An estimating equation approach to dimension reduction for longitudinal data

BY KELIN XU

*State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University,
2005 Songhu Road, Shanghai 200438, China*
kelinxu12@fudan.edu.cn

WENSHENG GUO

*Division of Biostatistics, University of Pennsylvania, 423 Guardian Drive, Philadelphia,
Pennsylvania 19104, U.S.A.*
wguo@mail.med.upenn.edu

MOMIAO XIONG

*Division of Biostatistics, Human Genetics Center, The University of Texas Health Science
Center at Houston, 1200 Herman Pressler, Houston, Texas 77030, U.S.A.*
momiao.xiong@uth.tmc.edu

LIPING ZHU

*Institute of Statistics and Big Data, Renmin University of China, 59 Zhongguancun Avenue,
Beijing 100872, China*
zhu.liping@mail.shufe.edu.cn

AND LI JIN

*State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University,
2005 Songhu Road, Shanghai 200438, China*
lijin@fudan.edu.cn

SUMMARY

Sufficient dimension reduction has been extensively explored in the context of independent and identically distributed data. In this article we generalize sufficient dimension reduction to longitudinal data and propose an estimating equation approach to estimating the central mean subspace. The proposed method accounts for the covariance structure within each subject and improves estimation efficiency when the covariance structure is correctly specified. Even if the covariance structure is misspecified, our estimator remains consistent. In addition, our method relaxes distributional assumptions on the covariates and is doubly robust. To determine the structural dimension of the central mean subspace, we propose a Bayesian-type information criterion. We show that the estimated structural dimension is consistent and that the estimated basis directions are root- n consistent, asymptotically normal and locally efficient. Simulations and an analysis of the Framingham Heart Study data confirm the effectiveness of our approach.

Some key words: Central mean subspace; Dimension reduction; Estimating equation; Longitudinal data; Semiparametric efficiency; Sliced inverse regression.

1. INTRODUCTION

Sufficient dimension reduction is an effective means of reducing covariate dimension in regression. Since the seminal work of [Li \(1991\)](#), many sufficient dimension reduction methods have been developed to deal with independent and identically distributed data. See, for example, [Cook & Weisberg \(1991\)](#), [Xia et al. \(2002\)](#), [Ma & Zhu \(2012\)](#) and [Lee et al. \(2013\)](#). For comprehensive reviews, see [Cook \(1998b\)](#), [Cook & Yin \(2001\)](#), [Ma & Zhu \(2013\)](#) and references therein.

In this article we generalize sufficient dimension reduction to longitudinal data, which arise frequently in biometric, epidemiological, social and economic studies that typically involve repeated measurements collected from individuals over time. Consider a typical longitudinal dataset, where there are n subjects, the i th subject is observed J_i times, and the observations are $\{(X_{ij}, Y_{ij}) : j = 1, \dots, J_i; i = 1, \dots, n\}$. Here X_{ij} is a $p \times 1$ covariate vector and Y_{ij} is a univariate response variable. In the sufficient dimension reduction context, we assume that the covariates are multivariate and there is a $p \times d_0$ matrix β , for some $d_0 < p$, such that

$$E(Y_{ij} | X_{ij}) = E(Y_{ij} | \beta^T X_{ij}). \quad (1)$$

Model (1) indicates that the covariates X_{ij} can be replaced with d_0 linear combinations, $\beta^T X_{ij}$, and that such a replacement will not cause any information loss in estimating the conditional mean of the response. We assume that the dimension p of X_{ij} is large while the column dimension d_0 of β is small, so that the dimension-reduced mean function $E(Y_{ij} | \beta^T X_{ij})$ is nonparametrically manageable. The minimum column space of β is called the central mean subspace, denoted by $\mathcal{S}_{E(Y|X)}$; see [Cook & Li \(2002\)](#). Write $\mathcal{S}_{E(Y|X)} = \text{span}(\beta)$. The column dimension d_0 of β is called the structural dimension of $\mathcal{S}_{E(Y|X)}$.

There are many approaches to estimating $\mathcal{S}_{E(Y|X)}$ when the observations are independent and identically distributed. See, for example, [Li & Duan \(1989\)](#) and [Härdle & Stoker \(1989\)](#) for the case where $\mathcal{S}_{E(Y|X)}$ is one-dimensional, and [Li \(1992\)](#), [Cook \(1998a\)](#) and [Cook & Li \(2002\)](#) for when $\mathcal{S}_{E(Y|X)}$ is not necessarily one-dimensional. The methods in the aforementioned works are computationally efficient, yet their application is limited by distributional assumptions on the covariates, such as the linearity and constant variance conditions. [Xia et al. \(2002\)](#) proposed minimum average variance estimation of $\mathcal{S}_{E(Y|X)}$. [Ma & Zhu \(2012\)](#) proposed a semiparametric approach which removes distributional assumptions on the covariates, and [Ma & Zhu \(2014\)](#) showed that an efficient estimator of $\mathcal{S}_{E(Y|X)}$ is not available even when the observations are independent and identically distributed.

The amount of research on sufficient dimension reduction for longitudinal data is far from adequate. When $\mathcal{S}_{E(Y|X)}$ is one-dimensional, [Bai et al. \(2009\)](#) proposed a quadratic inference function approach. [Jiang & Wang \(2011\)](#) adapted the minimum average variance estimation of [Xia et al. \(2002\)](#) to longitudinal data by assuming the covariates to be stochastic continuous functions. [Li & Yin \(2009\)](#) generalized several existing sufficient dimension reduction methods to longitudinal data; they assumed that the covariates satisfy the linearity condition and ignored the covariance structure within each subject in estimation. [Bi & Qu \(2015\)](#) used quadratic inference functions on transformed response variables to perform dimension reduction for longitudinal data. Their approach incorporated the covariance structure; however, their focus was on estimation of the central subspace rather than inference for the central mean subspace.

In this article, we adapt the generalized estimating equations of [Liang & Zeger \(1986\)](#) and the semiparametric approach of [Ma & Zhu \(2012\)](#) to longitudinal data analysis. This approach has several advantages over existing sufficient dimension reduction methods. First, the estimating equation approach accounts for the covariance structure within each subject, which is generally unknown and has to be estimated from data. If the estimated covariance matrix is consistent, we gain efficiency in estimating $\mathcal{S}_{E(Y|X)}$; if the estimated covariance matrix is inconsistent, we can still obtain a consistent estimator of $\mathcal{S}_{E(Y|X)}$. Second, the estimating equation approach has the double robustness property, which will be described in § 2. This ensures that we can obtain a consistent estimator of $\mathcal{S}_{E(Y|X)}$ even when the estimated mean function is inconsistent. It also implicitly explains the role that the linearity condition plays in the classical sufficient dimension reduction literature; in particular, linearity helps to simplify estimation of $\mathcal{S}_{E(Y|X)}$. Third, the estimating equation approach relaxes the distributional assumptions on the covariates. It can be readily used even when the linearity condition or the constant variance condition is violated. Fourth, because the asymptotic covariance matrix of our estimator can be estimated without much difficulty, the estimating equation approach facilitates inference for $\mathcal{S}_{E(Y|X)}$.

2. AN ESTIMATING EQUATION APPROACH

In the context of independent and identically distributed data, [Ma & Zhu \(2012\)](#) designed a semiparametric approach to estimating $\mathcal{S}_{E(Y|X)}$. In this section we follow the idea of generalized estimating equations proposed by [Liang & Zeger \(1986\)](#) and further adapt the semiparametric approach of [Ma & Zhu \(2012\)](#) to the estimation of $\mathcal{S}_{E(Y|X)}$ for longitudinal data.

