Joint estimation of mean-covariance model for longitudinal data with basis function approximations

Jie Mao\textsuperscript{a}, Zhongyi Zhu\textsuperscript{a,\*}, Wing K. Fung\textsuperscript{b}

\textsuperscript{a} Department of Statistics, School of Management, Fudan University, Shanghai, China
\textsuperscript{b} Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong, China

\textbf{ARTICLE INFO}

\textbf{Article history:}  
Received 30 December 2009  
Received in revised form 15 July 2010  
Accepted 9 August 2010  
Available online 18 August 2010

\textbf{Keywords:}  
Basis function  
BIC  
B-splines  
Modified Cholesky decomposition  
Partially linear model

\textbf{ABSTRACT}

When the selected parametric model for the covariance structure is far from the true one, the corresponding covariance estimator could have considerable bias. To balance the variability and bias of the covariance estimator, we employ a nonparametric method. In addition, as different mean structures may lead to different estimators of the covariance matrix, we choose a semiparametric model for the mean so as to provide a stable estimate of the covariance matrix. Based on the modified Cholesky decomposition of the covariance matrix, we construct the joint mean-covariance model by modeling the smooth functions using the spline method and estimate the associated parameters using the maximum likelihood approach. A simulation study and a real data analysis are conducted to illustrate the proposed approach and demonstrate the flexibility of the suggested model.

\textcopyright 2010 Elsevier B.V. All rights reserved.

1. Introduction

The estimation of the covariance matrix is important in a longitudinal study. A good estimator for the covariance can improve the efficiency of the regression coefficients. Furthermore, the covariance estimation itself is also of interest (Diggle and Verbyla, 1998). A number of authors have studied the problem of estimating the covariance matrix. Pourahmadi (1999, 2000) considered generalized linear models for the components of the modified Cholesky decomposition of the covariance matrix. Fan et al. (2007) and Fan and Wu (2008) proposed to use a semiparametric model for the covariance function. However, the mean and covariance estimators could have considerable bias when the specified parametric or semiparametric model for the covariance structure is far from the truth (Huang et al., 2007).

To balance the variability and bias of the covariance estimator, nonparametric estimators of the covariance structures are being proposed. There are several nonparametric methods used in estimating the covariance matrix. Diggle and Verbyla (1998) provided a nonparametric estimator for the covariance structure without assuming stationarity, but their estimator is not guaranteed to be positive definite. To overcome the positive-definiteness constraint, Wu and Pourahmadi (2003) proposed a nonparametric smoothing to regularize the estimation of a large covariance matrix based on the modified Cholesky decomposition method, but their first step raw estimate is too noisy and thus an inefficient estimate may result. Huang et al. (2007) proposed to apply a smoothing-based regularization after using the modified Cholesky decomposition of the covariance matrix and found that their estimation could be more efficient than Wu and Pourahmadi’s. However, they only considered balanced data which is not common in practice. Thus we present an extension of their method to unbalanced data. In addition, all these works focus on the estimation of the covariance matrix and pay little attention to the mean structure. As shown in Pan and Mackenzie (2003), a misspecified estimator of the mean structure may well lead to
2. Joint mean-covariance model

Assume that we have a sample of $n$ subjects. For the $i$th subject, $i = 1, \ldots, n$, the response variable $y_i(t_{ij})$ and the covariate vector $x_i(t_{ij})$ are collected at time points $t = t_{ij}, j = 1, \ldots, n_i$, where $n_i$ is the total number of observations for the $i$th subject. The following partially linear model is considered,

$$y_{ij} = x_{ij}'\beta + \alpha(t_{ij}) + \epsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, n,$$  \hspace{1cm} (2.1)

where $\beta$ is a $p$-dimensional unknown parameter vector, $\alpha(t_{ij})$ is an unknown smooth function, $E(\epsilon_{ij}) = 0$. Let $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{im})'$ and $\text{Cov}(\epsilon_{ij}) = \Sigma$, Donate $\mu_{ij} = x_{ij}'\beta + \alpha(t_{ij})$.

Model (2.1) retains the flexibility of a nonparametric model for the baseline function and maintains the explanatory power of a parametric model. Therefore, there is a rich literature on model (2.1) and its variations (Fan and Li, 2004).

In what follows, we first approximate the nonparametric term $\alpha(.)$ by a linear combination of $B$-spline basis functions, and then model the within-subject correlation and variation using the spline smoothing method, and finally construct the joint mean-covariance model.

