Nonparametric Regression Analysis of Multivariate Longitudinal Data

Dongdong Xiang[1], Peihua Qiu[2] and Xiaolong Pu[1]

[1]School of Finance and Statistics, East China Normal University

[2]School of Statistics, University of Minnesota

Abstract

Multivariate longitudinal data are common in medical, industrial and social science research. However, statistical analysis of such data in the current literature is restricted to linear or parametric modeling, which is inappropriate for applications in which the assumed parametric models are invalid. On the other hand, all existing nonparametric methods for analyzing longitudinal data are for univariate cases only. When longitudinal data are multivariate, nonparametric modeling becomes challenging, because we need to properly handle the association among the observed data across different time points and across different components of the multivariate response as well. Motivated by a real data from the National Hearth Lung and Blood Institute, this paper proposes a nonparametric modeling approach for analyzing multivariate longitudinal data. Our method is based on multivariate local polynomial smoothing. Both theoretical and numerical results show that it is useful in various cases.

Keywords: Longitudinal Data, Local Polynomial Regression, Multivariate Regression, Cluster Data.

1 Introduction

Some nonparametric methods have been proposed in the literature for the analysis of longitudinal data. Most of them restrict their attention to the analysis of one single outcome variable measured repeatedly over time. However, experiments in medical, industrial and social science research are often complex and characterized by several outcomes measured repeatedly over time. This paper focuses on statistical modeling of multivariate longitudinal data that are obtained from such experiments.

The example that motivates our research is the SHARe Framingham Heart Study of the National Hearth Lung and Blood Institute (cf., Cupples et al., 2007), in which 1826 participants were followed 7 times each at different ages. Multiple medical indices that are all important risk factors of stroke, including the systolic blood pressure (mmHg), diastolic blood pressure (mmHg), total cholesterol level (mg/100ml), and glucose level (mg/100ml), were measured at each time for each participant, and it was the interest of the medical researchers to know how these indices change over time. Similar studies have been reported in the literature. See, for instance, Godleski et al. (2000), Roy and Lin (2000), and Fieuws and Verbeke (2006).

In the literature, there is some existing research about statistical analysis of multivariate longitudinal data. However, almost all of them assume that the mean response follows a parametric model (cf. Gray and Brookmeyer, 2000; O'Brien and Fitzmaurice, 2004), or the error term follows a given parametric distribution (cf. Coull and Staudenmayer, 2004; Fieuws and Verbeke, 2006; Roy and Lin, 2000). In cases when all the model assumptions are valid, these methods should be effective. But, in practice, it is often difficult to obtain sufficient prior information for specifying the parametric models properly. There is some existing research on nonparametric or semiparametric modeling of longitudinal data. See, for instance, Liang and Zeger (1986), Lin and Carroll (2000), Lin and Carroll (2001), Wang (2003), Fitzmaurice et al. (2004), Weiss (2005), Chen and Jin (2005) and Li (2011). All such existing nonparametric or semiparametric methods are for analyzing univariate longitudinal data. So far, we have not found any existing research on nonparametric modeling of multivariate longitudinal data.

In this paper, we develop a nonparametric modeling approach for analyzing multivariate longitudinal data. By our approach, possible correlation among different components of the response is properly accommodated, along with possible correlation across different time points. Our method is based on local polynomial kernel smoothing. It is described in detail in Section 2. In section 3, some of its theoretical properties are discussed. In section 4, a simulation study is presented. Furthermore, our method is applied to the real data of the SHARe Framingham Heart Study in that section. Some concluding remarks are given in Section 5. Some technical details are provided in an appendix.

2 Proposed Method

Assume that $\boldsymbol{y}_{ij} = (y_{ij1}, y_{ij2}, \ldots, y_{ijq})^T$ are q-dimensional response observed at the jth time point t_{ij} from the ith subject, for $j = 1, 2, \ldots, J$ and $i = 1, 2, \ldots, n$. Further, $\boldsymbol{y}_{ij}$ is assumed to follow the multivariate nonparametric regression model

$$\boldsymbol{y}_{ij} = \boldsymbol{m}(t_{ij}) + \boldsymbol{\varepsilon}_{ij}, \qquad j = 1, 2, ...J,\ i = 1, 2, ...n, \tag{1}$$

where $\boldsymbol{m}(t_{ij}) = (m_1(t_{ij}), m_2(t_{ij}), \ldots, m_q(t_{ij}))^T$ denotes the mean of $\boldsymbol{y}_{ij}$, and $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \ldots, \varepsilon_{ijq})^T$ is the q-dimensional random error. Let

$$Y_i = (\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iJ})^T, \qquad \varepsilon_i = (\boldsymbol{\varepsilon}_{i1}, \ldots, \boldsymbol{\varepsilon}_{iJ})^T,$$

$\text{vec}(Y_i)$ be a long vector created by connecting all columns of Y_i one after another, and $\text{vec}(\varepsilon_i)$ be a long vector created from the columns of ε_i in the same way. Then, Y_i and ε_i are $J \times q$ matrices, and $\text{vec}(Y_i)$ and $\text{vec}(\varepsilon_i)$ are Jq-dimensional long vectors. In model (1), we assume that, for $i = 1, 2, \ldots, n$,

$$E(\text{vec}(\varepsilon_i)|t_{i1}, \ldots, t_{iJ}) = 0, \tag{2}$$

and

$$\text{Cov}(\text{vec}(\varepsilon_i)|t_{i1}, \ldots, t_{iJ}) = \text{Cov}(\text{vec}(Y_i)|t_{i1}, \ldots, t_{iJ}) =: V_i, \tag{3}$$

where V_i is the conditional covariance matrix of $\text{vec}(Y_i)$ containing $q \times q$ sub-matrices. Each sub-matrix is a $J \times J$ matrix. The diagonal sub-matrices measure the correlation among different components of the response at individual time points for the ith subject, and the off-diagonal sub-matrices measure the correlation among response vectors at different time points. Therefore, model (1) is quite general that accommodates the correlation among the observed data across different time points and across different components of the multivariate response vector as well.

To estimate model (1), we consider using the local polynomial kernel smoothing approach that has been used in the literature for handling cases with univariate longitudinal data (e.g., Lin and Carroll 2001, Wang 2003, Chen and Jin 2005). With multivariate longitudinal data, it would be much more complicated to use this approach with the possible correlation among different components of the response accommodated. To this end, let us first define some notations. In this paper, we use $\text{diag}\{a_{jl}, j = 1, ...J, l = 1, ..., q\}$ to denote a diagonal matrix with the $[j + (l-1)J]$th diagonal element to be a_{jl}. The inverse of a matrix throughout this paper means the Moore-Penrose generalized inverse of the matrix, and t denotes an arbitrary but fixed interior point of the domain of t_{ij}. The kernel function is denoted by $K(\cdot)$ which is chosen to be a symmetric density function with support [-1,1]. Typical choices of $K(\cdot)$ are the Epanechnikov kernel $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ and the uniform kernel $K(u) = 0.5I(|u| \leq 1)$, where $I(\cdot)$ is the indicator function. Define $K_h(u) = K(u/h)/h$, where h is a bandwidth. In multivariate setting, we need to use a q-dimensional bandwidth vector $\boldsymbol{H}$ for the q components of the response to allow different degrees of smoothing in different components. Let $\boldsymbol{H} = (h_1, ..., h_q)^T$,

$$K_{i\boldsymbol{H}} = \text{diag}\{K_{h_l}(t_{ij} - t), j = 1, ..., J, l = 1, ...q\}$$

and

$$W_i = \left(K_{i\boldsymbol{H}}^{-\frac{1}{2}} \widehat{V}_i K_{i\boldsymbol{H}}^{-\frac{1}{2}}\right)^{-1} = K_{i\boldsymbol{H}}^{\frac{1}{2}} \left(\tilde{I}_i \widehat{V}_i \tilde{I}_i\right)^{-1} K_{i\boldsymbol{H}}^{\frac{1}{2}},$$

where $\widehat{V}_i$ is an estimator of V_i, and

$$\begin{aligned} \tilde{I}_i &= \text{diag}\left\{I\left(K_{h_l}(t_{ij} - t) > 0\right), \ j = 1, ...J, \ l = 1, ..., q\right\} \\ &= \text{diag}\left\{I\left(|t_{ij} - t| \leq h_l\right), \ j = 1, ...J, \ l = 1, ..., q\right\}. \end{aligned}$$

For a positive integer p, let us consider the pth order local polynomial kernel smoothing procedure

$$\min_{\text{vec}(\beta) \in R^{q(p+1)}} \sum_{i=1}^{n} \left[\text{vec}(Y_i) - (I_q \otimes X_i)\text{vec}(\beta)\right]^T W_i \left[\text{vec}(Y_i) - (I_q \otimes X_i)\text{vec}(\beta)\right], \tag{4}$$

where $\otimes$ denotes the Kronecker product, and

$$X_i = \begin{pmatrix} 1 & (t_{i1} - t) & \dots & (t_{i1} - t)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_{iJ} - t) & \dots & (t_{iJ} - t)^p \end{pmatrix}_{J \times (p+1)}, \qquad \beta = \begin{pmatrix} \beta_0^{(1)} & \dots & \beta_0^{(q)} \\ \vdots & \ddots & \vdots \\ \beta_p^{(1)} & \dots & \beta_p^{(q)} \end{pmatrix}_{(p+1) \times q}.$$

In (4), the possible correlation among different response components has been accommodated by using $W_i = \left(K_{i\boldsymbol{H}}^{-\frac{1}{2}} \widehat{V}_i K_{i\boldsymbol{H}}^{-\frac{1}{2}}\right)^{-1}$. It can be checked that, in cases when we know that the q response components are independent of each other (i.e., V_i and $\widehat{V}_i$ are block diagonal matrices), the procedure (4) is equivalent to applying the univariate method by Chen and Jin (2005) to each component of the response vector.