Our main interest is in estimation and inference for $\mathcal{S}_{E(Y|X)}$. Inference for $\mathcal{S}_{E(Y|X)}$ is not straightforward, as the semiparametric tools developed by [Bickel et al. \(1993\)](#) are not applicable if the quantities of interest are spaces. Therefore, to make inference for $\mathcal{S}_{E(Y|X)}$, we are forced to convert space inference to parameter inference. This amounts to characterizing $\mathcal{S}_{E(Y|X)}$ with a unique set of parameters. Recall that $\mathcal{S}_{E(Y|X)} = \text{span}(\beta)$. Although $\mathcal{S}_{E(Y|X)}$ is unique, its basis β is not. Following [Ma & Zhu \(2014\)](#), we parameterize $\mathcal{S}_{E(Y|X)}$ with β taking the particular form $\beta = (I_{d_0}, \beta_{l_0}^\top)^\top$, where I_{d_0} stands for the $d_0 \times d_0$ identity matrix, β_{l_0} stands for the lower $(p - d_0) \times d_0$ submatrix of β , and d_0 is the dimension of $\mathcal{S}_{E(Y|X)}$. Estimating $\mathcal{S}_{E(Y|X)}$ amounts to estimating d_0 and β_{l_0} . For clarity, we discuss the estimation of β_{l_0} given d_0 in this section. Estimation of d_0 will be presented in § 3.

We propose an estimating equation approach to estimating β_{l_0} . Recall the notation X_{ij} and Y_{ij} defined in § 1. We write $m(\beta^\top X_{ij}) = E(Y_{ij} | \beta^\top X_{ij})$ and $m'(\beta^\top X_{ij})$ for the first derivative of $m(\beta^\top X_{ij})$ with respect to $(\beta^\top X_{ij})$. By definition, $m'(\beta^\top X_{ij})$ is a $1 \times d_0$ row vector. Let $\otimes$ stand for the Kronecker product operator. We let $X_{ij,-d_0}$ denote the last $p - d_0$ components of X_{ij} and V_i the $J_i \times J_i$ working covariance matrix within the i th subject. To be precise, V_i stands for the working covariance matrix of $Y_{i1}, \dots, Y_{iJ_i}$ when $X_{i1}, \dots, X_{iJ_i}$ are given. Following [Liang & Zeger \(1986\)](#) and slightly modifying the estimating equation approach of [Ma & Zhu \(2012\)](#) designed for independent and identically distributed data, we propose to estimate $\mathcal{S}_{E(Y|X)}$ by solving the estimating equations

$$\sum_{i=1}^n \begin{pmatrix} Y_{i1} - \hat{m}(b^\top X_{i1}) \\ Y_{i2} - \hat{m}(b^\top X_{i2}) \\ \vdots \\ Y_{iJ_i} - \hat{m}(b^\top X_{iJ_i}) \end{pmatrix}^\top \hat{V}_i^{-1} \begin{pmatrix} \hat{m}'(b^\top X_{i1}) \otimes \{X_{i1,-d_0} - \hat{E}(X_{i1,-d_0} | b^\top X_{i1})\}^\top \\ \hat{m}'(b^\top X_{i2}) \otimes \{X_{i2,-d_0} - \hat{E}(X_{i2,-d_0} | b^\top X_{i2})\}^\top \\ \vdots \\ \hat{m}'(b^\top X_{iJ_i}) \otimes \{X_{iJ_i,-d_0} - \hat{E}(X_{iJ_i,-d_0} | b^\top X_{iJ_i})\}^\top \end{pmatrix} = 0, \quad (2)$$

where the upper $d_0 \times d_0$ submatrix of b is fixed as the identity matrix I_{d_0} and the free parameters are the entries in the lower $(p - d_0) \times d_0$ submatrix of b . The estimating equations (2) are unbiased if $E(Y_{ij} | X_{i1}, \dots, X_{iJ_i}) = E(Y_{ij} | X_{ij})$, which is widely assumed in longitudinal data analysis and often referred to as the population average condition (Liang & Zeger, 1986). We solve (2) to obtain the estimator $\hat{\beta}_V = (I_{d_0}, \hat{\beta}_{l_0, V}^T)^T$. The subscript V indicates that the resulting solution to (2), $\hat{\beta}_{l_0, V}$ and hence $\hat{\beta}_V$, depends on the working covariance structure V_i ($i = 1, \dots, n$). To put (2) into practice, we assume that $E(Y_{ij} | b^T X_{ij})$ and $E(X_{ij} | b^T X_{ij})$ are smooth functions. We propose to estimate m and m' simultaneously by local polynomial regression (Fan & Gijbels, 1996) and to estimate $E(X_{ij} | b^T X_{ij})$ by kernel regression, yielding $\hat{m}(b^T X_{ij})$, $\hat{m}'(b^T X_{ij})$ and $\hat{E}(X_{ij, -d} | b^T X_{ij})$, respectively. To implement (2), it now remains to estimate the covariance matrix V_i . Liang & Zeger (1986) assumed that the working covariance matrix V_i has the form

$$\text{diag}\{\text{std}(Y_{i1} | X_{i1}), \dots, \text{std}(Y_{iJ_i} | X_{iJ_i})\} R_i(\alpha) \text{diag}\{\text{std}(Y_{i1} | X_{i1}), \dots, \text{std}(Y_{iJ_i} | X_{iJ_i})\},$$

where $\text{diag}(\cdot)$ is a diagonal matrix with its arguments as the diagonal elements, $\text{std}(Y_{ij} | X_{ij}) = \text{var}(Y_{ij} | X_{ij})^{1/2}$ is a constant, and $R_i(\alpha)$ stands for a $J_i \times J_i$ correlation matrix which can be characterized by a vector α . The strategy of using a working covariance is widely employed in longitudinal data analysis. See, for example, Horton & Lipsitz (1999) for a review, and Wedderburn (1974) and McCullagh (1983) for the quasiliikelihood theory. Following Wedderburn (1974) and McCullagh (1983), we consider the following three structures for $\text{var}(Y_{ij} | X_{ij})$.

- (i) If Y_{ij} is binary, we estimate $\text{var}(Y_{ij} | X_{ij})$ by

$$\hat{\text{var}}(Y_{ij} | X_{ij}) = \hat{m}(\hat{\beta}_V^T X_{ij}) \{1 - \hat{m}(\hat{\beta}_V^T X_{ij})\},$$

where $\hat{m}(\hat{\beta}_V^T X_{ij}) = \hat{E}(Y_{ij} | \hat{\beta}_V^T X_{ij})$.

- (ii) If Y_{ij} is discrete and nonnegative and takes infinitely many values, we estimate $\text{var}(Y_{ij} | X_{ij})$ by

$$\hat{\text{var}}(Y_{ij} | X_{ij}) = \hat{m}(\hat{\beta}_V^T X_{ij}),$$

which implicitly assumes that $(Y_{ij} | X_{ij})$ follows a Poisson distribution.

- (iii) If Y_{ij} is continuous, we estimate $\text{var}(Y_{ij} | X_{ij})$ by

$$\hat{\text{var}}(Y_{ij} | X_{ij}) = \sigma_0^2$$

for all i and j , where σ_0 is an unknown constant.