2.1. Smoothing nonparametric term

Following He et al. (2002, 2005) and Zhu et al. (2008), we approximate $\alpha(.)$ by the $B$-spline method

$$\alpha(t_{ij}) \approx \sum_{k=1}^{J_i} B_{jk}(t_{ij})\eta_k, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m,$$  \hspace{1cm} (2.2)

where $B_{jk}(t_{ij}) = (B_{j1}(t_{ij}), \ldots, B_{jk}(t_{ij}))'$ is the $B$-spline basis function, $\eta = (\eta_1, \ldots, \eta_m)'$ is an unknown spline coefficient vector to be estimated and $J_i$ is the number of spline coefficients, the selection of which will be discussed later in Section 3.2. Using (2.1) and (2.2), we have

$$y_{ij} \approx x_{ij}'\beta + B_j(t_{ij})\eta + \epsilon_{ij}.$$  \hspace{1cm} (2.3)

Let $\tilde{y}_{ij} = (x_{ij}', B_j(t_{ij}))'$ and $\theta = (\beta', \eta')'$, then (2.3) becomes

$$y_{ij} \approx \tilde{y}_{ij}'\theta + \epsilon_{ij}.$$  

2.2. Smoothing covariance matrix term

To alleviate the positive-definiteness constraint, we adopt the approach based on the modified Cholesky decomposition. The key idea is that a symmetric matrix $\Sigma$ is positive definite if and only if there exists a unique unit lower triangular matrix $T_i$, with 1’s as diagonal entries, and a unique diagonal matrix $D_i$ with positive definite entries such that

$$T_i\Sigma T_i' = D_i,$$

where $T_i$ and $D_i$ are easy to compute and interpret statistically: the below-diagonal entries of $T_i$ are the negatives of the coefficients of $\tilde{y}_{ij} = \mu_{ij} + \sum_{k=1}^{J_i}\phi_{ijk}(y_k - \mu_{jk})$, the linear least-squares predictor of $y_{ij}$ based on its predecessors $y_{i1}, \ldots, y_{i(i-1)}$, and the diagonal entries of $D_i$ are the prediction error variances $\sigma^2_{\epsilon_{ij}} = \text{Var}(\epsilon_{ij})$, where $\epsilon_{ij} = y_{ij} - \tilde{y}_{ij}$, for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Throughout this paper we refer to $\phi_{ijk}$ as generalized autoregressive parameters and to $\sigma^2_{\epsilon_{ij}}$ as innovation variances (Pourahmadi, 1999).
Since $\phi_{ik}$ and $\log(\sigma^2_{ij})$ are unconstrained, we can model them by spline regression. More specifically, as in smoothing $\alpha(.)$, we approximate $\log(\sigma^2_{ij})$ by spline functions

$$\log(\sigma^2_{ij}) \approx \sum_{s=1}^{J} B_{0s}(t_{ij}) \lambda_s = B'_0(t_{ij}) \lambda, \quad j = 1, \ldots, n_i, \ i = 1, \ldots, n,$$  

where $B_0(t_{ij}) = (B_{0s1}(t_{ij}), \ldots, B_{0sk}(t_{ij}))'$ is the vector of $B$-spline basis functions, $\lambda = (\lambda_1, \ldots, \lambda_J)'$ is the unknown spline coefficient vector to be estimated and $f_0$ is the number of $B$-spline basis functions.

On the other hand, similar to Huang et al. (2007), we smooth $T_j$ along its subdiagonals. But we here assume that the within-subject correlation only depends on the elapsed time. The $k$th $(k = 1, \ldots, n_i - 1)$ subdiagonal of all $T_j$ are modeled as realizations of some smooth function $f_k$.

$$\phi_{ijk} = f_k(|t_{ij} - t_{ik}|), \quad j = 1, \ldots, n_i, \ i = 1, \ldots, n.$$  

