuan Y, Shen X, Pan W. Maximum likelihood estimation over directed acyclic Gaussian graphs. *Statistical Analysis and Data Mining*. 2012; 5:523–530.
- Yuan Y, Shen X, Pan W, Wang Z. Reconstruction of a directed acyclic Gaussian graph. 2013 Manuscript.
- Zhou J, Qu A. Informative estimation and selection of correlation structure for longitudinal data. *Journal of the American Statistical Association*. 2012; 107:725–736.
- Zou H. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429.
- Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models (with discussion). *The Annals of Statistics*. 2008; 36:1509–1533. [PubMed: 19823597]
- Zhu Y, Shen X, Pan W. Constrained/regularized Maximum likelihood estimation over undirected graphs. 2014 Manuscript.

(a) Independence structure

(b) AR-1 structure

(c) Exchangeable structure

Figure 1.

Three commonly used working correlations in the GEE and QIF approach.

Figure 2. Plots of eigenvectors of correlation matrices for normally distributed data with block-wise exchangeable correlation structure.

Figure 3. Plots of eigenvectors of correlation matrices for normally distributed data with block-wise correlation structure. The first two blocks are exchangeable and the third block is AR-1.

(a) The inverse of the true correlation matrix.

(b) QIFE approach: two eigenvectors are used.

(c) Bien and Tibshirani's approach: the tuning parameter $\lambda=0.1$.

(d) Rohman's approach: the tuning parameter is chosen by c.v.

Figure 4.

Comparison of the proposed QIFE approach with two existing methods in simulated normal data.

(a) Empirical correlation

(b) QIFE estimates

(c) Empirical estimates

Figure 5.

Estimation of the inverse correlation matrix in simulated binary data.

Figure 6. The eigenvalues and the first eigenvector of the empirical correlation matrix in the air pollution data.

(a) Empirical correlation

(b) QIFE estimates

(c) Empirical estimates

Figure 7.

Empirical correlation matrix and estimation of the inverse correlation matrix in the sonar signal data.

Table 1

Comparisons in the case of normally distributed data with block diagonal correlation structures.

methods	mean($\hat{\beta}$)	se($\hat{\beta}$)	$\sqrt{\text{mse}(\hat{\beta})}$
2-block exchangeable: $N = 100$ and $J = 20$			
IND	0.97	4.89×10^{-1}	4.83×10^{-1}
EXC	0.99	4.57×10^{-1}	4.50×10^{-1}
AR1	1.01	4.70×10^{-1}	4.63×10^{-1}
SMV	1.03	3.34×10^{-1}	3.35×10^{-1}
QIFE	1.02	3.22×10^{-1}	3.18×10^{-1}
TRUE	1.04	3.24×10^{-1}	3.22×10^{-1}
2-block exchangeable: $N = 100$ and $J = 50$			
IND	0.98	2.31×10^{-1}	2.28×10^{-1}
EXC	0.98	2.16×10^{-1}	2.14×10^{-1}
AR1	0.98	2.17×10^{-1}	2.15×10^{-1}
SMV	1.00	1.08×10^{-1}	1.07×10^{-1}
QIFE	0.99	9.30×10^{-2}	9.18×10^{-2}
TRUE	1.00	9.05×10^{-2}	8.92×10^{-2}
3-block (2 exchangeable and 1 AR1): $N = 100$ and $J = 20$			
IND	0.99	4.12×10^{-1}	4.06×10^{-1}
EXC	0.99	4.16×10^{-1}	4.10×10^{-1}
AR1	0.98	3.81×10^{-1}	3.75×10^{-1}
SMV	0.98	3.87×10^{-1}	3.81×10^{-1}
QIFE	0.99	3.65×10^{-1}	3.60×10^{-1}
TRUE	1.01	3.56×10^{-1}	3.50×10^{-1}
3-block (2 exchangeable and 1 AR1): $N = 100$ and $J = 50$			
IND	1.00	1.75×10^{-1}	1.72×10^{-1}
EXC	1.02	1.47×10^{-1}	1.46×10^{-1}
AR1	1.01	1.67×10^{-1}	1.64×10^{-1}
SMV	1.01	1.44×10^{-1}	1.43×10^{-1}
QIFE	1.01	1.15×10^{-1}	1.14×10^{-1}
TRUE	1.02	1.00×10^{-1}	1.00×10^{-1}

Table 2

The number of eigenvector bases selected in various cases.