Note that $R_i(\alpha)$ is a constant correlation matrix but V_i is not. The diagonal elements in $R_i(\alpha)$ are identically 1, and the off-diagonal elements depend on an unknown parameter α . We consider the following four working correlation structures, as suggested by Liang & Zeger (1986) and Horton & Lipsitz (1999):

- (i) the triangular working correlation structure, where all elements on the main diagonal of $R_i(\alpha)$ equal 1, all elements on the first diagonals below and above the main diagonal equal α , and all the other elements equal 0;
- (ii) the AR(1) working correlation structure, where the (j, j') th element of $R_i(\alpha)$ is $\alpha^{|j-j'|}$ for $1 \leq j, j' \leq J_i$;
- (iii) the independent working correlation structure, where the off-diagonal elements are identically zero, so that no parameters have to be estimated in $R_i(\alpha)$;

- (iv) the unstructured working correlation structure, where the (j, j') th element of $R_i(\alpha)$ is exactly $\alpha(j, j') = \text{corr}(\epsilon_{ij}, \epsilon_{ij'})$, with $\epsilon_{ij} = Y_{ij} - m(\beta^T X_{ij})$ for $1 \leq j, j' \leq J_i$.

For the triangular and AR(1) working correlation structures, we estimate α by

$$\hat{\alpha} = \left[\{N - d_0 \times (p - d_0)\} \sum_{i=1}^n \sum_{j=1}^{J_i-1} \{Y_{ij} - \hat{m}(\hat{\beta}_V^T X_{ij})\} \{Y_{i,j+1} - \hat{m}(\hat{\beta}_V^T X_{i,j+1})\} \right] \\ / \left[\{N - n - d_0 \times (p - d_0)\} \sum_{i=1}^n \sum_{j=1}^{J_i} \{Y_{ij} - \hat{m}(\hat{\beta}_V^T X_{ij})\}^2 \right],$$

where $N = \sum_{i=1}^n J_i$. For the unstructured working correlation structure, we estimate $\alpha(j, j')$ by

$$\hat{\alpha}(j, j') = \left[\{N - d_0 \times (p - d_0)\} \sum_{i=1}^n \{Y_{ij} - \hat{m}(X_{ij}^T \hat{\beta}_V)\} \{Y_{ij'} - \hat{m}(X_{ij'}^T \hat{\beta}_V)\} \right] \\ / \left[\{n_{jj'}^* - d_0 \times (p - d_0)\} \sum_{i=1}^n \sum_{j=1}^{J_i} \{Y_{ij} - \hat{m}(X_{ij}^T \hat{\beta}_V)\}^2 \right],$$

where $n_{jj'}^* = \sum_{i=1}^n I(J_i \geq j)I(J_i \geq j')$, with $I(\cdot)$ denoting the indicator function. We estimate $\text{var}(Y_{ij} | X_{ij})$ and $R_i(\alpha)$ as above to obtain $\hat{V}_i$. Inserting $\hat{V}_i$, $\hat{m}(b^T X_{ij})$, $\hat{m}'(b^T X_{ij})$ and $\hat{E}(x_{ij, -d_0} | b^T X_{ij})$ into the estimating equations (2), we can obtain $\hat{\beta}_{l_0, V}$ and, accordingly, $\hat{\beta}_V$ by solving (2) iteratively.

Let us make the following remarks about the estimating equations (2). First, the estimating equation approach is doubly robust. To be precise, if the population average condition $E(Y_{ij} | X_{i1}, \dots, X_{iJ_i}) = E(Y_{ij} | X_{ij})$ holds and $\hat{m}(\beta^T X_{ij})$ is a consistent estimator of $m(\beta^T X_{ij})$, then the estimator obtained from (2), namely $\hat{\beta}_V$, is consistent for β whether or not $\hat{E}(X_{ij, -d_0} | \beta^T X_{ij})$ is a consistent estimator of $E(X_{ij, -d_0} | \beta^T X_{ij})$. If the working covariance matrix V_i is a function of $\beta^T X_{ij}$ and $\hat{E}(X_{ij, -d_0} | \beta^T X_{ij})$ is a consistent estimator of $E(X_{ij, -d_0} | \beta^T X_{ij})$, then $\hat{\beta}_V$ is consistent for β whether or not $\hat{m}(\beta^T X_{ij})$ is a consistent estimator of $m(\beta^T X_{ij})$. In other words, if one of $\hat{m}(\beta^T X_{ij})$ and $\hat{E}(X_{ij, -d_0} | \beta^T X_{ij})$ is consistent, $\hat{\beta}_{l_0, V}$ will be consistent, echoing the findings of [Ma & Zhu \(2012\)](#). However, if neither $\hat{m}(\beta^T X_{ij})$ nor $\hat{E}(X_{ij, -d_0} | \beta^T X_{ij})$ is consistent, $\hat{\beta}_V$ will be inconsistent ([Kang & Schafer, 2007](#)). We advocate using local smoothing to estimate both $m(\beta^T X_{ij})$ and $E(X_{ij, -d_0} | \beta^T X_{ij})$, because nonparametric smoothing techniques yield consistent estimates under mild conditions.

Second, the double robustness property has an important implication. Specifically, if the linearity condition is indeed true, i.e., $E(X_{ij} | \beta^T X_{ij})$ is a linear function of X_{ij} , and we estimate $E(X_{ij} | \beta^T X_{ij})$ by linear regression, which is typically consistent under mild conditions, we can misspecify $\hat{m}(\beta^T X_{ij})$ as the sample mean of $y_i = (Y_{i1}, \dots, Y_{iJ_i})^T$ and $\hat{m}'(\beta^T X_{ij})$ as an arbitrary constant, and then the solution to (2) is the ordinary least squares estimator when $\mathcal{S}_{E(Y|X)}$ is one-dimensional. In other words, when the linearity condition holds, the ordinary least squares estimator lies in $\mathcal{S}_{E(Y|X)}$, which generalizes the result of [Li & Duan \(1989\)](#) to the longitudinal data scenario. This also indicates that the linearity condition, when it holds, can simplify the estimation of $\mathcal{S}_{E(Y|X)}$.

Third, to specify the covariance structure V_i , we assume three working variance functions, depending on the nature of the response variable, and four working correlation functions. Although these working structures may not reflect the underlying true covariance structures, the resulting estimator $\hat{\beta}_V$ is always consistent. If these working structures happen to be correct, we will show that $\hat{\beta}_V$ is efficient. Even if V_i is incorrect, $\hat{\beta}_V$ is consistent and the estimated V_i , denoted by $\hat{V}_i$, approximates the true covariance matrix. In addition, $\hat{\beta}_V$ is consistent for β whether or not $\hat{m}'(\beta^T X_{ij})$ is consistent for $m'(\beta^T X_{ij})$.

To state the asymptotic properties of $\hat{\beta}_{l_0, V}$, we assume the following regularity conditions.

Condition 1. The univariate kernel function $K(\cdot)$ is a bounded symmetric probability density function, has compact support and is Lipschitz-continuous on its support. It satisfies

$$\int \mu^i K(\mu) d\mu = 0 \quad (i = 1, \dots, m-1), \quad 0 \neq \int \mu^m K(\mu) d\mu < \infty.$$

Condition 2. The density function $f(\beta^T X)$ of $(\beta^T X)$ is bounded away from zero on a finite compact set. The $(m-1)$ th derivatives of $f(X^T \beta)$, $m(\beta^T X)f(\beta^T X)$ and $E(X | X^T \beta)f(\beta^T X)$ are Lipschitz-continuous.