It can be checked that the solution of (4) is

$$\widehat{\text{vec}(\beta)} = \left[\sum_{i=1}^{n} (I_q \otimes X_i)^T W_i (I_q \otimes X_i)\right]^{-1} \left[\sum_{i=1}^{n} (I_q \otimes X_i)^T W_i \text{vec}(Y_i)\right]. \tag{5}$$

Then, the pth order local polynomial kernel estimators of $\boldsymbol{m}^{(k)}(t) = (m_1^{(k)}(t), \dots, m_q^{(k)}(t))^T$, for $k = 0, ..., p$, are

$$\widehat{\boldsymbol{m}}^{(k)}(t) = k! \widehat{\text{vec}(\beta)}^T (I_q \otimes \boldsymbol{e}_{k+1}), \tag{6}$$

where $\boldsymbol{e}_{k+1}$ is a $(p+1)$-dimensional vector that has the value of 1 at the $(k+1)$th position and 0 at all other positions. In the special case when $k=0$, (6) becomes to be

$$\widehat{\boldsymbol{m}}^{(0)}(t) = \widehat{\mathrm{vec}(\beta)}^T (I_q \otimes \boldsymbol{e}_1)$$

and it is the pth order local polynomial kernel estimator of $\boldsymbol{m}(t)$.

In (4), we need to provide a reasonable estimator $\widehat{V}_i$ of the covariance matrix V_i. In practice, for each subject, if there are replicated observations at each time point, then V_i can be estimated by their empirical estimators (i.e., sample covariance matrices). Otherwise, some assumptions on V_i are necessary. For instance, if it is reasonable to assume that V_i are the same for all i, then the common covariance matrix can be estimated by the the method described as follows. First, we use the local linear kernel smoothing procedure to estimate individual components of $\boldsymbol{m}(\cdot)$ separately, using the Epanechnikov kernel function and the bandwidths determined by the conventional cross-validation (CV) procedure. The estimators are denoted as $\widetilde{\boldsymbol{m}}(\cdot) = (\widetilde{m}_1(\cdot), ..., \widetilde{m}_q(\cdot))$. Then, we compute the residuals

$$\widetilde{\varepsilon}_{ijl} = y_{ijl} - \widetilde{m}_l(t_{ij}), \qquad i = 1,2,\ldots,n, j = 1,2,\ldots,J, l = 1,2,\ldots,q.$$

The $([(l-1)J+j], [(s-1)J+k])$th element of V_i can be estimated by the following kernel estimator

$$\widehat{\mathrm{Cov}}(\varepsilon_{ijl}, \varepsilon_{iks}) = \begin{cases} \dfrac{\sum_{v=1}^n \widetilde{\varepsilon}_{vjl}\widetilde{\varepsilon}_{vks} K\left(\frac{t_{vj}-t_{ij}}{g_l}\right) K\left(\frac{t_{vk}-t_{ik}}{g_s}\right)}{\sum_{v=1}^n K\left(\frac{t_{vj}-t_{ij}}{g_l}\right) K\left(\frac{t_{vk}-t_{ik}}{g_s}\right)}, & j \neq k \text{ or } l \neq s; \\ \dfrac{\sum_{v=1}^n \widetilde{\varepsilon}_{vjl}^2 K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)}{\sum_{v=1}^n K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)}, & j = k, l = s. \end{cases} \tag{7}$$

where g_l is the bandwidth for the response component l, $j,k = 1,2,\ldots,J$, and $l,s = 1,2,\ldots,q$. In (7), we can still use the Epanechnikov kernel function, and the bandwidths $(g_1, ..., g_J)^T$ can be chosen as follows. Define new data

$$y^*_{ijl} = \widetilde{\varepsilon}^2_{ijl}, \qquad i = 1,2,\ldots,n, j = 1,2,\ldots,J, l = 1,2,\ldots,q.$$

Then, the mean of $\boldsymbol{y}^*_{ij} = (y^*_{ij1}, y^*_{ij2}, \ldots, y^*_{ijq})^T$ is a good approximation of the variance of $\boldsymbol{y}_{ij}$, denoted as $\sigma^2(t_{ij})$. Then, we can use the CV procedure to choose the bandwidths for the local linear kernel smoothing of the new data when the Epanechnikov kernel function is used. The resulting bandwidths can be used as the chosen values of $(g_1, ..., g_J)^T$. To specify V_i properly, we can also consider the method to use a time series model (e.g., an ARMA model) for specifying the possible correlation of the observed data across different time points, as mentioned by Chen and Jin (2005) in univariate cases.

In certain applications, it is possible that some response components are not measured, or their values are missing, at some time points. To handle such cases, our proposed method should be modified accordingly, described as follows. Let δ_{ijl} be a binary variable taking the value of 0 when the observation of the lth component of $\boldsymbol{y}(t_{ij})$ is missing and 1 otherwise. Define

$$\Delta_i = \mathrm{diag}\{\delta_{ijl}, j = 1, ..., J, l = 1, ...q\}.$$

Then, the quantity $\widehat{\mathrm{Cov}}(\varepsilon_{ijl}, \varepsilon_{iks})$ in (7) should be changed to

$$\widehat{\mathrm{Cov}}'(\varepsilon_{ijl}, \varepsilon_{iks}) = \begin{cases} \dfrac{\sum_{v=1}^{n} \widetilde{\varepsilon}_{vjl}\widetilde{\varepsilon}_{vks}\delta_{vjl}\delta_{vks}K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)K\left(\frac{t_{vk}-t_{ik}}{g_s}\right)}{\sum_{v=1}^{n} K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)\delta_{vjl}\delta_{vks}K\left(\frac{t_{vk}-t_{ik}}{g_s}\right)}, & j \neq k \text{ or } l \neq s; \\ \dfrac{\sum_{v=1}^{n} \widetilde{\varepsilon}_{vjl}^2\delta_{vjl}K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)}{\sum_{v=1}^{n} \delta_{vjl}K\left(\frac{t_{vj}-t_{ij}}{g_l}\right)}, & j = k, l = s. \end{cases} \tag{8}$$

The resulting estimator of V_i is denoted as $\widehat{V}_i'$. Then, the formula (5) should be modified to

$$\widehat{\mathrm{vec}(\beta)}' = \left[\sum_{i=1}^{n}(I_q \otimes X_i)^T \Delta_i W_i' \Delta_i (I_q \otimes X_i)\right]^{-1} \left[\sum_{i=1}^{n}(I_q \otimes X_i)^T \Delta_i W_i' \Delta_i \mathrm{vec}(Y_i)\right], \tag{9}$$

where

$$W_i' = \left(K_{i\boldsymbol{H}}^{-\frac{1}{2}} \Delta_i \widehat{V}_i' \Delta_i K_{i\boldsymbol{H}}^{-\frac{1}{2}}\right)^{-1} = K_{i\boldsymbol{H}}^{\frac{1}{2}} \left(\tilde{I}_i \Delta_i \widehat{V}_i' \Delta_i \tilde{I}_i\right)^{-1} K_{i\boldsymbol{H}}^{\frac{1}{2}}.$$

Finally, the pth order local polynomial kernel estimators of $\boldsymbol{m}^{(k)}(t)$, for $k = 0, ..., p$, can still be computed by (6), after $\widehat{\mathrm{vec}(\beta)}$ is replaced by $\widehat{\mathrm{vec}(\beta)}'$ in (9). The resulting estimators are denoted as $\widehat{\boldsymbol{m}}^{(k)\prime}(t)$.

3 Asymptotic Properties

In this section, we study the theoretical properties of our proposed method described in the previous section. These properties requires some regularity conditions on the local distribution of the design points, which are first described along with the necessary notations.

Let Ω_v, for $1 \leq v \leq 2^J - 1$, be the $2^J - 1$ distinct non-empty subsets of $\{1, ..., J\}$, and $B(t, \delta)$ denote the interval $[t - \delta, t + \delta]$. Assume that the design points $(t_{i1}, ..., t_{iJ})^T$, for $i = 1, ..., n$, are independent and identically distributed, and that their partial density at any given point t in the design space exists. The concept of partial density was discussed in Chen and Jin (2005), and it says that there exists a constant $\delta_0 > 0$ such that, for all $u \in B(t, \delta_0)$ and all $v = 1, ..., 2^J - 1$, we have

$$
\begin{aligned}
&\Pr\{t_{1j} \in B(u,\delta), \text{ and the elements in } \{t_{1j}, j \in \Omega_v\} \text{ are all equal, and } t_{1j_1} \neq t_{1j} \text{ if } j_1 \notin \Omega_v \text{ and } j \in \Omega_v\} \\
&\qquad = \int_{-\delta}^{\delta} f_v(z+u)dz \\
&\qquad = \Pr\{t_{1j} \in B(u,\delta) \text{ for all } j \in \Omega_v, \text{ and } t_{1j} \notin B(u,\delta) \text{ for all } j \notin \Omega_v\} + o(\delta)
\end{aligned}
$$

for all $0 < \delta < 2\delta_0$, where $f_v(.)$, for $1 \leq v \leq 2^J - 1$, are nonnegative continuous functions on $B(t_0, 2\delta_0)$ such that $\sum_{v=1}^{2^J-1} f_v(z) > 0$ for all $z \in B(t, 2\delta_0)$. This condition ensures that the chance for two design points to take values both in a small neighborhood of t is negligible unless they belong to the same Ω_v.