	0	1	2	3	4	5	6	7
Normal (blocks: 2 exchangeable): $N = 100, J = 20$	0	44	53	3	0	0	0	0
Normal (blocks: 2 exchangeable): $N = 100, J = 50$	0	47	50	3	0	0	0	0
Normal (blocks: 2 exchangeable and 1 AR1): $N = 100, J = 20$	0	0	11	79	8	1	1	0
Normal (blocks: 2 exchangeable and 1 AR1): $N = 100, J = 50$	0	0	70	25	3	2	0	0
Binary (blocks: 2 AR1 and 1 exchangeable): $N = 100, J = 20$	9	3	21	38	24	3	2	0
Binary (blocks: 2 exchangeable and 1 AR1): $N = 100, J = 20$	0	0	23	74	3	0	0	0

Table 3Comparisons in the case of binary data with $I = 100$

method	$\text{mean}(\hat{\beta})$	$\text{sd}(\hat{\beta})$	$\sqrt{\text{mse}(\hat{\beta})}$
three blocks, 2AR1, 1 exchangeable, $J = 40$			
IND	4.98×10^{-1}	5.87×10^{-2}	5.79×10^{-2}
EXC	5.00×10^{-1}	5.90×10^{-2}	5.90×10^{-2}
AR1	5.04×10^{-1}	5.93×10^{-2}	5.93×10^{-2}
SMV	5.04×10^{-1}	5.72×10^{-2}	5.72×10^{-2}
QIFE	5.00×10^{-1}	5.45×10^{-2}	5.47×10^{-2}
TRUE	5.14×10^{-1}	5.82×10^{-2}	5.82×10^{-2}
three blocks, 2 exchangeable, 1 AR1, $J = 30$			
IND	4.98×10^{-1}	7.05×10^{-2}	6.94×10^{-2}
EXC	5.03×10^{-1}	6.24×10^{-2}	6.15×10^{-2}
AR1	5.04×10^{-1}	6.72×10^{-2}	6.62×10^{-2}
SMV	5.02×10^{-1}	5.47×10^{-2}	5.39×10^{-2}
QIFE	5.02×10^{-1}	4.74×10^{-2}	4.67×10^{-2}
TRUE	5.05×10^{-1}	4.38×10^{-2}	4.35×10^{-2}

Table 4

Model estimation results of the air pollution data set using various correlation basis matrices. The parameter estimators (the standard errors) are shown.

factors	IND	EXC	ARI	SMV	QIFE
MTEMP	-2.49×10^{-1} (0.42)	-1.35×10^{-1} (0.1)	3.8×10^{-2} (0.11)	-1.5×10^{-1} (0.09)	-1.73×10^{-1} (0.1)
NO	2.86×10^{-1} (0.47)	1.54×10^{-1} (0.09)	-3.01×10^{-2} (0.13)	1.77×10^{-1} (0.08)	1.98×10^{-1} (0.09)
NO2	-1.05×10^{-2} (0.09)	-4.79×10^{-3} (0.03)	1.66×10^{-2} (0.03)	7.77×10^{-4} (0.03)	-2.26×10^{-3} (0.03)
NOX	-2.72×10^{-1} (0.36)	-5.09×10^{-2} (0.04)	-3.72×10^{-2} (0.07)	-5.38×10^{-2} (0.04)	-6.26×10^{-2} (0.04)
TRS	-1.78×10^{-1} (0.26)	-1.94×10^{-2} (0.05)	-5.66×10^{-4} (0.06)	-1.33×10^{-2} (0.04)	-1.45×10^{-2} (0.04)
OZ	1.27×10^{-1} (0.14)	7.8×10^{-2} (0.04)	9.32×10^{-2} (0.04)	5.96×10^{-2} (0.04)	6.98×10^{-2} (0.04)
CO	-1.22×10^{-2} (0.3)	-2.31×10^{-2} (0.05)	-7.08×10^{-2} (0.07)	-1.3×10^{-2} (0.05)	-1.44×10^{-2} (0.05)
COH	1.19×10^{-1} (0.15)	-3.29×10^{-2} (0.03)	9.73×10^{-3} (0.02)	-2.03×10^{-2} (0.03)	-2.18×10^{-2} (0.03)

Table 5

Model estimation results of the sonar data set using various correlation basis matrices. The parameter estimators (the standard errors) are shown.

	IND	SMV	QIFE
Intercept	0.2739 (0.0045)	0.2487 (0.0052)	0.2838 (0.0042)
Slope	1.1131 (0.5429)	2.4067 (0.6933)	0.4758 (0.3588)