Condition 3. The bandwidth h satisfies $h = O(n^{-\kappa})$ for $(4m)^{-1} < \kappa < (2d_0)^{-1}$.

Condition 4. The moments $E(X^T X)$, $E(Y^2)$, $E\{m(\beta^T X)\}^2$ and $E\{m'(\beta^T X)m'(\beta^T X)^T\}$ are finite.

Condition 5. The repetition number J_i is uniformly bounded away from infinity.

THEOREM 1. Under Conditions 1–5, as $n \rightarrow \infty$,

$$\sqrt{n}\{\text{vec}(\hat{\beta}_{l_0, V}) - \text{vec}(\beta_{l_0})\} \rightarrow N\{0, A_V^{-1} B_V A_V^{-1}\}$$

in distribution, with

$$A_V = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i^T V_i^{-1} C_i, \quad B_V = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i^T V_i^{-1} \Sigma_i V_i^{-1} C_i,$$

where

$$C_i = \begin{pmatrix} m'(\beta^T X_{i1}) \otimes \{X_{i1, -d_0} - E(X_{i1, -d_0} | \beta^T X_{i1})\}^T \\ m'(\beta^T X_{i2}) \otimes \{X_{i2, -d_0} - E(X_{i2, -d_0} | \beta^T X_{i2})\}^T \\ \vdots \\ m'(\beta^T X_{iJ_i}) \otimes \{X_{iJ_i, -d_0} - E(X_{iJ_i, -d_0} | \beta^T X_{iJ_i})\}^T \end{pmatrix},$$

V_i denotes the expectation of $\hat{V}_i$, and $\Sigma_i = \text{cov}(Y_{i1}, \dots, Y_{iJ_i} | X_{i1}, \dots, X_{iJ_i})$ is the underlying true covariance structure of $Y_{i1}, \dots, Y_{iJ_i}$ when $X_{i1}, \dots, X_{iJ_i}$ are given.

Let $\hat{\beta}_{l_0, \Sigma}$ denote the estimator when either the true or the consistently estimated covariance matrix, Σ_i or $\hat{\Sigma}_i$, is used in (2). By Theorem 1, the asymptotic variance of $\hat{\beta}_{l_0, \Sigma}$ is

$n^{-1}A_{\Sigma}^{-1}B_{\Sigma}A_{\Sigma}^{-1}$, where

$$A_{\Sigma} = B_{\Sigma} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i^{\top} \Sigma_i^{-1} C_i.$$

PROPOSITION 1. *If A_V and A_{Σ} are nonsingular, then $A_V^{-1}B_VA_V^{-1} \geq A_{\Sigma}^{-1}B_{\Sigma}A_{\Sigma}^{-1} = A_{\Sigma}^{-1}$.*

Proposition 1 says that if we opt to use Σ_i or $\hat{\Sigma}_i$ as the working covariance matrix V_i in (2) to estimate $\mathcal{S}_{E(Y|X)}$, the estimators are semiparametrically efficient.

3. ESTIMATION OF STRUCTURAL DIMENSION

In this section, we extend the work of Xia et al. (2002) and Zhu et al. (2006) and design a Bayesian-type information criterion to estimate the structural dimension d_0 in the longitudinal setting. Unlike the sequential test of Cook & Li (2004), our estimator does not require distributional assumptions on the covariates and is consistent.

For an arbitrary working dimension d , we define

$$\begin{aligned} \mathcal{L}(d) &= \sum_{i=1}^n \begin{pmatrix} Y_{i1} - \hat{m}(\hat{\beta}_d^{\top} X_{i1}) \\ Y_{i2} - \hat{m}(\hat{\beta}_d^{\top} X_{i2}) \\ \vdots \\ Y_{iJ_i} - \hat{m}(\hat{\beta}_d^{\top} X_{iJ_i}) \end{pmatrix}^{\top} \hat{V}_{d,i}^{-1} \begin{pmatrix} Y_{i1} - \hat{m}(\hat{\beta}_d^{\top} X_{i1}) \\ Y_{i2} - \hat{m}(\hat{\beta}_d^{\top} X_{i2}) \\ \vdots \\ Y_{iJ_i} - \hat{m}(\hat{\beta}_d^{\top} X_{iJ_i}) \end{pmatrix}, \\ G(d) &= \mathcal{L}(d) + dC_n, \end{aligned} \quad (3)$$

where $\hat{\beta}_d$ and $\hat{V}_{d,i}$ are estimated as described in §2 and the subscript d indicates that they are estimated under the working dimension d . In general the loss function $\mathcal{L}(d)$ is monotonically decreasing because, typically, with more dimensions and more parameters, we can estimate the mean function more precisely. Intuitively, if $d < d_0$, then $\mathcal{L}(d) - \mathcal{L}(d_0) = O_p(n)$, and if $d \geq d_0$, then $\mathcal{L}(d_0) - \mathcal{L}(d) = o_p(n)$. This inspires us to choose the penalty term C_n to estimate the structural dimension d_0 . We define the estimated dimension as

$$\hat{d} = \arg \min_{1 \leq d \leq p} G(d).$$

The consistency of this Bayesian-type information criterion is stated in Theorem 2.

THEOREM 2. *If, in addition to the conditions of Theorem 1, $C_n/n^{1/2} \rightarrow \infty$ and $C_n/n \rightarrow 0$ as $n \rightarrow \infty$, then $\hat{d}$ converges to d_0 in probability.*

Theorem 2 implies that the consistency of this Bayesian-type information criterion holds for a wide range of C_n . How to choose an optimal C_n is a challenging problem, but from our limited experience, the choice of

$$C_n = N^{-1/3} \sum_{i=1}^n \sum_{j=1}^{J_i} (Y_{ij} - \bar{Y})^2, \quad N = \sum_{i=1}^n J_i, \quad \bar{Y} = N^{-1} \sum_{i=1}^n \sum_{j=1}^{J_i} Y_{ij} \quad (4)$$

seems to work well in our numerical studies.

4. SIMULATIONS

We first evaluate the performance of the estimating equation approach. The number of observations J_i for each subject is drawn from a uniform distribution on $\{3, \dots, 12\}$. Let p be the covariate dimension and n the number of subjects. We fix $p = 6$ and $n = 50$, and generate Y_{ij} from the following two-dimensional central mean subspace model (Li, 1991):

$$Y_{ij} = \frac{X_{ij}^T \beta_1}{0.5 + (X_{ij}^T \beta_2 + 1.5)^2} + \epsilon_{ij}, \quad (5)$$

where $\beta_1 = (1, 0, 0.5, 0.3, -0.3, 0.8)^T$ and $\beta_2 = (0, 1, 0.5, -0.3, 0.3, -0.8)^T$.

Following Hedeker & Gibbons (2006), we consider two scenarios of generating the covariate vector $X_{ij} \in \mathbb{R}^6$. In the first scenario, all covariates are time-varying. We generate the first five components from a multivariate normal distribution with mean zero and covariance matrix $(0.5^{|k-l|})_{5 \times 5}$, and generate the sixth component from a Bernoulli distribution with success probability 0.5. In the second scenario, some covariates are time-invariant. We generate the first four components and the last component in the same way as in the previous time-varying scenario; to generate the fifth component, we first generate an n -vector from the standard normal distribution, and then repeat the i th value of this vector J_i times, for $i = 1, \dots, n$. Thus, the fifth component is time-invariant.