Let $\mathcal{S}_v(0) = \{t_{1j} = t \text{ for all } j \in \Omega_v, \text{ and } t_{1j} \neq t \text{ for all } j \notin \Omega_v\}$, and define

$$
\xi_v^{(sk)} = E\{(\tilde{\boldsymbol{e}}_s \otimes \mathbf{1_0})^T (\tilde{I}_{v0} \widehat{V}_1 \tilde{I}_{v0})^{-1} (\tilde{\boldsymbol{e}}_k \otimes \mathbf{1_0}) | \mathcal{S}_v(0)\}, \qquad \text{for } s, k = 1, ..., q,
$$

where $\tilde{I}_{v0} = I_q \otimes \text{diag}\{I(1 \in \Omega_v), ..., I(J \in \Omega_v)\}$ is a $qJ \times qJ$ nonrandom matrix, $\tilde{\boldsymbol{e}}_k$ is a q-dimensional vector with 1 at the kth position and 0 at all other positions, and $\mathbf{1_0}$ is a J-dimensional vector with all components equal to 1. We further define

$$
\begin{aligned}
V_0(t) &= \text{Cov}(\text{vec}(\varepsilon_1) | t_{11} = t, \ldots, t_{1J} = t) \\
\tilde{\xi}_v^{(sk)}(t) &= E\{(\tilde{\boldsymbol{e}}_s \otimes \mathbf{1_0})^T (\tilde{I}_{v0} \widehat{V}_1 \tilde{I}_{v0})^{-1} V_0(t) (\tilde{I}_{v0} \widehat{V}_1 \tilde{I}_{v0})^{-1} (\tilde{\boldsymbol{e}}_k \otimes \mathbf{1_0}) | \mathcal{S}_v(0)\} \\
\bar{\xi}_{v,l_1 l_2}^{(sk)}(t) &= E\{(\tilde{\boldsymbol{e}}_s \otimes \mathbf{1_0})^T (\tilde{I}_{v_0} \widehat{V}_1 \tilde{I}_{v_0})^{-1} (E_{l_1} \otimes I_J) V_0(t) (E_{l_2} \otimes I_J) (\tilde{I}_{v_0} \widehat{V}_1 \tilde{I}_{v_0})^{-1} (\tilde{\boldsymbol{e}}_k \otimes \mathbf{1_0}) | \mathcal{S}_v(0)\},
\end{aligned}
$$

where E_l is a $q \times q$ matrix with a single 1 at the lth diagonal position and with 0 at all other positions, for $l = 1, ..., q$. Set $h_{\max} = \max\{h_1, ..., h_q\}$, and assume that $h_l = c_l h_{\max}$, where $0 < c_l \leq 1$ are constants, for $l = 1, ..., q$. Define

$$
\begin{aligned}
\mu_j(h_s, h_k) &= (h_s h_k)^{-\frac{1}{2}} \int z^j K^{\frac{1}{2}}(h_{\max} z / h_s) K^{\frac{1}{2}}(h_{\max} z / h_k) dz \\
\nu_j(h_s, h_k, h_{l_1}, h_{l_2}) &= (h_s h_k h_{l_1} h_{l_2})^{-\frac{1}{2}} \int z^j K^{\frac{1}{2}}(h_{\max} z / h_s) \\
&\quad \times K^{\frac{1}{2}}(h_{\max} z / h_k) K^{\frac{1}{2}}(h_{\max} z / h_{l_1}) K^{\frac{1}{2}}(h_{\max} z / h_{l_2}) dz, \\
\nu_{m+l,v}^{(sk)}(t) &= \sum_{l_1, l_2 = 1}^{q} \bar{\xi}_{v,l_1 l_2}^{(sk)}(t) \nu_{m+l}(h_s, h_k, h_{l_1}, h_{l_2}).
\end{aligned}
$$

Then, it can be checked that $\mu_j(h_s, h_k) = O(h_{\max}^{-1})$ and $\nu_j(h_s, h_k, h_{l_1}, h_{l_2}) = O(h_{\max}^{-2})$ for any $s, k, l_1, l_2 \in \{1, ..., q\}$. Let S and $\bar{S}$ be both $[q(p+1)] \times [q(p+1)]$ matrices with the $[(p+1)(s-1)+m+$

$1, (p+1)(k-1)+l+1]$th elements equal to $\sum_{v=1}^{2^J-1} f_v(t)\xi_v^{(sk)}\mu_{m+l}(h_s, h_k)$ and $\sum_{v=1}^{2^J-1} f_v(t)\nu_{m+l,v}^{(sk)}(t)$, respectively, for $s, k, m, l \in \{1, ..., q\}$. Then, we have the following results.

Proposition 1 Let $\mathcal{F}_n$ denote the σ-algebra generated by $(t_{i1}, ..., t_{iJ})$, for $i = 1, ..., n$. Assume that the design points $(t_{i1}, ..., t_{iJ})^T$, for $i = 1, ..., n$, are independent and identity distributed and that their partial density exists at any given point t in the design space. The elements of V_i defined in (3) are assumed to be continuous functions of $(t_{i1}, ..., t_{iJ})$, and the components of the $(p+1)$th derivative $\boldsymbol{m}^{(p+1)}(t)$ of $\boldsymbol{m}(t)$ are assumed to be continuous functions of t, for $i = 1, ..., n$. Moreover, it is assumed that $h_l = c_l h_{\max}$, where $0 < c_l \leq 1$ are constants, for $l = 1, ..., q$, $h_{\max} = o(1)$, and $1/(nh_{\max}) = o(1)$. Then, we have the following results.

(i) The conditional covariance of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$\mathrm{Cov}\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k!^2}{nh_{\max}^{1+2k}}[(I_q \otimes \boldsymbol{e}_{k+1}^T)S^{-1}\bar{S}S^{-1}(I_q \otimes \boldsymbol{e}_{k+1})] + o_P\left(\frac{1}{nh_{\max}^{1+2k}}\right). \tag{10}$$

(ii) The conditional bias of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$Bias\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k!}{(p+1)!}h_{\max}^{p+1-k}[(I_q \otimes \boldsymbol{e}_{k+1}^T)S^{-1}\boldsymbol{D}] + o_P(h_{\max}^{p+1-k}), \tag{11}$$

where $\boldsymbol{D} = (d_{10}, ..., d_{1p}, ..., d_{q0}, ..., d_{qp})^T$, and

$$d_{sk} = \sum_{v=1}^{2^J-1}\sum_{l=1}^{q} f_v(t)m_l^{(p+1)}(t)\xi_v^{(sl)}\mu_{k+p+1}(h_s, h_l), \qquad \text{for } s = 1, ..., q,\ k = 0, ..., p.$$

Proposition 1 shows that the conditional covariance and the conditional bias of $\widehat{\boldsymbol{m}}^{(k)}(t)$ converge to 0 with the rates $O_P\left(\frac{1}{nh_{\max}^{1+2k}}\right)$ and $O_P(h_{\max}^{p+1-k})$, respectively, which are the same as those in univariate cases provided by Chen and Jin (2005). These results are derived in a quite general setting. In some special cases, they can have simpler expressions. For instance, in cases when different components of $\boldsymbol{m}(\cdot)$ have similar smoothness, we can use a bandwidth vector with $h_1 \sim \cdots \sim h_q \sim h_{\max}$. In such cases, $\mu_j(h_s, h_k) \approx \frac{1}{h_{\max}}\int u^j K(u)du =: \frac{1}{h_{\max}}\mu_j$, and $\nu_j(h_s, h_k, h_{l_1}, h_{l_2}) \approx \frac{1}{h_{\max}^2}\int u^j K^2(u)du =: \frac{1}{h_{\max}^2}\nu_j$, where "$\approx$" means that some higher order terms have been omitted in the related expressions. Then, define $\boldsymbol{c}_p = (\mu_{p+1}, ..., \mu_{2p+1})^T$, $S_1 = (\mu_{i+j})_{0\leq i,j\leq p}$, $\bar{S}_1 = (\nu_{i+j})_{0\leq i,j\leq p}$, and let

$$\boldsymbol{C} = \mathrm{diag}\left\{\sum_{v=1}^{2^J-1}\sum_{l=1}^{q} f_v(t)m_l^{(p+1)}(t)\xi_v^{(1l)}, \ldots, \sum_{v=1}^{2^J-1}\sum_{l=1}^{q} f_v(t)m_l^{(p+1)}(t)\xi_v^{(ql)}\right\},$$

$$N = \left(\sum_{v=1}^{2^J-1} f_v(t) \bar{\xi}_v^{(sk)}(t) \right)_{q\times q} \qquad \text{and} \qquad M = \left(\sum_{v=1}^{2^J-1} f_v(t) \xi_v^{(sk)} \right)_{q\times q},$$

where $\bar{\xi}_v^{(sk)}(t) = \sum_{l_1,l_2=1}^{q} \bar{\xi}_{v,l_1l_2}^{(sk)}(t)$. In such cases, the results in Proposition 1 can be simplified to those in Corollary 1 below.

Corollary 1 Besides the conditions in Proposition 1, we further assume that $h_1 \sim \cdots \sim h_q$. Then, we have the following results.

(i) The conditional covariance of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$\mathrm{Cov}\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k!^2}{nh_{\max}^{1+2k}} \boldsymbol{e}_{k+1}^T S_1^{-1} \bar{S}_1 S_1^{-1} \boldsymbol{e}_{k+1} M^{-1} N M^{-1} + o_P\left(\frac{1}{nh_{\max}^{1+2k}}\right). \tag{12}$$

(ii) The conditional bias of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$Bias\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k! h_{\max}^{p+1-k}}{(p+1)!} \boldsymbol{e}_{k+1}^T S_1^{-1} \boldsymbol{c}_p M^{-1} \boldsymbol{C} + o_P(h_{\max}^{p+1-k}). \tag{13}$$

In cases when $p-k$ is even, the first term on the right-hand-side of the above expression is actually 0.