Each of these smooth functions can be approximated as a spline function

$$f_k(|t_{ij} - t_{ik}|) \approx \sum_{s=1}^{J} B_{sp}(|t_{ij} - t_{ik}|) \gamma_{sk} = B'_k(|t_{ij} - t_{ik}|) \gamma_k,$$

where $B_k(u) = (B_{k1}(u), \ldots, B_{kJ}(u))'$ is the vector of $B$-spline basis functions, $\gamma_k = (\gamma_{k1}, \ldots, \gamma_{kJ})'$ is the unknown spline coefficient vector to be estimated and $f_k$ is the number of $B$-spline basis functions. Note that here we map the $\sum_{s=1}^{n_i} (n_i - k)$ numbers on the $k$th subdiagonal of all $T_j$ to the function $f_k$ evaluated at $\sum_{s=1}^{n_i} (n_i - k)$ points. Similar to Huang et al. (2007), we smooth only the first $m_0$ subdiagonal of all $T_j$ for some nonnegative integer $m_0$ and set all the elements of the last $n_i - 1 - m_0$ subdiagonal of $T_j$ to be 0. Note that if there exists one $n_i$ satisfying $n_i - 1 < m_0$, then we model the $\min(n_i, n_i - 1)$ subdiagonal of $T_j$ as realizations of some smooth function $f_{\min(n_i)}$. This is the similar case for several numbers $n_i$ satisfying $n_i - 1 < m_0$. To simplify the implementation, we take $J = \cdots = J_{m_0} = J$ and $B_1 = \cdots = B_{m_0} = B$. Then it follows that

$$\phi_{ijk} \approx \gamma_{j-k} B(|t_{ij} - t_{ik}|), \quad j = 2, \ldots, n_i, \ i = 1, \ldots, n,$$

$$\max(1, j - m_0) \leq k \leq j - 1; \quad \phi_{ijk} = 0, \quad j = 2, \ldots, n_i, \ k < j - m_0.$$  

Such simplification performs well in the simulation study and greatly reduces the computational cost. It is worth mentioning that model (2.5) is more general than that in Huang et al. ’s model which is confined to the analysis of balanced longitudinal data. Besides that, generalized autoregressive parameters are supposed to only depend on the elapsed time.

2.3. Joint mean-covariance model

Following Pourahmadi (1999, 2000), we construct the following joint mean-covariance model for modeling the mean, innovation variance and generalized autoregressive parameters:

$$\bar{\mu}_j = \bar{x}_j \theta, \quad \log(\sigma^2_{ij}) = \bar{z}_j' \lambda, \quad \phi_{ijk} = \bar{z}_j' \gamma'$$

where $\bar{x}_j = (\bar{x}_{j1}, \ldots, \bar{x}_{jn})'$, $\bar{z}_j = B_0(t_{ij})$ for $j = 1, \ldots, n_i$ and

$$\bar{z}_{ij} = \begin{pmatrix} 0, \ldots, 0, B_1(|t_{ij} - t_{ik}|), \ldots, B_j(|t_{ij} - t_{ik}|), \ldots, 0, \ldots, 0 \end{pmatrix}_{(i-j+k)},$$

for $k = 1, \ldots, j - 1, j = 2, \ldots, n_i$, and $\theta = (\beta', \eta)'$, $\lambda$ and $\gamma = (\gamma'_{1}, \ldots, \gamma'_{m_0}, \ldots, \gamma'_{n_i - 1})'$ are parameters for the means, innovation variances and generalized autoregressive parameters, respectively, where $\gamma_s = 0$ for all $s > m_0$.

3. Estimation procedure

3.1. Estimation of $\theta$, $\gamma$ and $\lambda$

In the above section, we have modeled the nonparametric term, and the main diagonal of $D_i$ and the subdiagonals of $T_j$ by spline functions and then we will employ the likelihood approach as given in the following.

Let $y_i = (y_{i1}, \ldots, y_{in})'$, $X_i = (x_{i1}, \ldots, x_{in})'$, $\mu_i = (\mu_{i1}, \ldots, \mu_{in})$ and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{in})'$. Then, the logarithm of the likelihood function of $y_1, \ldots, y_n$ can be written as

$$L = -\frac{1}{2} \sum_{i=1}^{n} \log|\Sigma_i| - \frac{1}{2} \sum_{i=1}^{n} \text{tr}(\Sigma^{-1}_i S_i)$$

(3.1)
up to a constant that can be neglected, where $S_i = (y_i - \mu_i)(y_i - \mu_i)'$. With $\mu_i$ being replaced by $\tilde{\mu}_i = (\tilde{\mu}_i1, \ldots, \tilde{\mu}_im)$, we have

$$L = -\frac{1}{2} \sum_{i=1}^{n} \log |\Sigma_i| - \frac{1}{2} \sum_{i=1}^{n} (y_i - \tilde{\mu}_i)' \Sigma_i^{-1} (y_i - \tilde{\mu}_i).$$  \hfill (3.2)