The errors ϵ_{ij} in (5) are generated from a multivariate normal distribution with mean zero and covariance matrix Σ_i of the following three types. The first is a tridiagonal covariance structure: $\text{cov}(\epsilon_{ij}, \epsilon_{ij'})$ equals 1 if $|j - j'| = 0$, 0.4 if $|j - j'| = 1$ and 0 if $|j - j'| > 1$. The second is an AR(1) covariance structure: $\text{cov}(\epsilon_{ij}, \epsilon_{ij'}) = 0.4^{|j-j'|}$. The third is an independent structure: $\text{cov}(\epsilon_{ij}, \epsilon_{ij'})$ equals 1 if $|j - j'| = 0$ and equals 0 otherwise.

We compare the performance of the following eight estimators:

- (i) the partial dimension reduction method proposed by Li & Yin (2009);
- (ii) the quadratic inference function method proposed by Bi & Qu (2015), using the true covariance structure as the working covariance structure;
- (iii) the estimating equation approach (2), with the estimated working covariance structure $\hat{V}_i$ specified as a tridiagonal structure;
- (iv) the estimating equation approach (2), with $\hat{V}_i$ specified as an AR(1) structure;
- (v) the estimating equation approach (2), with $\hat{V}_i$ specified as an independent structure;
- (vi) the estimating equation approach (2), with $\hat{V}_i$ specified as an unstructured structure;
- (vii) the estimating equation approach (2), where the centralization for $X_{ij,-d}$ is misspecified; to be specific, we solve the estimating equations

$$\sum_{i=1}^n \begin{pmatrix} Y_{i1} - \hat{m}(b^T X_{i1}) \\ Y_{i2} - \hat{m}(b^T X_{i2}) \\ \vdots \\ Y_{iJ_i} - \hat{m}(b^T X_{iJ_i}) \end{pmatrix}^T \Sigma_i^{-1} \begin{pmatrix} \hat{m}'(b^T X_{i1}) \otimes (X_{i1,-d} - \hat{X}_{i1,-d})^T \\ \hat{m}'(b^T X_{i2}) \otimes (X_{i2,-d} - \hat{X}_{i2,-d})^T \\ \vdots \\ \hat{m}'(b^T X_{iJ_i}) \otimes (X_{iJ_i,-d} - \hat{X}_{iJ_i,-d})^T \end{pmatrix} = 0,$$

where $\hat{X}_{ij,-d}$ is the fitted value of $X_{ij,-d}$ through regressing $X_{ij,-d}$ on $b^T X_{ij}$ linearly ($i = 1, \dots, n$; $j = 1, \dots, J_i$);

Table 1. *Simulation results based on 1000 repetitions in the case where all the covariates are time-varying and the underlying true covariance admits a tridiagonal structure*

		β_{13}	β_{14}	β_{15}	β_{16}	β_{23}	β_{24}	β_{25}	β_{26}
PDR	bias	-0.25	-0.05	-0.04	-0.38	-1.07	2.03	-0.75	1.59
	std	0.79	0.78	0.62	0.63	1.71	1.95	1.65	1.40
QIF	bias	0.07	-0.02	0.04	-1.23	0.29	-0.14	0.16	-3.30
	std	0.22	0.19	0.19	0.57	0.50	0.44	0.44	1.27
TRI	bias	0.01	0.01	-0.01	0.04	0.03	-0.02	0.02	-0.04
	std	0.12	0.10	0.09	0.13	0.10	0.09	0.08	0.14
	$\hat{\text{std}}$	0.11	0.09	0.09	0.13	0.10	0.08	0.08	0.14
	cvp	0.94	0.94	0.95	0.95	0.94	0.94	0.95	0.96
AR(1)	bias	0.01	0.01	-0.01	0.05	0.03	-0.02	0.02	-0.04
	std	0.12	0.10	0.09	0.13	0.10	0.09	0.08	0.14
	$\hat{\text{std}}$	0.11	0.10	0.09	0.14	0.10	0.08	0.08	0.14
	cvp	0.93	0.94	0.96	0.95	0.94	0.95	0.94	0.95
IND	bias	0.01	0.02	-0.02	0.07	0.04	-0.02	0.02	-0.07
	std	0.14	0.11	0.10	0.16	0.11	0.10	0.09	0.17
	$\hat{\text{std}}$	0.12	0.11	0.10	0.15	0.11	0.09	0.09	0.17
	cvp	0.94	0.95	0.94	0.95	0.94	0.94	0.94	0.95
UN	bias	0.02	0.01	-0.01	0.10	0.03	-0.01	0.03	-0.09
	std	0.13	0.11	0.10	0.17	0.12	0.10	0.09	0.18
	$\hat{\text{std}}$	0.11	0.09	0.09	0.14	0.10	0.09	0.08	0.15
	cvp	0.89	0.88	0.89	0.91	0.89	0.89	0.89	0.91
X_{MIS}	bias	-0.07	0.04	-0.03	0.11	-0.01	0.09	-0.03	0.29
	std	0.06	0.04	0.09	0.13	0.09	0.07	0.04	0.27
	$\hat{\text{std}}$	0.10	0.09	0.09	0.18	0.09	0.10	0.08	0.18
	cvp	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88
Y_{MIS}	bias	0.04	0.12	-0.11	0.21	0.16	-0.11	0.11	-0.07
	std	0.18	0.16	0.15	0.20	0.16	0.16	0.15	0.22
	$\hat{\text{std}}$	0.17	0.15	0.14	0.20	0.16	0.15	0.14	0.21
	cvp	0.94	0.91	0.92	0.92	0.88	0.91	0.90	0.94

PDR, method of Li & Yin (2009); QIF, method of Bi & Qu (2015); TRI, tridiagonal working covariance structure; AR(1), AR(1) working covariance structure; IND, independent working covariance structure; UN, unstructured working covariance structure; X_{MIS} , centralization for $X_{ij,-d}$ misspecified; Y_{MIS} , centralization for Y_{ij} misspecified; bias, difference between average of the estimates and true value of the parameter; std, Monte Carlo standard deviation; $\hat{\text{std}}$, average of the estimated standard deviations; cvp, empirical coverage probability at 95% nominal level.

- (viii) the estimating equation approach (2), where the centralization for Y_{ij} is misspecified; specifically, we solve the estimating equations

$$\sum_{i=1}^n \begin{pmatrix} Y_{i1} - \hat{Y}_{i1} \\ Y_{i2} - \hat{Y}_{i2} \\ \vdots \\ Y_{iJ_i} - \hat{Y}_{iJ_i} \end{pmatrix}^T \Sigma_i^{-1} \begin{pmatrix} \hat{m}'(b^T X_{i1}) \otimes \{X_{i1,-d} - \hat{E}(X_{i1,-d} | b^T X_{i1})\}^T \\ \hat{m}'(b^T X_{i2}) \otimes \{X_{i2,-d} - \hat{E}(X_{i2,-d} | b^T X_{i2})\}^T \\ \vdots \\ \hat{m}'(b^T X_{iJ_i}) \otimes \{X_{iJ_i,-d} - \hat{E}(X_{iJ_i,-d} | b^T X_{iJ_i})\}^T \end{pmatrix} = 0,$$

Table 2. *Simulation results based on 1000 repetitions in the case where all the covariates are time-varying and the underlying true covariance admits an AR(1) structure*