Compared to expressions (10) and (11), the leading terms of expressions (12) and (13) are much simpler. For practical purpose, we can use $h_1 = \cdots = h_q = h$ for simplicity. Another special case that deserves our attention is the one when different response components are independent of each other. In this case, the matrices $\widehat{V}_i$ are nearly block diagonal. Consequently, our proposed method is similar to the one that handles individual response components separately. Results of Proposition 1 in this case can be simplified to the ones in Corollary 2 below.

Corollary 2 Besides the conditions in Proposition 1, we further assume that different response components are independent of each other. Then, we have the following results.

(i) The conditional covariance of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$\begin{aligned} \mathrm{Cov}\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = {} & \frac{k!^2}{n} \boldsymbol{e}_{k+1}^T S_1^{-1} \bar{S}_1 S_1^{-1} \boldsymbol{e}_{k+1} \\ & \times diag\left\{ \frac{\sum_{v=1}^{2^J-1} f_v(t) \tilde{\xi}_v^{(ll)}(t)}{\{\sum_{v=1}^{2^J-1} f_v(t) \xi_v^{(ll)}\}^2} h_l^{1+2k},\ l=1,...,q \right\} + o_P\left(\frac{1}{nh_{\max}^{1+2k}}\right). \end{aligned} \tag{14}$$

(ii) The conditional bias of $\widehat{\boldsymbol{m}}^{(k)}(t)$ is

$$Bias\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k!}{(p+1)!}\boldsymbol{e}_{k+1}^T S_1^{-1}\boldsymbol{c}_p \times diag\left\{m_l^{(p+1)}(t)h_l^{p+1-k},\, l=1,...,q\right\} + o_P(h_{\max}^{p+1-k}). \tag{15}$$

In cases when $p-k$ is even, the first term on the right-hand-side of the above expression is actually 0.

Next, we discuss the properties of our proposed method in cases when there are missing observations. First, we intriduce some extra notations. Let $\xi_v^{\prime(sk)}$ denote $\xi_v^{(sk)}$ after the quantity $(\tilde{I}_{v0}\widehat{V}_1\tilde{I}_{v0})^{-1}$ is replaced by $(\tilde{I}_{v0}\Delta_1\widehat{V}_1\Delta_1\tilde{I}_{v0})^{-1}$ in its definition, and $\bar{\xi}_{v,l_1l_2}^{\prime(sk)}(t)$ denote $\bar{\xi}_{v,l_1l_2}^{(sk)}(t)$ after $(\tilde{I}_{v0}\widehat{V}_1\tilde{I}_{v0})^{-1}$ and $\widehat{V}_0(t)$ are replaced by $(\tilde{I}_{v0}\Delta_1\widehat{V}_1\Delta_1\tilde{I}_{v0})^{-1}$ and $\Delta_1\widehat{V}_0(t)\Delta_1$, respectively, in its definition. Furthermore, let S' and $\bar{S}'$ be both $[q(p+1)]\times[q(p+1)]$ matrices with their $[(p+1)(s-1)+m+1, (p+1)(k-1)+l+1]$th elements to be $\sum_{v=1}^{2^J-1} f_v(t)\xi_v^{\prime(sk)}\mu_{m+l}(h_s,h_k)$ and $\sum_{v=1}^{2^J-1} f_v(t)\nu_{m+l,v}^{\prime(sk)}(t)$, respectively, for $s,k,m,l\in\{1,...,q\}$, where

$$\nu_{m+l,v}^{\prime(sk)}(t) = \sum_{l_1,l_2=1}^{q} \bar{\xi}_{v,l_1l_2}^{(sk)}(t)\nu'_{m+l}(h_s,h_k,h_{l_1},h_{l_2}).$$

Then, we have the following results about the estimated model in cases with missing data that is described at the end of Section 2.

Corollary 3 Assume that the assumptions in Proposition 1 all hold and $P(\delta_{ijl}=0)=p_l$, for $i=1,2,\ldots,n$, $j=1,2,\ldots,J$, and $l=1,2,\ldots,q$, where $p_l\in[0,1)$ are probablity values that do not depend on i and j. Then, we have the following results.

(i) The conditional covariance of $\widehat{\boldsymbol{m}}^{(k)'}(t)$ is

$$\mathrm{Cov}\{\widehat{\boldsymbol{m}}^{(k)'}(t)|\mathcal{F}_n\} = \frac{k!^2}{nh_{\max}^{1+2k}}[(P^{-1}\otimes\boldsymbol{e}_{k+1}^T)S'^{-1}\bar{S}'S'^{-1}(P^{-1}\otimes\boldsymbol{e}_{k+1})] + o_P\left(\frac{1}{nh_{\max}^{1+2k}}\right).$$

where $P=diag\{p_1,...,p_q\}$.

(ii) The conditional bias of $\widehat{\boldsymbol{m}}^{(k)'}(t)$ is

$$Bias\{\widehat{\boldsymbol{m}}^{(k)'}(t)|\mathcal{F}_n\} = \frac{k!}{(p+1)!}h_{\max}^{p+1-k}[(P^{-1}\otimes\boldsymbol{e}_{k+1}^T)S'^{-1}\boldsymbol{D}'] + o_P(h_{\max}^{p+1-k}),$$

where $\boldsymbol{D}'=(d'_{10},...,d'_{1p},...,d'_{q0},...,d'_{qp})^T$, and

$$d'_{sk} = \sum_{v=1}^{2^J-1}\sum_{l=1}^{q} f_v(t)m_l^{(p+1)}(t)\xi_v^{\prime(sl)}\mu_{k+p+1}(h_s,h_l), \qquad \text{for } s=1,...,q,\ k=0,...,p.$$

4 Numerical Study

In this section, we investigate the numerical performance of the proposed method using several simulation examples and one real-data example. We also discuss estimation of the true covariance matrices V_i defined in (3) and selection of the bandwidth vector used in our proposed local smoothing estimators.

We first consider cases when no missing values are present in the observed data. In such cases, the simulated data are generated from the model (1) with $J = 3$, $q = 3$, and

$$m_1(t) = 2 \times \exp\{\sin(10t)\}, \;\; m_2(t) = 1 - \exp\{-t\}, \;\; m_3(x) = 1 - \exp\{-t\} + 2\sin(10t).$$

The error term $\mathrm{vec}(\varepsilon_i)$ follows the normal distribution with mean $\mathbf{0}$. Its correlation matrix is specified as follows: for $j, k, l, s = 1, 2, 3$,

$$\mathrm{corr}(\varepsilon_{1jl}, \varepsilon_{1ks}) = \begin{cases} 1, & \text{if } j = k,\, l = s \\ \rho_1, & \text{if } j \neq k,\, l = s \\ \rho_2, & \text{if } j = k,\, l \neq s \\ \rho_1\rho_2, & \text{if } j \neq k,\, l \neq s, \end{cases}$$

and the variances of its components are

$$\mathrm{var}(\varepsilon_{111}) = \mathrm{var}(\varepsilon_{112})/2 = \mathrm{var}(\varepsilon_{113})/3 = 0.25,$$

$$\mathrm{var}(\varepsilon_{121}) = \mathrm{var}(\varepsilon_{122})/2 = \mathrm{var}(\varepsilon_{123})/3 = 0.64,$$

$$\mathrm{var}(\varepsilon_{131}) = \mathrm{var}(\varepsilon_{132})/2 = \mathrm{var}(\varepsilon_{133})/3 = 0.36.$$

The design points $\{t_{ij}, j = 1, 2, 3, \;\; i = 1, 2, \ldots, n\}$ are generated from the uniform distribution $U[-2, 2]$, and they are independent from the random errors.

From the above definition, we can see that ρ_1 specifies the association of individual response components over different time points, and ρ_2 specifies the association among different response components at a given time point. To demonstrate the effectiveness of our proposed method, we consider the following three cases:

Case I: $\rho_1 = 0.8$ and $\rho_2 = 0$, in which individual response components are associated across different time points but the components are independent of each other at a given time point.

Case II: $\rho_1 = 0$ and $\rho_2 = 0.8$, in which individual response components are associated at a given time point but they are independent across different time points.

Case III: $\rho_1 = 0.8$ and $\rho_2 = 0.8$, in which individual response components are associated across different time points and the components are associated at a given time point as well.

With the multivariate longitudinal data, besides the proposed method described in the previous two sections, which is denoted as MULTIVARIATE here, there are a number of alternative approaches. One alternative approach is to apply the univariate method by Chen and Jin (2005) to each dimension of the multivariate longitudinal data, and obtain estimators of the individual components of $\boldsymbol{m}(\cdot)$ separately. This method is denoted as INDIVIDUAL in this section. Another alternative method is to use a simplified version of MULTIVARIATE, in which the same bandwidth h is used in all dimensions, as described in Corollary 1 in Section 3. This simplified version is denoted as SIMPLIFIED here. For each method, we compute the values of the estimator $\widehat{\boldsymbol{m}}(t)$ at 101 grid points $\{t_j = -1.8 + 0.036 \times j,\ j = 0, ..., 100\}$. Then, the following three performance measures are computed:

$$\begin{aligned}
\text{Bias}_l &= \frac{1}{101}\sum_{j=0}^{100} |m_l(t_j) - \widehat{m}_l(t_j)| \\
\text{SD}_l &= \text{sample standard deviation of } \{m_l(t_j) - \widehat{m}_l(t_j),\ j = 0, 1, 2, \ldots, 100\} \\
\text{MISE}_l &= \frac{4}{101}\sum_{j=0}^{100} \left(m_l(t_j) - \widehat{m}_l(t_j)\right)^2,
\end{aligned}$$

where $l = 1, 2, 3$ is the index of the response components. To remove some randomness, all presented values of these measures in this section are averages computed from 100 replicated simulations.