Similar to Pourahmadi (2000), the function has three representations corresponding to the three submodels in (2.6)

$$-2L(\theta, \lambda, \gamma) = \sum_{i=1}^{n} \log |\Sigma_i| + \sum_{i=1}^{n} (y_i - \tilde{X}_i\theta)' \Sigma_i^{-1} (y_i - \tilde{X}_i\theta)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \log(\sigma^2_{ij}) + \frac{g_{ij}}{\sigma^2_{ij}} \right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \log(\sigma^2_{ij}) + \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{\sigma^2_{ij}} \sum_{k=1}^{m} \sum_{p=1}^{m} \phi_{ijk} \phi_{ijp} s_{ijkp},$$

where $\tilde{X}_i = (\tilde{x}_{i1}, \ldots, \tilde{x}_{im})$, $g_{ij}$ is the $(j, j)$th element of $G_i = T_i \tilde{S}_i T_i'$ with $\tilde{S}_i = (y_i - \tilde{X}_i\theta)(y_i - \tilde{X}_i\theta)'$ and $s_{ijk}$ is the $(k, p)$th element of $\tilde{S}_i$. We have the following updating procedures.

Updating $\theta$. Given $\Sigma$, the estimator of $\theta$ can be updated by

$$\tilde{\theta} = \left\{ \left( \sum_{i=1}^{n} \tilde{X}_i \Sigma_i^{-1} \tilde{X}_i \right)^{-1} \sum_{i=1}^{n} \tilde{X}_i \Sigma_i^{-1} y_i \right\}^{-1} \sum_{i=1}^{n} \tilde{X}_i \Sigma_i^{-1} y_i.$$  \hfill (3.3)

Updating $\lambda$. Since function $L(\theta, \lambda, \gamma)$ is nonlinear in $\lambda$, an iterative method for computing the estimator of $\lambda$ is needed. Given $\theta$, $\gamma$ and the current values of $\lambda$, the estimate of $\lambda$ can be updated using: $\lambda \leftarrow \lambda - H_\lambda^{-1} U_\lambda$, where

$$U_\lambda = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} [B_0(t_{ij}) - \exp(-\lambda'[B_0(t_{ij})]B_0(t_{ij}) g_{ij}']$$

and

$$H_\lambda = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} [\exp(-\lambda'[B_0(t_{ij})]B_0(t_{ij}) B_0(t_{ij}) g_{ij}]).$$

Updating $\gamma$. Given $\theta$ and $\lambda$, and fixing the rest of the parameters, we have

$$\gamma = A^{-1} b,$$

where

$$A = \sum_{i=1}^{n} \sum_{j=1}^{n} \exp(-\lambda'[B_0(t_{ij})]) \sum_{k=1}^{j-1} \sum_{l=1}^{j-1} s_{ikl} (z_{i}^{ik} z_{i}^{jk} + z_{i}^{ik} z_{i}^{jk})$$

and

$$b = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \exp(-\lambda'[B_0(t_{ij})]) \sum_{k=1}^{j-1} s_{ikl}.$$

A convenient initial value for $\lambda$ and $\gamma$ is $\lambda^{(0)} = 0$ and $\gamma^{(0)} = 0$. In other words, $l_{n \times n}$ is chosen as the starting value for the covariance matrix $\Sigma_i$. This algorithm is more general than that given in Huang et al. (2007). Their algorithm is confined to the analysis of balanced longitudinal data with equal time intervals. So they assumed that generalized autoregressive parameters are modeled as the realizations of some smooth function evaluated at equally spaced grid points on the $[0, 1]$ interval. Here generalized autoregressive parameters are supposed to only depend on the elapsed time.

Similar to that in He et al. (2005), the asymptotic covariance matrix of $\tilde{\beta}$ can be estimated by

$$\text{Cov}(\tilde{\beta}) = (A_n)^{-1} K_n(A_n)^{-1} \hfill (3.4)$$

where $A_n$ and $K_n$ are defined by

$$A_n = \sum_{i=1}^{n} \Sigma_i^{-1} X_i,$$

and

$$K_n = \sum_{i=1}^{n} \Sigma_i^{-1} (y_i - \tilde{\mu}_i)(y_i - \tilde{\mu}_i)' \Sigma_i^{-1} X_i,$$

$$M = (B_0'(t_{11}), \ldots, B_0'(t_{1m}), \ldots, B_0'(t_{mn})), X^* = (I - P) X, P = M(M' \Sigma M)^{-1} M' \Sigma$$ and $\Sigma = \text{diag}(\Sigma_1, \ldots, \Sigma_n)$. 