		β_{13}	β_{14}	β_{15}	β_{16}	β_{23}	β_{24}	β_{25}	β_{26}
PDR	bias	0.69	3.20	-1.84	1.68	1.97	7.98	-5.05	5.66
	std	0.92	0.86	0.77	0.72	1.79	1.82	1.78	1.40
QIF	bias	0.01	0.05	-0.21	-0.85	-0.13	0.23	-0.45	-2.01
	std	0.38	0.38	0.34	1.15	1.05	1.02	0.93	2.82
TRI	bias	0.01	0.01	-0.01	0.05	0.03	-0.02	0.02	-0.05
	std	0.12	0.10	0.09	0.14	0.10	0.09	0.08	0.15
	sfd	0.11	0.10	0.09	0.14	0.10	0.09	0.08	0.15
	cvp	0.93	0.94	0.95	0.95	0.94	0.95	0.95	0.95
AR(1)	bias	0.01	0.01	-0.01	0.05	0.03	-0.02	0.02	-0.05
	std	0.12	0.10	0.09	0.14	0.10	0.09	0.08	0.15
	sfd	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.16
	cvp	0.96	0.96	0.96	0.97	0.96	0.96	0.97	0.97
IND	bias	0.01	0.02	-0.02	0.07	0.04	-0.03	0.03	-0.08
	std	0.13	0.11	0.10	0.16	0.11	0.10	0.09	0.16
	sfd	0.12	0.11	0.10	0.15	0.11	0.09	0.09	0.16
	cvp	0.94	0.93	0.94	0.95	0.94	0.94	0.95	0.96
UN	bias	0.01	0.02	-0.02	0.08	0.04	-0.02	0.03	-0.09
	std	0.14	0.12	0.11	0.17	0.12	0.10	0.09	0.18
	sfd	0.11	0.10	0.09	0.14	0.10	0.09	0.08	0.16
	cvp	0.91	0.90	0.92	0.92	0.91	0.93	0.91	0.94
X_{MIS}	bias	-0.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.13
	std	0.11	0.10	0.10	0.15	0.10	0.09	0.08	0.15
	sfd	0.11	0.10	0.09	0.14	0.10	0.09	0.08	0.17
	cvp	0.94	0.93	0.93	0.93	0.95	0.94	0.94	0.93
Y_{MIS}	bias	0.04	0.12	-0.11	0.20	0.16	-0.11	0.11	-0.07
	std	0.17	0.15	0.15	0.20	0.16	0.16	0.13	0.21
	sfd	0.17	0.15	0.14	0.20	0.16	0.14	0.13	0.22
	cvp	0.95	0.92	0.92	0.92	0.90	0.91	0.90	0.93

where $\hat{Y}_{ij}$ is the fitted value of Y_{ij} through regressing Y_{ij} on $b^T Y_{ij}$ linearly ($i = 1, \dots, n$; $j = 1, \dots, J_i$).

In terms of estimation bias, our proposed approaches using (2) perform better than the partial dimension reduction method of Li & Yin (2009) and the quadratic inference function method of Bi & Qu (2015), as can be seen from Tables 1–4. For example, in Table 1, when the true covariance structure admits a tridiagonal structure in the time-varying scenario, the bias of Li & Yin's method in estimating β_{24} is as large as 2.03, and the bias of Bi & Qu's method in estimating β_{26} is as large as -3.30. These methods require distributional assumptions on the covariates, which are not satisfied in our setting.

The six methods that use estimating equations, corresponding to (iii)–(viii) above, yield estimators with small biases even when the covariance structure is misspecified. Indeed, we use methods (iii)–(vi) to demonstrate how the working covariance matrices affect estimation efficiency, as measured by the Monte Carlo standard deviation. The estimating equation approach achieves its best performance when the working covariance matrix is the true or the consistently estimated covariance structure, which agrees with Proposition 1. For example, when

Table 3. *Simulation results based on 1000 repetitions in the case where all the covariates are time-varying and the underlying true covariance admits an independent structure*

		β_{13}	β_{14}	β_{15}	β_{16}	β_{23}	β_{24}	β_{25}	β_{26}
PDR	bias	0.49	-0.70	0.23	0.64	1.48	-1.89	0.66	3.19
	std	0.86	0.84	0.72	0.51	2.02	1.51	1.60	1.05
QIF	bias	0.06	-0.01	0.04	-1.55	0.10	-0.05	0.12	-3.52
	std	0.22	0.21	0.19	0.60	0.53	0.46	0.44	1.35
TRI	bias	0.01	0.02	-0.02	0.07	0.04	-0.03	0.03	-0.07
	std	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.16
	sfd	0.12	0.11	0.10	0.16	0.11	0.10	0.09	0.16
	cvp	0.95	0.95	0.94	0.95	0.94	0.94	0.94	0.95
AR(1)	bias	0.01	0.02	-0.02	0.07	0.04	-0.03	0.03	-0.07
	std	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.16
	sfd	0.12	0.11	0.10	0.16	0.11	0.10	0.09	0.16
	cvp	0.95	0.95	0.94	0.95	0.94	0.94	0.94	0.95
IND	bias	0.01	0.02	-0.02	0.07	0.04	-0.03	0.03	-0.07
	std	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.15
	sfd	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.16
	cvp	0.95	0.96	0.94	0.95	0.94	0.94	0.94	0.95
UN	bias	0.01	0.02	-0.02	0.09	0.04	-0.03	0.03	-0.11
	std	0.14	0.12	0.11	0.18	0.12	0.11	0.10	0.19
	sfd	0.13	0.11	0.11	0.17	0.12	0.10	0.10	0.18
	cvp	0.95	0.94	0.94	0.95	0.94	0.94	0.94	0.94
X_{MIS}	bias	-0.00	0.01	-0.00	-0.00	-0.00	-0.00	0.00	0.17
	std	0.12	0.11	0.11	0.15	0.11	0.10	0.09	0.15
	sfd	0.12	0.11	0.10	0.15	0.11	0.10	0.09	0.17
	cvp	0.95	0.94	0.94	0.94	0.96	0.94	0.94	0.88
Y_{MIS}	bias	0.06	0.12	-0.12	0.21	0.17	-0.11	0.12	-0.08
	std	0.17	0.16	0.16	0.20	0.16	0.15	0.14	0.22
	sfd	0.19	0.16	0.15	0.21	0.17	0.16	0.14	0.23
	cvp	0.96	0.93	0.92	0.93	0.91	0.93	0.92	0.95

estimating β_{13} with the tridiagonal covariance structure under the time-varying scenario, the Monte Carlo standard deviations of the estimators of our proposed methods are 0.12, 0.14 and 0.13 when the working covariance has, respectively, AR(1), independent and unstructured structures, none of which is as stable as the method with the correctly specified covariance structure.

We use the estimators (vii) and (viii) to demonstrate the double robustness property of the estimating equation approach. When either the centralization for X_{ij} or that for Y_{ij} is misspecified, the estimators obtained from the estimating equation approach remain consistent as long as one of the centralization quantities is consistently estimated.

The estimating equation approach also offers the possibility of making inferences about the free parameters of model (1), because we introduce a one-to-one parameterization of the central mean subspace. This makes our approach more appealing than the methods of Li & Yin (2009), Jiang & Wang (2011) and Bi & Qu (2015). The estimated standard deviations are very close to the Monte Carlo standard deviations, and the empirical coverage probabilities are very close to 95% in most scenarios, indicating that inferences based on our approaches are fairly reliable.