In all three methods considered, the Epanechnikov kernel function described in Section 2 and $p = 1$ (i.e., local linear smoothing) are used in the local polynomial kernel smoothing procedures (cf., (4)). For a fair comparison, we first use the true covariance matrices V_i, instead of their estimates, in all methods. The optimal bandwidths of each method are then searched by minimizing the MISE value. The searched optimal bandwidths and the corresponding values of the three performance measures in the three cases considered are presented in Tables 1 and 2, respectively, for two sample sizes $n = 100$ and $n = 200$. For each measure, its values corresponding to the three response components are presented separately, together with their summation, denoted as SUM.

From Table 1, it can be seen that, in case I when the three response components are independent of each other, the methods MULTIVARIATE and INDIVIDUAL perform exactly the same. As a

Table 1: Averaged performance measures Bias, SD, and MISE, based on 100 replicated simulations, of the methods MULTIVARIATE, INDIVIDUAL, and SIMPLIFIED in the case when $n = 100$. The numbers in $\boldsymbol{H}$ are the searched optimal bandwidths.

		MULTIVARIATE			INDIVIDUAL			SIMPLIFIED		
Case	Components	Bias	SD	MISE	Bias	SD	MISE	Bias	SD	MISE
I		$\boldsymbol{H} = (0.08, 0.65, 0.11)^T$			$\boldsymbol{H} = (0.08, 0.65, 0.11)^T$			$\boldsymbol{H} = (0.11, 0.11, 0.11)^T$		
	1	0.103	0.224	0.246	0.103	0.224	0.246	0.159	0.199	0.286
	2	0.036	0.114	0.057	0.036	0.114	0.057	0.020	0.269	0.267
	3	0.140	0.317	0.455	0.140	0.317	0.455	0.117	0.334	0.470
	SUM	0.279	0.655	0.758	0.279	0.655	0.758	0.295	0.802	1.023
II		$\boldsymbol{H} = (0.08, 0.45, 0.11)^T$			$\boldsymbol{H} = (0.09, 0.5, 0.11)^T$			$\boldsymbol{H} = (0.1, 0.1, 0.1)^T$		
	1	0.102	0.186	0.201	0.139	0.211	0.280	0.169	0.196	0.304
	2	0.067	0.128	0.080	0.043	0.112	0.056	0.024	0.264	0.259
	3	0.119	0.253	0.297	0.151	0.307	0.449	0.127	0.325	0.459
	SUM	0.288	0.567	0.578	0.333	0.630	0.785	0.320	0.785	1.022
III		$\boldsymbol{H} = (0.07, 0.5, 0.11)^T$			$\boldsymbol{H} = (0.09, 0.65, 0.12)^T$			$\boldsymbol{H} = (0.1, 0.1, 0.1)^T$		
	1	0.074	0.208	0.211	0.129	0.217	0.287	0.188	0.193	0.334
	2	0.067	0.139	0.091	0.044	0.112	0.058	0.018	0.251	0.232
	3	0.105	0.268	0.312	0.159	0.298	0.440	0.133	0.314	0.443
	SUM	0.245	0.615	0.615	0.332	0.627	0.785	0.339	0.758	1.009

matter of fact, it can be checked that these two methods are equivalent in such cases. Compared to the method SIMPLIFIED, their MISE values are smaller across all three response components. This part of the results demonstrates that when the curvature of the three components of $\boldsymbol{m}(\cdot)$ are quite different, the method SIMPLIFIED may not be appropriate to use. It also confirms that the proposed method MULTIVARIATE is appropriate to use even in cases when the response components are actually independent. In case II when observations across different time points are independent but different response components are correlated, we can see that the method MULTIVARIATE performs better than both methods INDIVIDUAL and SIMPLIFIED in terms of the SUMs of the three performance measures, although it is slightly worse then the method INDIVIDUAL for estimating $m_2(\cdot)$. In case III when observations across different time points are correlated and different response components are correlated as well, the method MULTIVARIATE also performs better than both methods INDIVIDUAL and SIMPLIFIED in terms of the SUMs of the three performance measures. Similar conclusions can be made from results in Table 2.

In practice, the covariance matrices V_i are often unknown and they need to be estimated from observed data. Next, we investigate the performance of the three methods when V_i, $i = 1, ...n$, are assumed to be the same and are estimated by the procedure (7). The estimated V_i by (7) is used in

Table 2: Averaged performance measures Bias, SD, and MISE, based on 100 replicated simulations, of the methods MULTIVARIATE, INDIVIDUAL, and SIMPLIFIED in the case when $n = 200$. The numbers in $\boldsymbol{H}$ are the searched optimal bandwidths.

		MULTIVARIATE			INDIVIDUAL			SIMPLIFIED		
Case	Components	Bias	SD	MISE	Bias	SD	MISE	Bias	SD	MISE
I		$\boldsymbol{H} = (0.06, 0.5, 0.09)^T$			$\boldsymbol{H} = (0.06, 0.5, 0.09)^T$			$\boldsymbol{H} = (0.09, 0.09, 0.09)^T$		
	1	0.064	0.169	0.127	0.064	0.169	0.127	0.137	0.140	0.171
	2	0.031	0.083	0.031	0.031	0.083	0.031	0.015	0.184	0.125
	3	0.091	0.227	0.226	0.091	0.227	0.226	0.091	0.227	0.226
	SUM	0.185	0.479	0.383	0.185	0.479	0.383	0.242	0.551	0.522
II		$\boldsymbol{H} = (0.06, 0.4, 0.08)^T$			$\boldsymbol{H} = (0.06, 0.45, 0.09)^T$			$\boldsymbol{H} = (0.08, 0.08, 0.08)^T$		
	1	0.059	0.137	0.091	0.066	0.170	0.132	0.113	0.146	0.151
	2	0.041	0.091	0.038	0.035	0.079	0.030	0.017	0.198	0.144
	3	0.070	0.200	0.169	0.106	0.228	0.242	0.085	0.244	0.251
	SUM	0.170	0.428	0.298	0.207	0.477	0.404	0.214	0.587	0.546
III		$\boldsymbol{H} = (0.06, 0.45, 0.10)^T$			$\boldsymbol{H} = (0.06, 0.55, 0.1)^T$			$\boldsymbol{H} = (0.09, 0.09, 0.09)^T$		
	1	0.059	0.136	0.087	0.061	0.168	0.123	0.134	0.137	0.164
	2	0.054	0.101	0.050	0.031	0.082	0.030	0.019	0.186	0.128
	3	0.093	0.191	0.171	0.120	0.221	0.244	0.099	0.233	0.243
	SUM	0.206	0.427	0.308	0.212	0.470	0.397	0.252	0.555	0.535

all three methods in place of the true matrices V_i. The corresponding results of the three methods in cases when $n = 200$ and when the bandwidths are chosen to be optimal by minimizing the MISE values are presented in Table 3. From the table, we can see that similar conclusions can be made here to those from Tables 1 and 2, regarding the relative performance of the three methods. By comparing the results of MULTIVARIATE in Tables 2 and 3, we can see that they are almost the same, which implies that the procedure (7) for specifying V_i is quite efficient. Corresponding results in the case when $n = 100$ are similar and thus omitted here.

In practice, the optimal bandwidths are also unknown. To implement our proposed method in such cases, we propose using the cross-validation (CV) procedure to determine the bandwidths as follows. Let

$$CV_l(h_l) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}\left(y_l(t_{ij}) - \widehat{m}_{l,-i}(t_{ij})\right)^2, \qquad \text{for } l = 1, 2, 3,$$

and

$$CV(\boldsymbol{H}) = CV_1(h_1) + CV_2(h_2) + CV_3(h_3),$$

where $\widehat{m}_{l,-i}(\cdot)$ is the "leave-one-subject-out" estimator of $m_l(\cdot)$ obtained when the observations of the ith subject are not used. Then, the three bandwidths can be determined by minimizing $CV(\boldsymbol{H})$ over R_+^3. However, this minimization process might be time-consuming. To simplify the

Table 3: Averaged performance measures Bias, SD, and MISE, based on 100 replicated simulations, of the methods MULTIVARIATE, INDIVIDUAL, and SIMPLIFIED in the case when $n = 200$ and V_i are estimated. The numbers in $\boldsymbol{H}$ are the searched optimal bandwidths.

		MULTIVARIATE			INDIVIDUAL			SIMPLIFIED		
Case	Components	Bias	SD	MISE	Bias	SD	MISE	Bias	SD	MISE
I		$\boldsymbol{H} = (0.06, 0.5, 0.09)^T$			$\boldsymbol{H} = (0.06, 0.5, 0.09)^T$			$\boldsymbol{H} = (0.08, 0.08, 0.08)^T$		
	1	0.065	0.170	0.129	0.064	0.170	0.128	0.112	0.147	0.147
	2	0.032	0.083	0.032	0.032	0.082	0.032	0.016	0.197	0.143
	3	0.092	0.227	0.227	0.092	0.227	0.226	0.072	0.243	0.239
	SUM	0.189	0.480	0.387	0.189	0.479	0.386	0.200	0.587	0.529
II		$\boldsymbol{H} = (0.06, 0.4, 0.09)^T$			$\boldsymbol{H} = (0.06, 0.4, 0.09)^T$			$\boldsymbol{H} = (0.08, 0.08, 0.08)^T$		
	1	0.066	0.138	0.096	0.066	0.171	0.133	0.114	0.147	0.153
	2	0.041	0.091	0.038	0.029	0.084	0.030	0.017	0.199	0.146
	3	0.091	0.193	0.172	0.107	0.229	0.244	0.086	0.246	0.256
	SUM	0.197	0.421	0.306	0.202	0.483	0.407	0.217	0.591	0.555
III		$\boldsymbol{H} = (0.06, 0.45, 0.09)^T$			$\boldsymbol{H} = (0.06, 0.55, 0.1)^T$			$\boldsymbol{H} = (0.09, 0.09, 0.09)^T$		
	1	0.063	0.138	0.091	0.061	0.168	0.124	0.138	0.138	0.171
	2	0.040	0.095	0.040	0.033	0.081	0.031	0.019	0.187	0.130
	3	0.084	0.200	0.178	0.122	0.222	0.247	0.101	0.235	0.248
	SUM	0.187	0.432	0.309	0.216	0.470	0.402	0.258	0.560	0.549