Author's personal copy
3.2. Model selection

A question which has not been addressed in the implementation of the foregoing procedure is the choice of tuning parameters \((m_0, j_0, J_0, \eta)\). The number of spline coefficients \(j_0\), \(J_0\) and \(J\) are determined by the degree of splines and the number of knots. In the following simulation study and data analysis, we use the sample quantiles of the time points as knots. For the sake of simplicity, we opt for convenient choices of knot placement in this article. Following He et al. (2005), we use the sample quartiles of \(\{t_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i\}\) as knots. For example, if we use three internal knots, then these are taken to be the three quartiles of the observed \(t_{ij}\). We use cubic splines and take the number of internal knots to be the integer part of \(M^{1/3}\), where \(M\) is the number of distinct values in \(\{t_{ij}, i = 1, \ldots, n, j = 1, \ldots, n_i\}\).

In this article, we use cubic splines in simulation study and real data analysis. The tuning parameter \(m_0\) is selected based on the BIC criterion, which is defined as

\[
\text{BIC}(m_0) = -\frac{2}{n} \hat{L}_{\text{max}} + \frac{\log n}{n} (m_0^{\ell}),
\]

where \(\hat{L}_{\text{max}}\) is the maximized \(\hat{L}\) for the model with the specified tuning parameter \(m_0\). So we can get

\[
m_0^{\ell} = \arg \min_{m_0} \{\text{BIC}(m_0)\}.
\]

4. Simulation study

In this section, we investigate the performance of the proposed method by the Monte Carlo simulation. For comparison, we estimate \(\beta\) and \(\alpha(.)\) using a working independence covariance structure (WI) and the true covariance structure (True). We also include the sample covariance estimator in the comparison. Moreover, we demonstrate the flexibility and efficiency of model (2.1) by comparing with the linear model (4.3) and investigate the effect of a misspecification of the mean structure on the covariance estimates.

4.1. Simulation models

The data sets are generated from the following model

\[
y_{ij} = x_{ij} \beta + \alpha(t_{ij}) + e_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, 50,
\]

\[(4.1)\]

\(\beta_1 = 1, \beta_2 = 2, \alpha(t_{ij}) = \sin(2\pi t_{ij}/12),\) the observation times are generated as follows: each individual has a set of scheduled time points \(\{0, 1, \ldots, 12\}\), and each scheduled time, except time 0, has a probability of 20\% being skipped. The actual observation time is a random perturbation of the scheduled time: a uniform \((0, 1)\) random variable is added to non-skipped scheduled time to obtain the actual observation time. The covariates are chosen as follows: \(x_{1,ij}\) is a normal random variable with mean 0, variance 1, and \(x_{2,ij}\) follows the Bernoulli-distributed random variable with success probability 0.5 and independent of \(x_{1,ij}\). In addition, \(e_{ij}\) are generated from a Gaussian process with mean 0 and two following covariance matrices which are similar to that considered in Huang et al. (2007) and Wu and Pourahmadi (2003):

- \(\Sigma_1: \phi_{ij-k} = 0, \sigma_{ij} = 1\), corresponding to the identity covariance matrix.
- \(\Sigma_2: \phi_{ij-j-1} = 0.25((t_{ij} - t_{(ij-1)})^2 - 0.5, \phi_{ij-k} = 0, k \geq 2, \sigma_{ij} = \log(t_{ij}/12 + 2),\) corresponding to varying coefficient AR(1).

Based on the model selection mentioned in Section 3.2.2, we opt for \(J_0 = 7, J_0 = 7\) and \(J = 7\). This particular choice is by no means an optimal choice, but substantially reduces the computational cost. The following simulation results are all based on 1000 independent repetitions. In each simulation, we use BIC criterion to select tuning parameter \(m_0\).

4.2. Performance of estimating \(\beta\) and \(\alpha(.)\)

Table 1 summarizes the results over 1000 repetitions. In this table, ‘Bias’ represents the sample average over 1000 estimates subtracting the true value of \(\beta\) and ‘SD’ and ‘MSE’ represent the sample standard deviation and mean squared error of the estimator respectively. From Table 1, we can see that the spline approach yields estimates for \(\beta\) as good as the estimates obtained using the true covariance structure. Table 1 also indicates that the proposed estimates are more efficient than those obtained using the working independent covariance matrix.