Table 4. *Simulation results based on 1000 repetitions in the case where some covariates are time-invariant and the underlying true covariance admits a tridiagonal structure*

		β_{13}	β_{14}	β_{15}	β_{16}	β_{23}	β_{24}	β_{25}	β_{26}
PDR	bias	0.14	-0.08	-0.19	0.45	-1.11	1.28	-0.46	1.83
	std	0.69	0.73	0.62	0.61	1.06	1.66	1.28	1.33
QIF	bias	0.05	0.00	0.02	-1.05	0.20	-0.06	0.08	-2.64
	std	0.21	0.17	0.20	0.47	0.48	0.39	0.45	1.06
TRI	bias	0.01	0.02	-0.02	0.05	0.03	-0.02	0.02	-0.04
	std	0.11	0.09	0.09	0.15	0.09	0.08	0.07	0.13
	sfd	0.11	0.09	0.10	0.14	0.09	0.07	0.07	0.13
	cvp	0.94	0.95	0.96	0.94	0.94	0.94	0.95	0.95
AR(1)	bias	0.01	0.02	-0.02	0.05	0.03	-0.02	0.02	-0.05
	std	0.12	0.09	0.10	0.15	0.10	0.08	0.07	0.13
	sfd	0.11	0.09	0.10	0.14	0.09	0.08	0.07	0.13
	cvp	0.94	0.95	0.97	0.94	0.93	0.94	0.95	0.96
IND	bias	0.02	0.02	-0.02	0.07	0.04	-0.03	0.02	-0.07
	std	0.14	0.11	0.10	0.17	0.11	0.09	0.08	0.15
	sfd	0.13	0.10	0.09	0.16	0.11	0.08	0.08	0.15
	cvp	0.94	0.95	0.93	0.94	0.94	0.94	0.94	0.96
UN	bias	0.02	0.03	-0.03	0.08	0.03	-0.02	0.03	-0.07
	std	0.13	0.10	0.11	0.17	0.11	0.09	0.08	0.16
	sfd	0.11	0.09	0.10	0.14	0.09	0.08	0.07	0.14
	cvp	0.90	0.91	0.93	0.91	0.91	0.91	0.90	0.92
X_{MIS}	bias	0.01	0.00	-0.01	0.01	0.00	-0.00	0.00	0.09
	std	0.12	0.08	0.09	0.12	0.10	0.08	0.07	0.15
	sfd	0.11	0.08	0.11	0.13	0.09	0.07	0.07	0.15
	cvp	0.94	0.94	0.97	0.95	0.94	0.95	0.97	0.92
Y_{MIS}	bias	0.06	0.12	-0.14	0.25	0.16	-0.11	0.10	-0.11
	std	0.18	0.15	0.15	0.23	0.17	0.14	0.12	0.22
	sfd	0.19	0.15	0.18	0.22	0.16	0.13	0.12	0.21
	cvp	0.93	0.91	0.94	0.92	0.88	0.90	0.93	0.94

Finally, we evaluate the accuracy of (3) in estimating the structural dimension of $\mathcal{S}_{E(Y|X)}$. From an asymptotic perspective, the consistency of the Bayesian-type information criterion holds for a wide range of C_n . From a finite-sample perspective, we tried several different choices of C_n and obtained similar results. Table 5 reports results for C_n as defined in (4). The frequencies of the estimated structural dimensions are very similar under different working covariance structures, and our proposed Bayesian-type information criterion can estimate the structural dimension quite well. Table 5 also shows the frequencies of estimated structural dimensions obtained from the maximal eigenvalue ratio criterion suggested by Bi & Qu (2015). Our simulation results indicate that the Bayesian-type information criterion is superior to the maximal eigenvalue ratio criterion.

5. AN APPLICATION

We demonstrate the usefulness of our approach through an analysis of the Framingham Heart Study data (Larson et al., 2007). This study included 373 subjects. For each subject, several clinical examinations were performed every two years over 14 years. The response variable

Table 5. Frequency (%) of the estimated dimension $\hat{d}$. In the time-varying scenario, all covariates are time-varying; in the time-invariant scenario, some covariates are time-invariant

True	Working	Time-varying			Time-invariant		
		$\hat{d} = 1$	$\hat{d} = 2$	$\hat{d} \geq 3$	$\hat{d} = 1$	$\hat{d} = 2$	$\hat{d} \geq 3$
TRI	MERC	9	77	13	10	74	16
	TRI	2	92	6	2	90	8
	AR(1)	5	87	8	0	90	10
	IND	2	94	4	1	93	6
	UN	9	84	7	1	92	7
AR(1)	MERC	22	74	4	23	74	3
	TRI	1	90	9	0	95	5
	AR(1)	1	93	6	0	93	7
	IND	3	89	8	0	92	8
	UN	2	87	11	0	97	3
IND	MERC	11	89	0	8	92	0
	TRI	0	97	3	2	92	6
	AR(1)	4	90	6	3	85	12
	IND	3	89	8	1	91	8
	UN	4	92	4	1	91	8

MERC, maximal eigenvalue ratio criterion suggested by [Bi & Qu \(2015\)](#); TRI, tridiagonal working covariance structure; AR(1), AR(1) working covariance structure; IND, independent working covariance structure; UN, unstructured working covariance structure.

was systolic blood pressure, and the four risk factors were body mass index, age of individual when the data were collected, serum cholesterol concentration, and haemoglobin concentration. In total, 1919 complete observations were included in our analysis. The aim is to understand how several risk factors affect hypertension. Because different within-subject correlation matrices yield similar conclusions, in what follows we report only the results obtained from assuming an unstructured working covariance matrix $\hat{V}_i$. The Bayesian-type information criterion yields $\hat{d} = 1$. In other words, the systolic blood pressure depends on these risk factors through a single linear combination. There are many reports in the literature showing that body mass index has a strong association with high blood pressure ([Macmohan et al., 1987](#); [Brown et al., 2000](#)). This allows us to fix the coefficient of body mass index at 1 to ensure the identifiability of model (2) when $\hat{d} = 1$. The estimated coefficients of age, serum cholesterol concentration and haemoglobin concentration are 0.37, 0.15 and 0.18, respectively, with standard deviations of 0.03, 0.04 and 0.16. Both age and serum cholesterol concentration are important risk factors for systolic blood pressure, but we failed to observe any strong association between haemoglobin concentration and hypertension. This confirms the findings of other studies ([Selby et al., 1990](#); [Vasan et al., 2002](#)). Hypertension is age-related and is typically associated with dyslipidemia, and there may be underlying factors that influence both blood pressure and serum cholesterol simultaneously ([Selby et al., 1990](#)).

6. DISCUSSION

We suggest estimating the correlation structure by the usual moment estimators ([Liang & Zeger, 1986](#)). There are several alternative ways to improve the accuracy of estimating the within-subject covariance matrix, including quadratic inference functions ([Qu et al., 2000](#)), transformation of time scales ([Pan & Mackenzie, 2003](#)) and Cholesky decomposition ([Pourahmadi &](#)

Daniels, 2002; Pourahmadi, 2011). Westgate (2013) proposed a bias correlation approach to estimating an unstructured correlation matrix. These procedures can readily be used in our context.