computation, we suggest using a two-step CV procedure instead, by noticing from results in Tables 1 and 2 that the optimal bandwidths of the two methods MULTIVARIATE and INDIVIDUAL are actually quite close to each other. In the first step, we determine the individual bandwidths $\{h_l, l = 1, 2, 3\}$ separately, by applying the CV procedure to the method INDIVIDUAL. The selected bandwidths from this step are denoted as $\{h_{l,0}, l = 1, 2, 3\}$. Then, in the second step, we determine the three bandwidths by minimizing $CV(\boldsymbol{H})$ in a small neighborhood of $(h_{1,0}, h_{2,0}, h_{3,0})^T$. In our simulation study, we use the neighborhood $\{(h_1, h_2, h_3)|h_1 = h_{10} + 0.01\delta_l,\ h_2 = h_{20} + 0.05\delta_2,\ h_3 = h_{30} + 0.01\delta_3,\ \delta_1, \delta_2, \delta_3 = 0, \pm 1, \pm 2\}$. The method MULTIVARIATE with the bandwidths chosen by the CV procedure and the covariance matrix estimated by (7) is denoted as MULTIVARIATE-CV. Its results corresponding to the cases considered in Tables 1 and 2 are presented in Table 4. By comparing the tables, we can see that MULTIVARIATE-CV performs a little worse than MULTIVARIATE, but it still performs favorably, compared to the methods INDIVIDUAL and SIMPLIFIED, in cases when the response components are correlated, even if the methods INDIVIDUAL and SIMPLIFIED use their optimal bandwidths.

Next, we consider an example in which missing observations are present in the observed data. The setup of this example is the same as that of Table 4, except that $n = 200$, the probabilities

Table 4: Averaged performance measures Bias, SD, and MISE, based on 100 replicated simulations, of the method MULTIVARIATE-CV in the cases when $n = 100$ or 200.

		Case I			Case II			Case III		
n	Components	Bias	SD	MISE	Bias	SD	MISE	Bias	SD	MISE
100	1	0.118	0.236	0.297	0.136	0.192	0.258	0.137	0.197	0.252
	2	0.062	0.117	0.075	0.083	0.127	0.093	0.084	0.123	0.094
	3	0.157	0.326	0.497	0.151	0.274	0.389	0.144	0.283	0.376
	SUM	0.336	0.679	0.870	0.371	0.593	0.740	0.365	0.604	0.722
200	1	0.079	0.165	0.138	0.090	0.130	0.118	0.101	0.125	0.117
	2	0.048	0.081	0.041	0.065	0.076	0.046	0.060	0.078	0.041
	3	0.109	0.211	0.227	0.117	0.179	0.187	0.121	0.192	0.211
	SUM	0.236	0.457	0.407	0.272	0.385	0.351	0.282	0.396	0.369

of missing observations for the three components are assumed to be $p_1 = p_2 = p_3 = \pi$, and $\pi = 0.05, 0.1$, or 0.2. The corresponding results are presented in Table 5, which are computed by (6) after $\widehat{\text{vec}(\beta)}$ is replaced by $\widehat{\text{vec}(\beta)}'$ in (9). From the table, it can be seen that (i) the MISE value increases when p increases, which is intuitively reasonable, and (ii) our proposed method performs reasonably well in such cases.

Table 5: Averaged performance measures Bias, SD, and MISE, based on 100 replicated simulations, of the method MULTIVARIATE-CV in the cases when $n = 200$, the probabilities of missing observations for the three components are $p_1 = p_2 = p_3 = \pi = 0.05, 0.1$, or 0.2.

		Case I			Case II			Case III		
π	Components	Bias	SD	MISE	Bias	SD	MISE	Bias	SD	MISE
0.05	1	0.061	0.188	0.148	0.059	0.152	0.105	0.052	0.151	0.098
	2	0.014	0.123	0.056	0.034	0.124	0.061	0.036	0.131	0.069
	3	0.092	0.241	0.251	0.081	0.213	0.195	0.070	0.222	0.203
	SUM	0.167	0.552	0.455	0.174	0.489	0.361	0.158	0.504	0.370
0.1	1	0.062	0.193	0.158	0.058	0.158	0.114	0.067	0.151	0.118
	2	0.015	0.122	0.056	0.034	0.123	0.061	0.045	0.121	0.069
	3	0.094	0.248	0.265	0.078	0.220	0.204	0.079	0.214	0.213
	SUM	0.171	0.563	0.479	0.170	0.501	0.379	0.191	0.486	0.400
0.2	1	0.066	0.207	0.182	0.063	0.169	0.132	0.076	0.158	0.128
	2	0.014	0.131	0.063	0.034	0.138	0.082	0.049	0.138	0.091
	3	0.096	0.268	0.305	0.092	0.237	0.247	0.115	0.224	0.311
	SUM	0.176	0.606	0.550	0.189	0.544	0.461	0.240	0.520	0.530

Finally, we apply our proposed method to the real-data example about the SHARe Framingham Heart Study that is described in Section 1. The raw data can be downloaded from the web page $http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\_id = phs000007.v4.p2$.

After deleting the patients with missing observations. A total of $n = 1028$ non-stroke patients with ages from 14 to 85 are included in our analysis. In this example, the response is 4-dimensional (i.e., $q = 4$), and each patient was followed 7 times (i.e., $J = 7$). In our proposed method, we consider using $p = 1$ (i.e., use the multivariate local linear kernel smoothing in (4)). The covariance matrices V_i are determined by the procedure (7). The bandwidth vector $\boldsymbol{H}$ is chosen using the two-step CV procedure described above, and the chosen bandwidth vector is $\boldsymbol{H} = (9, 5, 4, 6)^T$. The four estimated components of $\boldsymbol{m}(\cdot)$ are shown in the four plots of Figure 1 by the solid curves. After obtaining the estimator $\widehat{\boldsymbol{m}}(\cdot)$, we use the following method to estimate the variance functions of the components of the multivariate response $\boldsymbol{y}(t)$. First, we compute the residuals

$$\widehat{\varepsilon}_{ijl} = y_{ijl} - \widehat{m}_l(t_{ij}), \qquad i = 1, 2, \ldots, n, j = 1, 2, \ldots, J, l = 1, 2, \ldots, q.$$

Then, the estimators of the variance functions can be obtained, after we apply the proposed method described in Section 2 to the new data $\{\widehat{\varepsilon}^2_{ijl}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, J, l = 1, 2, \ldots, q\}$. Using the estimated variance functions, the pointwise 95% confidence bands of the components of $\boldsymbol{m}(\cdot)$ are constructed and presented in Figure 1 (a)-(d) by the dashed curves, along with the observed longitudinal data of the first 20 patients shown by little circles connected by thin lines. From the plots, it can be seen that our estimators describe the observed data reasonably well.

5 Concluding Remarks

In this paper, we have proposed a local smoothing method for analyzing multivariate longitudinal data. Our method can accommodate not only the correlation among observations across different time points, but also the correlation among different response components. The numerical results presented in the paper show that our proposed method performs well in applications. Although we focus on cases when the explanatory variable t is univariate in this paper, it is possible to generalize our proposed method for handling cases with multiple explanatory variables, using methods similar to the one by Ruppert and Wand (1994).

There are several issues that have not been addressed in this paper yet, which could be good future research topics. First, our numerical results show that the cross-validation procedure for choosing the bandwidths works reasonable well. However, as pointed out by Hall and Robinson (2009), the bandwidths chosen by this approach usually have a larger variability. Hall and Robinson (2009) proposed two procedures to overcome this limitation. Unfortunately, these procedures are