To explore the robustness of the proposed methods to the normality assumption, we generated data from the multivariate \(t\) distribution, with random error coming from \(\varepsilon/\sqrt{Z/2}\), where \(\varepsilon \sim N(0, \Sigma_2), Z \sim \chi^2(4), \varepsilon\) and \(Z\) are independent, and all other components in this model are simulated the same as in the previous case. The simulation result is displayed in the bottom part of Table 1. The proposed method substantially improves over the working independence associated with the performance of \(\beta\) and performs similarly to that using the true covariance.

Table 2 demonstrates that the standard error formula (3.4) works well for two different working correlation structures whether the random error follows normal distribution or not.
Table 1
Performance of $\hat{\beta}$.

<table>
<thead>
<tr>
<th>Covariance Method</th>
<th>$\hat{\beta}_1$ Bias</th>
<th>SD</th>
<th>MSE</th>
<th>$\hat{\beta}_2$ Bias</th>
<th>SD</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>0.0506</td>
<td>4.5207</td>
<td>0.2042</td>
<td>0.1587</td>
<td>8.6742</td>
<td>0.7519</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0477</td>
<td>4.5348</td>
<td>0.2055</td>
<td>0.1549</td>
<td>8.7513</td>
<td>0.7653</td>
</tr>
<tr>
<td>True</td>
<td>0.0083</td>
<td>3.5123</td>
<td>0.1233</td>
<td>0.0452</td>
<td>6.6902</td>
<td>0.4472</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0074</td>
<td>4.6800</td>
<td>0.2188</td>
<td>0.1600</td>
<td>9.0562</td>
<td>0.8196</td>
</tr>
<tr>
<td><strong>t distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>0.0955</td>
<td>4.4534</td>
<td>0.1982</td>
<td>0.1617</td>
<td>8.7606</td>
<td>0.7670</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0790</td>
<td>4.2864</td>
<td>0.1836</td>
<td>0.1296</td>
<td>8.5528</td>
<td>0.7309</td>
</tr>
<tr>
<td>True</td>
<td>0.0175</td>
<td>3.5222</td>
<td>0.1239</td>
<td>0.0968</td>
<td>7.2740</td>
<td>0.5287</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0320</td>
<td>3.5661</td>
<td>0.1271</td>
<td>0.1347</td>
<td>7.2673</td>
<td>0.5278</td>
</tr>
</tbody>
</table>

*Values in the columns of Bias, SD and MSE are multiplied by a factor of 100.

Table 2
Assessment of the standard errors using formula (3.4).

<table>
<thead>
<tr>
<th>Covariance</th>
<th>$\hat{\beta}_1$ SD</th>
<th>SE(Std)</th>
<th>$\hat{\beta}_2$ SD</th>
<th>SE(Std)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>0.0453</td>
<td>0.0422 (0.0050)</td>
<td>0.0875</td>
<td>0.0842 (0.0088)</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0365</td>
<td>0.0331 (0.0044)</td>
<td>0.0700</td>
<td>0.0658 (0.0076)</td>
</tr>
<tr>
<td><strong>t distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True</td>
<td>0.0429</td>
<td>0.0409 (0.0053)</td>
<td>0.0855</td>
<td>0.0819 (0.0094)</td>
</tr>
<tr>
<td>Spline</td>
<td>0.0357</td>
<td>0.0320 (0.0048)</td>
<td>0.0727</td>
<td>0.0632 (0.0081)</td>
</tr>
</tbody>
</table>

We next evaluate the performance of $\hat{\alpha}$. It can be assessed by the square root of the average squared errors (RASE),

$$RASE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} (\hat{\alpha}(t_{ij}) - \alpha(t_{ij}))^2}.$$  

Table 3 displays the RASE for $\alpha$. When the true covariance is $\Sigma_2$, compared with the working independence, the RASE using the true covariance and the proposed covariance structure are smaller. This shows that when using the spline method, we can reduce the RASE by accounting for the covariance structure. This implies that we can improve the performance of $\hat{\alpha}$ after incorporating the covariance matrix when we use the spline method, which is consistent with Zhu et al. (2008).

4.3. Performance of estimating $\Sigma$

To investigate the performance of covariance estimate, we use two loss functions, namely the entropy loss $\Delta_1(\Sigma, G) = n^{-1} \sum_{i=1}^{n} \{ \text{tr}(\Sigma^{-1} G_i - \log |\Sigma| G_i - n\} |$ and the quadratic loss $\Delta_2(\Sigma, G) = n^{-1} \sum_{i=1}^{n} \{ \text{tr}(\Sigma^{-1} G_i - I)^2 \}$, where $\Sigma_i$ is the true covariance matrix and $G_i$ is a positive definite matrix. Each of these losses is 0 when $G_i = \Sigma_i$ and is positive when $G_i \neq \Sigma_i$. The corresponding risk functions are defined by

$$R_i(\Sigma, G) = E_{\Sigma} \{ \Delta_i(\Sigma, G) \}, \quad i = 1, 2.$$
Table 4
Risk comparison in covariance estimates.