For simplicity, we have assumed that the covariates are measured at the same time and there are no missing data. Covariance estimation with complete data is straightforward, as we described in § 2. When the data are subject to missingness, however, estimating the covariance matrix is non-trivial. When the data are multivariate normal and are missing at random or missing completely at random (Rubin, 1976), estimation and inference can be accomplished by maximum likelihood facilitated by the EM algorithm (Rovine, 1994; Arbuckle et al., 1996). For nonnormal data, Yuan & Bentler (2000) proposed several likelihood-based methods for covariance estimation when the data are missing completely at random. When the data are missing at random and X_{ij} is continuous over time, we may borrow information across different time-points to impute the missing values.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online contains technical proofs of the theorems and proposition as well as additional simulation results.

ACKNOWLEDGEMENT

Wensheng Guo was supported by the U.S. National Institutes of Health. Liping Zhu was supported by the National Natural Science Foundation of China, the Henry Fok Education Foundation Fund of Young College Teachers, Programs for Innovative Research Team in University of China and the Key Laboratory of Mathematical Economics, Ministry of Education of China. Li Jin was supported by the National Science Foundation of China, the Ministry of Science and Technology and the 111 Project. The authors thank the editor, associate editor and two referees for constructive suggestions and comments which have led to significant improvements of the manuscript. We thank Annie Qu for kindly sharing with us the codes for the quadratic inference function method proposed in Bi & Qu (2015).

REFERENCES

- ARBUCKLE, J. L., MARCOULIDES, G. A. & SCHUMACKER, R. E. (1996). Full information estimation in the presence of incomplete data. In *Advanced Structural Equation Modeling: Issues and Techniques*, A. George & E. S. Randall, eds. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 243–77.
- BAI, Y., FUNG, W. K. & ZHU, Z. Y. (2009). Penalized quadratic inference functions for single-index models with longitudinal data. *J. Mult. Anal.* **100**, 152–61.
- BI, X. & QU, A. (2015). Sufficient dimension reduction for longitudinal data. *Statist. Sinica* **25**, 787–807.
- BICKEL, P. J., KLAASSEN, C. A., BICKEL, P. J., RITOV, Y., KLAASSEN, J., WELLNER, J. A. & RITOV, Y. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.
- BROWN, C. D., HIGGINS, M., DONATO, K. A., ROHDE, F. C., GARRISON, R., OBARZANEK, E., ERNST, N. D. & HORAN, M. (2000). Body mass index and the prevalence of hypertension and dyslipidemia. *Obesity Res.* **8**, 605–19.
- COOK, R. D. (1998a). Principal Hessian directions revisited. *J. Am. Statist. Assoc.* **93**, 84–94.
- COOK, R. D. (1998b). *Regression Graphics: Ideas for Studying Regressions through Graphics*. New York: Wiley.
- COOK, R. D. & LI, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30**, 455–74.
- COOK, R. D. & LI, B. (2004). Determining the dimension of iterative Hessian transformation. *Ann. Statist.* **32**, 2501–31.
- COOK, R. D. & WEISBERG, S. (1991). Comment on “Sliced inverse regression for dimension reduction”. *J. Am. Statist. Assoc.* **86**, 328–32.
- COOK, R. D. & YIN, X. (2001). Dimension reduction and visualization in discriminant analysis (with Discussion). *Aust. New Zeal. J. Statist.* **43**, 147–99.
- FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications*. Boca Raton, Florida: Chapman & Hall/CRC.
- HÄRDLE, W. & STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Am. Statist. Assoc.* **84**, 986–95.

- HEDEKER, D. & GIBBONS, R. D. (2006). *Longitudinal Data Analysis*. New York: Wiley.
- HORTON, N. J. & LIPSITZ, S. R. (1999). Review of software to fit generalized estimating equation regression models. *Am. Statistician* **53**, 160–9.
- JIANG, C.-R. & WANG, J.-L. (2011). Functional single index models for longitudinal data. *Ann. Statist.* **39**, 362–88.
- KANG, J. D. & SCHAFER, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statist. Sci.* **22**, 523–39.
- LARSON, M. G., ATWOOD, L. D., BENJAMIN, E. J., CUPPLES, L. A., D'AGOSTINO, R. B. SR., FOX, C. S., GOVINDARAJU, D. R., GUO, C.-Y., HEARD-COSTA, N. L., HWANG, S.-J., ET AL. (2007). Framingham Heart Study 100K project: Genome-wide associations for cardiovascular disease outcomes. *BMC Med. Genet.* **8**, S5–9.
- LEE, K.-Y., LI, B. & CHIAROMONTE, F. (2013). A general theory for nonlinear sufficient dimension reduction: Formulation and estimation. *Ann. Statist.* **41**, 221–49.
- LI, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Am. Statist. Assoc.* **86**, 316–27.
- LI, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. *J. Am. Statist. Assoc.* **87**, 1025–39.
- LI, K.-C. & DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17**, 1009–52.
- LI, L. & YIN, X. (2009). Longitudinal data analysis using sufficient dimension reduction method. *Comp. Statist. Data Anal.* **53**, 4106–15.
- LIANG, K.-Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- MA, Y. & ZHU, L. (2012). A semiparametric approach to dimension reduction. *J. Am. Statist. Assoc.* **107**, 168–79.
- MA, Y. & ZHU, L. (2013). A review on dimension reduction. *Int. Statist. Rev.* **81**, 134–50.
- MA, Y. & ZHU, L. (2014). On estimation efficiency of the central mean subspace. *J. R. Statist. Soc. B* **76**, 885–901.
- MACMOHAN, S., CUTLER, J., BRITAIN, E. & HIGGINS, M. (1987). Obesity and hypertension: Epidemiological and clinical issues. *Eur. Heart J.* **8**, 57–70.
- MCCULLAGH, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59–67.
- PAN, J. & MACKENZIE, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika* **90**, 239–44.
- POURAHMADI, M. (2011). Covariance estimation: The GLM and regularization perspectives. *Statist. Sci.* **26**, 369–87.
- POURAHMADI, M. & DANIELS, M. (2002). Dynamic conditionally linear mixed models for longitudinal data. *Biometrics* **58**, 225–31.
- QU, A., LINDSAY, B. G. & LI, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–36.
- ROVINE, M. J. (1994). Latent variables models and missing data analysis. In *Latent Variables Analysis: Applications for Developmental Research*, A. von Eye & C. C. Clogg, eds. Thousand Oaks: Sage Publications, pp. 181–225.
- RUBIN, B. D. (1976). Inference and missing data. *Biometrika* **63**, 581–92.
- SELBY, J. V., FRIEDMAN, G. D. & QUESENBERRY, C. P. (1990). Precursors of essential hypertension: Pulmonary function, heart rate, uric acid, serum cholesterol, and other serum chemistries. *Am. J. Epidemiol.* **131**, 1017–27.
- VASAN, R. S., BEISER, A., SESHADRI, S., LARSON, M. G., KANNEL, W. B., D'AGOSTINO, R. B. & LEVY, D. (2002). Residual lifetime risk for developing hypertension in middle-aged women and men. *J. Am. Med. Assoc.* **287**, 1003–10.
- WEDDERBURN, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. *Biometrika* **61**, 439–47.
- WESTGATE, P. M. (2013). A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. *Statist. Med.* **32**, 2850–8.
- XIA, Y., TONG, H., LI, W. & ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Statist. Soc. B* **64**, 363–410.
- YUAN, K.-H. & BENTLER, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociol. Methodol.* **30**, 165–200.
- ZHU, L., MIAO, B. & PENG, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Am. Statist. Assoc.* **101**, 630–43.

[Received July 2014. Revised November 2015]