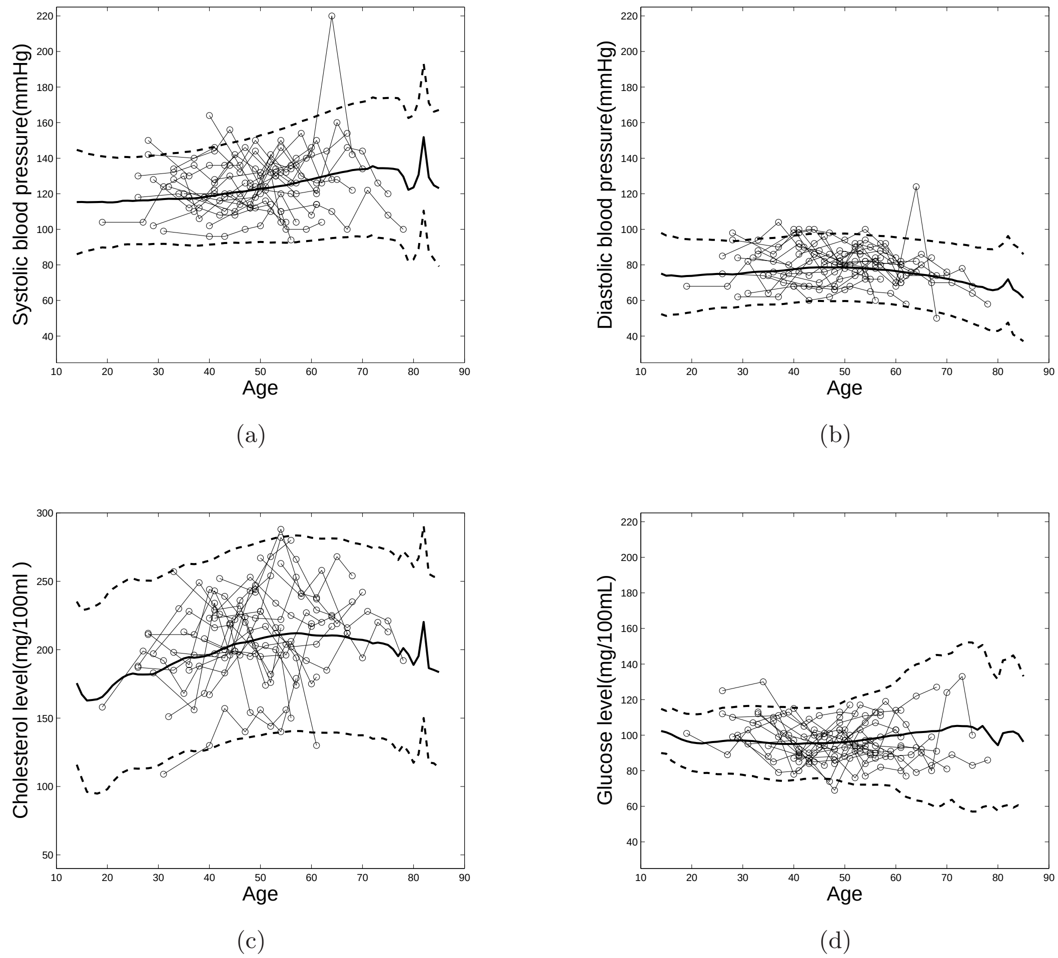

Figure 1: Estimated mean components, the pointwise 95% confidence bands of the true mean response components, and the observed longitudinal data of the first 20 patients in the dataset of the SHARe Framingham Heart Study. (a) systolic blood pressure, (b) diastolic blood pressure, (c) cholesterol level, and (d) glucose level.

computationally intensive. Therefore, it still requires much future research to propose an efficient and computationally simple procedure for choosing the bandwidths. Second, our proposed method may not be suitable for high-dimensional (e.g., $q \geq 20$) multivariate longitudinal data because of the complexity in computing estimators of V_i and in choosing the bandwidths. It requires much future research to develop appropriate methods for handling such cases as well. Third, in Corollary 3, it is assumed that the probabilities of missing observations of the response components are unchanged over time. In certain applications, this assumption may not be valid. If the probabilities of missing observations depend on observation times, then variable bandwidths might be more appropriate to use in our proposed method. At places with more missing observations, the bandwidths should be chosen larger; they can be chosen smaller at places with less missing observations. This topic is not trivial, and is left for our future research.

Acknowledgments: We thank the co-editor, Professor Naisyin Wang, the AE, and a referee for many helpful comments and suggestions, which greatly improved the quality of the paper. This research is supported in part by an NSF grant and by China Scholarship Council through its joint Ph.D program.

Appendix

Proof of Proposition 1

By the definitions of S, $\bar{S}$ and $\boldsymbol{D}$, we can show the following results:

$$S = O(h_{\max}^{-1}), \quad \bar{S} = O(h_{\max}^{-2}), \quad \boldsymbol{D} = O(h_{\max}^{-1}). \tag{A.1}$$

By the continuity of V_i and a direct algebraic manipulation, we can get the result that

$$\mathrm{Cov}(\widehat{\mathrm{vec}(\beta)}|\mathcal{F}_n) = A_n^{-1} B_n A_n^{-1}\{1 + o_P(1)\}, \tag{A.2}$$

where $A_n = \sum_{i=1}^n (I_q \otimes X_i)^T W_i (I_q \otimes X_i)$ and

$$B_n = \sum_{i=1}^n (I_q \otimes X_i^T) K_{i\boldsymbol{H}}^{\frac{1}{2}} (\tilde{I}_i \widehat{V}_i \tilde{I}_i)^{-1} K_{i\boldsymbol{H}}^{\frac{1}{2}} V_0(t) K_{i\boldsymbol{H}}^{\frac{1}{2}} (\tilde{I}_i \widehat{V}_i \tilde{I}_i)^{-1} K_{i\boldsymbol{H}}^{\frac{1}{2}} (I_q \otimes X_i).$$

Set $\widetilde{H} = \mathrm{diag}\{1, h_{\max}, ..., h_{\max}^p\}$ and $c_{ijsm} = (t_{ij} - t)^m K_{h_s}^{\frac{1}{2}}(t_{ij} - t)$, for $i = 1, ..., n$, $j = 1, ..., J$, $s = 1, ..., q$ and $m = 0, ..., p$. For every fixed $v = 1, ..., 2^J - 1$, let

$$\mathcal{S}_v(h_{\max}) = \{t_{1j} \in B(t, h_{\max}) \text{ for all } j \in \Omega_v, \text{ and } t_{1,j} \notin B(t, h_{\max}) \text{ for all } j \notin \Omega_v\}$$

Then, the existence condition of the partial density of $\{t_{ij}\}$ ensures that $\Pr\{t_{1j}$ are all equal for all $j \in \Omega_v | \mathcal{S}_v(h_{\max})\} = 1 + o(1)$ on $B(t, h_{\max})$, as $h_{\max} \to 0$. Let $a_{m+1,l+1}^{(sk)}$ denote the $((s-1)(p+1) + m + 1, (k-1)(p+1) + l + 1)$th element of A_n, $j_v \in \Omega_v$, and $\boldsymbol{C}_{ism} = (0, ..., 0, c_{i1sm}, ..., c_{iJsm}, 0, ..., 0)_{1 \times qJ}^T$.

Then,

$$
\begin{aligned}
E(a^{(sk)}_{m+1,l+1}) &= \sum_{i=1}^{n} E\{\boldsymbol{C}^T_{ism}(\tilde{I}_i\widehat{V}_i\tilde{I}_i)^{-1}\boldsymbol{C}_{ikl}\} \\
&= n\sum_{v=1}^{2^J-1} E\{\boldsymbol{C}^T_{1sm}(\tilde{I}_1\widehat{V}_1\tilde{I}_1)^{-1}\boldsymbol{C}_{1kl}|\mathcal{S}_v(h_{\max})\}\Pr\{\mathcal{S}_v(h_{\max})\} \\
&= n\sum_{v=1}^{2^J-1} E[(t_{1j_v}-t)^{m+l}K^{\frac{1}{2}}_{h_s}(t_{1j_v}-t)K^{\frac{1}{2}}_{h_k}(t_{1j_v}-t)I\{\mathcal{S}_v(h_{\max})\}] \\
&\quad \times E\{(\tilde{\boldsymbol{e}}_s\otimes\mathbf{1_0})^T(\tilde{I}_{v_0}\widehat{V}_1\tilde{I}_{v_0})^{-1}(\tilde{\boldsymbol{e}}_k\otimes\mathbf{1_0})|\mathcal{S}_v(0)\}\{1+o(1)\} \\
&= n\sum_{v=1}^{2^J-1}\frac{1}{\sqrt{h_sh_k}}\int_{t-h_{\max}}^{t+h_{\max}}(u-t)^{m+l}K^{\frac{1}{2}}\left(\frac{u-t}{h_s}\right)K^{\frac{1}{2}}\left(\frac{u-t}{h_k}\right)f_v(u)du \\
&\quad \times E\{(\tilde{\boldsymbol{e}}_s\otimes\mathbf{1_0})^T(\tilde{I}_{v_0}\widehat{V}_1\tilde{I}_{v_0})^{-1}(\tilde{\boldsymbol{e}}_k\otimes\mathbf{1_0})|\mathcal{S}_v(0)\}\{1+o(1)\} \\
&= n\sum_{v=1}^{2^J-1}\frac{h^{m+l+1}_{\max}}{\sqrt{h_sh_k}}\int_{-1}^{1}z^{m+l}K^{\frac{1}{2}}\left(\frac{h_{\max}}{h_s}z\right)K^{\frac{1}{2}}\left(\frac{h_{\max}}{h_k}z\right)f_v(t)dz \\
&\quad \times E\{(\tilde{\boldsymbol{e}}_s\otimes\mathbf{1_0})^T(\tilde{I}_{v_0}\widehat{V}_1\tilde{I}_{v_0})^{-1}(\tilde{\boldsymbol{e}}_k\otimes\mathbf{1_0})|\mathcal{S}_v(0)\}\{1+o(1)\} \\
&= nh^{m+l+1}_{\max}\mu_{m+l}(h_s,h_k)\sum_{v=1}^{2^J-1}f_v(t)\xi^{(sk)}_v\{1+o(1)\}.
\end{aligned}
$$

Similarly, we can show that $\{\text{var}(a^{(sk)}_{m+1,l+1}\}^{\frac{1}{2}} = o(nh^{m+l}_{\max})$. By combining these results, we have

$$
\begin{aligned}
a^{(sk)}_{m+1,l+1} &= E(a^{(sk)}_{m+1,l+1}) + O_p[\{\text{var}(a^{(sk)}_{m+1,l+1}\}^{\frac{1}{2}}] \\
&= nh^{m+l+1}_{\max}\mu_{m+l}(h_s,h_k)\textstyle\sum_{v=1}^{2^J-1}f_v(t)\xi^{(sk)}_v\{1+o(1)\}.
\end{aligned}
$$

Therefore,

$$
A_n = nh_{\max}[(I_q\otimes\widetilde{H})S(I_q\otimes\widetilde{H})]\{1+o_P(1)\}. \tag{A.3}
$$