<table>
<thead>
<tr>
<th>Covariance</th>
<th>Entropy loss</th>
<th>Quadratic loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample</td>
<td>Spline</td>
</tr>
<tr>
<td>Normal distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1$</td>
<td>Inf</td>
<td>0.0903</td>
</tr>
<tr>
<td>$\Sigma_2$</td>
<td>351.8983</td>
<td>0.4969</td>
</tr>
<tr>
<td>t distribution</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1$</td>
<td>Inf</td>
<td>0.4755</td>
</tr>
<tr>
<td>$\Sigma_2$</td>
<td>355.7442</td>
<td>0.9238</td>
</tr>
</tbody>
</table>

Note: Inf, infinity.

Table 5
Risk comparison in covariance estimates with the model (4.3).

<table>
<thead>
<tr>
<th>Covariance</th>
<th>Entropy loss</th>
<th>Quadratic loss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sample</td>
<td>Spline (2.1)</td>
</tr>
<tr>
<td>Case I: $\alpha(t) = \sin(2\pi t)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1$</td>
<td>Inf</td>
<td>0.8151</td>
</tr>
<tr>
<td>$\Sigma_2$</td>
<td>352.3187</td>
<td>4.3354</td>
</tr>
<tr>
<td>Case II: $\alpha(t) = t$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Sigma_1$</td>
<td>Inf</td>
<td>0.0885</td>
</tr>
<tr>
<td>$\Sigma_2$</td>
<td>352.1086</td>
<td>0.5839</td>
</tr>
</tbody>
</table>

Note: See Table 4 for abbreviations.

An estimator $\hat{\Sigma}_1$ is considered to be better than an estimator $\hat{\Sigma}_2$ if its risk function is smaller, that is, $R_i(\Sigma, \hat{\Sigma}_1) < R_i(\Sigma, \hat{\Sigma}_2)$. The risk function of the proposed estimator is approximated by the Monte Carlo simulation. For more information see Huang et al. (2007).

The results for the proposed spline smoothed estimator for two different covariance matrices are presented in Table 4. In the table, ‘sample’ and ‘spline’ represent respectively the sample covariance matrix and the covariance matrix estimator based on the spline method. It is obvious from Table 4 that the proposed covariance estimator outperforms the sample covariance matrix under both loss functions. To further show the robustness of the proposed covariance estimator, we generated data from the multivariate $t$ distribution with $\Sigma_2$. The results are shown in Table 4 and they are consistent with those for normal data. Moreover, although the sample covariance estimator is unbiased, it is often not invertible even when the dimensionality is no larger than the sample size. This has been proved in Table 4. Due to the singularity of the sample covariance matrix, the entropy loss of it diverges to infinity. Thus the sample covariance matrix is not a good estimator.

4.4. Comparison with linear models

In this section, to demonstrate the flexibility and efficiency of model (2.1), we compare this model with the following linear model,

$$y_{ij} = \beta_0 + \alpha t_{ij} + \epsilon_{ij}, \quad j = 1, \ldots, n_j, \quad i = 1, \ldots, n$$

based on the performance of the corresponding covariance estimators. We estimate $\alpha$ and $\beta$ in model (4.3) using the weighted least squares method. We generated 1000 datasets from model (2.1) as follows:

- Case I: $\beta = (1, 1)'$ and $\alpha(t) = \sin(2\pi t)$.
- Case II: $\beta = (1, 1)'$ and $\alpha(t) = t$, that is, $\alpha = 1$ in model (4.3).

All other components are the same as those specified in Section 4.1.

To illustrate the flexibility of model (2.1), we estimated the covariance matrix using the data generated under the setting of Case I. Simulation results of the estimated risks are reported in the upper part of Table 5, in which the notations are the same as those given in Table 4. In Table 5, spline (2.1) represents the covariance estimator based on the proposed partially linear model (2.1) while spline (4.3) is that based on the linear model (4.3). Table 5 indicates that a misspecification of the component of the mean $\alpha(.)$ can produce a larger risk on the estimation of the covariance component.