Let $b^{(sk)}_{m+1,l+1}$ denote the $((s-1)(p+1)+m+1,(k-1)(p+1)+l+1)$th element of B_n. Then, we have

$$\begin{aligned}
E(b^{(sk)}_{m+1,l+1}) &= \sum_{i=1}^{n} E\{\boldsymbol{C}^T_{ism}(\tilde{I}_i\widehat{V}_i\tilde{I}_i)^{-1}K^{\frac{1}{2}}_{i\boldsymbol{H}}V_0(t)K^{\frac{1}{2}}_{i\boldsymbol{H}}(\tilde{I}_i\widehat{V}_i\tilde{I}_i)^{-1}\boldsymbol{C}_{ikl}\} \\
&= n\sum_{v=1}^{2^J-1} E\{\boldsymbol{C}^T_{1sm}(\tilde{I}_1\widehat{V}_1\tilde{I}_1)^{-1}K^{\frac{1}{2}}_{1\boldsymbol{H}}V_0(t)K^{\frac{1}{2}}_{1\boldsymbol{H}}(\tilde{I}_1\widehat{V}_1\tilde{I}_1)^{-1}\boldsymbol{C}_{1kl}|\mathcal{S}_v(h_{\max})\}\Pr\{\mathcal{S}_v(h_{\max})\} \\
&= n\sum_{v=1}^{2^J-1}\{\sum_{l_1,l_2=1}^{q} E[(t_{1j_v}-t)^{m+l}K^{\frac{1}{2}}_{h_s}(t_{1j_v}-t)K^{\frac{1}{2}}_{h_k}(t_{1j_v}-t) \\
&\quad \times K^{\frac{1}{2}}_{h_{l_1}}(t_{1j_v}-t)K^{\frac{1}{2}}_{h_{l_2}}(t_{1j_v}-t)I\{\mathcal{S}_v(h_{\max})\}]E\{(\tilde{\boldsymbol{e}}_s\otimes\mathbf{1_0})^T(\tilde{I}_{v_0}\widehat{V}_1\tilde{I}_{v_0})^{-1} \\
&\quad \times(E_{l_1}\otimes I_J)V_0(t)(E_{l_2}\otimes I_J)(\tilde{I}_{v_0}\widehat{V}_1\tilde{I}_{v_0})^{-1}(\tilde{\boldsymbol{e}}_k\otimes\mathbf{1_0})|\mathcal{S}_v(0)\}\}\{1+o(1)\} \\
&= n\sum_{v=1}^{2^J-1} f_v(t)h^{m+l+1}_{\max}\{\sum_{l_1,l_2=1}^{q}\bar{\xi}^{(sk)}_{v,l_1l_2}(t)\frac{1}{\sqrt{h_sh_kh_{l_1}h_{l_2}}} \\
&\quad \times\int_{-1}^{1} z^{l+m}K^{\frac{1}{2}}(\frac{h_{\max}}{h_s}z)K^{\frac{1}{2}}(\frac{h_{\max}}{h_k}z)K^{\frac{1}{2}}(\frac{h_{\max}}{h_{l_1}}z)K^{\frac{1}{2}}(\frac{h_{\max}}{h_{l_2}}z)dz\}\{1+o(1)\} \\
&= n\sum_{v=1}^{2^J-1} f_v(t)h^{m+l+1}_{\max}\{\sum_{l_1,l_2=1}^{q}\bar{\xi}^{(sk)}_{v,l_1l_2}(t)\nu_{m+l}(h_s,h_k,h_{l_1},h_{l_2})\}\{1+o(1)\} \\
&= nh^{m+l+1}_{\max}\sum_{v=1}^{2^J-1} f_v(t)\nu^{(sk)}_{m+l,v}(t)\{1+o(1)\}.
\end{aligned}$$

Similar to (A.3), we have

$$B_n = nh_{\max}[(I_q\otimes\widetilde{H})\bar{S}(I_q\otimes\widetilde{H})]\{1+o_P(1)\}. \tag{A.4}$$

By combining (A.1)-(A.4), we have

$$\mathrm{Cov}\{\widehat{\boldsymbol{m}}^{(k)}(t)|\mathcal{F}_n\} = \frac{k!^2}{nh^{1+2k}_{\max}}[(I_q\otimes\boldsymbol{e}^T_{k+1})S^{-1}\bar{S}S^{-1}(I_q\otimes\boldsymbol{e}_{k+1})] + o_P\left(\frac{1}{nh^{1+2k}_{\max}}\right).$$

Similar to the asymptotic expansion of B_n in (A.4), we can show that

$$\begin{aligned}
\text{Bias}\{\widehat{\boldsymbol{m}}^{(k)}(t)\} &= k!(I_q \otimes \boldsymbol{e}_{k+1}^T)[E(\widehat{\text{vec}(\beta)}|\mathcal{F}_n) - \text{vec}(\beta)] \\
&= k!(I_q \otimes \boldsymbol{e}_{k+1}^T)A_n^{-1}\textstyle\sum_{i=1}^n (I_q \otimes X_i^T)W_i E[\text{vec}(Y_i) - (I_q \otimes X_i)\text{vec}(\beta)] \\
&= \tfrac{k!}{(p+1)!}(I_q \otimes \boldsymbol{e}_{k+1}^T)A_n^{-1}\textstyle\sum_{i=1}^n (I_q \otimes X_i^T)W_i \begin{pmatrix} m_1^{(p+1)}(t) \\ \vdots \\ m_q^{(p+1)}(t) \end{pmatrix} \\
&\quad \otimes \begin{pmatrix} (t_{i1} - t)^{p+1} \\ \vdots \\ (t_{iJ} - t)^{p+1} \end{pmatrix} \{1 + o(1)\} \\
&= \tfrac{nk!h_{\max}^{p+2}}{(p+1)!}(I_q \otimes \boldsymbol{e}_{k+1}^T)\{nh_{\max}^{-1}[(I_q \otimes \widetilde{H})S(I_q \otimes \widetilde{H})]\{1 + o_P(1)\}\}^{-1} \\
&\quad \times[(I_q \otimes \widetilde{H})\boldsymbol{D}]\{1 + o_P(1)\} \\
&= \tfrac{k!}{(p+1)!}h_{\max}^{p+1-k}[(I_q \otimes \boldsymbol{e}_{k+1}^T S^{-1}\boldsymbol{D}] + o_P(h_{\max}^{p+1-k}).
\end{aligned}$$

The last equation holds because $\boldsymbol{D} = O(h_{\max}^{-1})$, as specified in (A.1). By now, we have proved the results (10) and (11).

Proof of Corollary 1

Similar to the proof of Proposition 1, we can show that

$$A_n = n[M \otimes \widetilde{H}S_1\widetilde{H}]\{1 + o_P(1)\} \tag{A.5}$$

$$B_n = nh^{-1}[N \otimes \widetilde{H}\bar{S}_1\widetilde{H}]\{1 + o_P(1)\} \tag{A.6}$$

$$\sum_{i=1}^n (I_q \otimes X_i^T)W_i E[\text{vec}(Y_i) - (I_q \otimes X_i)\text{vec}(\beta)] = [\boldsymbol{C} \otimes \widetilde{H}c_p]\{1 + o_P(1)\}. \tag{A.7}$$

The results (12) and (13) can be obtained after combining (A.5)-(A.7).

Proof of Corollary 2

In cases when response components are independent, the covariance matrices V_i are block diagonal. By combining this result with those in (10) and (11) in Proposition 1, the conclusions (14) and (15) are straightforward.

Proof of Corollary 3

The proof of Corollary 3 is similar to the one of Proposition 1. Thus, it is omitted here.

References

Chen K and Jin Z, Local polynomial regression analysis of clustered data. Biometrika 2005; 92: 59-74.

Coull BA and Staudenmayer J. Self-modeling regression for multivariate curve data. Statistica Sinica 2004; 14: 695C711.

Cupples LA et al., The Framingham Heart Study 100K SNP genome-wide association study resource: overview of 17 phenotype working group reports. BMC Medical Genetics 2007; 8 (Suppl 1): S1.

Fieuws S and Verbeke G, Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. Biometrics 2006; 62: 424C31.

Fitzmaurice GM, Laird NM and Ware JH. Applied longitudinal analysis. John-Willey & Sons, New York; 2004.

Gray SM and Brookmeyer R, Multidimensional longitudinal data: estimating treatment effect from continuous, discrete or time-to-event response variable. Journal of the American Statistical Association 2000; 95: 396C406.

Godleski JJ, Verrier RL, Koutrakis P, Catalano P, et al. Mechanisms of morbidity and mortality from exposure to ambient air particles. Research Report, Health Effects Institute, 2000; 91: 5C103.

Hall P and Robinson AP, Reducing variability of cross-validation for smoothing-parameter choice. Biometrika, 2009; 96: 175-186.

Li Yehua, Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. Biometrika 2011; accepted.

Liang KY and Zeger SL, Longitudinal data analysis using generalized linear models. Biometrika 1986; 73: 13C22.

Lin X and Carroll R, Nonparametric function estimation for clustered data when the predictor is measured without/with error. J. Am. Statist. Assoc. 2000; 95: 520-534.

Lin X and Carroll R, Semiparametric regression for clustered data using generalized estimating equations. J. Am. Statist. Assoc. 2001; 96: 1045C1056.

O'Brien LM and Fitzmaurice GM, Analysis of longitudinal multiple-source binary data using generalized estimating equations. Applied Statistician 2004; 53: 177C193.

Ruppert D and Wand MP, Multivariate locally weighted least squares regression. Ann. Statist. 1994; 22: 1346-1370.

Rochon J. Analyzing bivariate repeated measures for discrete and continuous outcome variable. Biometrics 1996; 52: 740C750.

Roy J and Lin X. Latent variable model for longitudinal data with multiple continuous outcomes. Biometrics 2000; 56: 1047C1054.

WANG, N, Marginal nonparametric kernel regression accounting for within-subject correlation. Biometrika 2003; 90: 43C52.

Weiss RE, Modeling Longitudinal data. Springer, USA; 2005.