Simulation results of the estimated risks for Case II are displayed in the lower parts of Table 5. In this case, the risks of the resulting covariance estimates based on both models (2.1) and (4.3) are of similar magnitudes. Therefore, the proposed estimation procedure with model (2.1) offers a good balance between model flexibility and estimation efficiency for the mean as well as the covariance.
Table 6
Values of BIC.

<table>
<thead>
<tr>
<th>m₀</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIC</td>
<td>−1.8960</td>
<td>−1.9762</td>
<td>−2.0720</td>
<td>−1.5790</td>
<td>−1.3826</td>
</tr>
</tbody>
</table>

5. Application to real data

Here we apply the proposed method to the actual longitudinal data. The data is the longitudinal hormone study on progesterone (Zhang et al., 1998) which collected urine samples from 34 healthy women in a menstrual cycle and urinary progesterone on alternative days. A total of 492 observations were obtained from the 34 participants with each contributing from 11 to 28 observations. The menstrual cycle lengths of these women ranged from 23 to 56 days, with an average of 29.6 days. Biologically, it is meaningful to suppose that the change of the progesterone level for each woman depends on the time during a menstrual cycle relative to her cycle length. Therefore, each woman’s menstrual cycle length was standardized to a 28-day cycle. In addition, a log-transformation was applied to the progesterone level to make the normality assumption more plausible.

For the ith subject, denote x₁₁ to be age and x₁₂ to be body mass index (x₁₁ and x₁₂ are standardized variables with mean 0 and standard deviation 1). Set as the response yᵢⱼ the jth log-transformed progesterone value measured at standardized day tᵢⱼ since menstruation for the ith woman. We consider the following semi-parametric model

\[ y(tᵢⱼ) = \beta₁x₁(tᵢⱼ) + \beta₂x₂(tᵢⱼ) + \alpha(tᵢⱼ) + \varepsilon(tᵢⱼ), \]

where \( \alpha(t) \) is the nonparametric term.

Applying the proposed method, we use cubic splines to model the nonparametric term and the covariance matrix of the data. The BIC values in Table 6 suggest smoothing the first three subdiagonals of T and setting the rest as zeros. The nonparametric mean function and the diagonal of D are smoothed with six B-spline basis functions, and the first three subdiagonals of T are all smoothed with five B-spline basis functions.

From Fig. 1, we can see that intercept term \( \alpha(t) \) decreases in the first 9 days and increases largely later on. They reach a peak around the 23rd day in the cycle, and then decrease again. This result is very similar to that given by Zhang et al. (1998). Fig. 2(a)–(c) display the smoothed first three subdiagonals of T. Fig. 2(a) and (b) seem to fluctuate around a constant and (c) shows that the smoothed third subdiagonal of T remains constant if the time lag is less than 4 days and decreases sharply when the time lag becomes larger. The smoothed diagonal of D is shown in Fig. 1(d). Similar to the smoothed first subdiagonal of T, the innovation curve also fluctuate.

Next we estimate \( \beta \). The estimators of \( \beta \) are shown in the Table 7 with the SD in parentheses. Table 7 indicates that the association between the response and age is positive while the relationship between the response and body mass index
is negative. It also follows from Table 7 that the proposed method gives estimators with smaller standard error. For our approach, the impacts of age have no significant effect on progesterone level while body mass index is a significant variable. The results in Zhang et al. (1998) and He et al. (2002) indicate that both age and body mass index have no significant effect on progesterone level.

6. Conclusion

In this article, we proposed a partially linear model which keeps the flexibility of the nonparametric model while maintaining the explanatory power of the parametric model. We model the nonparametric term, the within-subject correlation and variation by spline functions after decomposing the covariance matrix based on the modified Cholesky decomposition. Here we consider the unbalanced data while Huang et al. (2007) considered balanced data with equal time intervals. In the simulation study, we have shown that the proposed method performs very well, even when the noise term does not follow the normal distribution. We also show that, when using B-splines, we can reduce the RASE of the nonparametric term by incorporating the covariance matrix. Finally, we have demonstrated that the model (2.1) offers a good balance between model flexibility and estimation efficiency.

In this paper, we only smooth a selected number of subdiagonals and set the remaining, shorter diagonals, as zeros. That means the proposed approach only allows the zeros in the Cholesky factor to be regularly placed. Further research on relaxing this condition is of interest.

Acknowledgements

The authors would like to thank the Editor and the referees for their constructive comments and helpful suggestions that largely improve the presentation of the paper. The research is supported by the Natural Science Foundation of China Grant 10931002, 1091120386.

References