Ann. Inst. Statist. Math.
Vol. 52, No. 3, 519–543 (2000)

A TWO-STEP SMOOTHING METHOD FOR VARYING-COEFFICIENT MODELS WITH REPEATED MEASUREMENTS

COLIN O. WU[1*], KAI FUN YU[2] AND CHIN-TSANG CHIANG[3]

[1] *Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, U.S.A.*
[2] *Division of Epidemiology, Statistics and Prevention Research, National Institute of Child Health and Human Development, Bethesda, MD 20852, U.S.A.*
[3] *Department of Statistics, Tunghai University, Taichung, Taiwan*

(Received May 6, 1998; revised February 15, 1999)

Abstract. Datasets involving repeated measurements over time are common in medical trials and epidemiological cohort studies. The outcomes and covariates are usually observed from randomly selected subjects, each at a set of possibly unequally spaced time design points. One useful approach for evaluating the effects of covariates is to consider linear models at a specific time, but the coefficients are smooth curves over time. We show that kernel estimators of the coefficients that are based on ordinary local least squares may be subject to large biases when the covariates are time-dependent. As a modification, we propose a two-step kernel method that first centers the covariates and then estimates the curves based on some local least squares criteria and the centered covariates. The practical superiority of the two-step kernel method over the ordinary least squares kernel method is shown through a fetal growth study and simulations. Theoretical properties of both the two-step and ordinary least squares kernel estimators are developed through their large sample mean squared risks.

Key words and phrases: Bandwidth, kernel, longitudinal data, mean squared error, ultrasound measurement, varying coefficient models.

1. Introduction

In many medical and epidemiological studies, such as growth studies, interests of statistical analyses are often focused on determining the relationship between a time dependent real-valued outcome variable $Y(t)$ and a set of random covariates $\boldsymbol{X}(t) = (X_1(t), \dots, X_k(t))^T$. Longitudinal observations of these studies are usually obtained from n randomly selected subjects and, for $i = 1, \dots, n$, the i-th subject has n_i repeated measurements recorded at possibly unequally spaced time points t_{ij}, $j = 1, \dots, n_i$. The j-th measurement of $(t, \boldsymbol{X}(t), Y(t))$ for the i-th subject is then denoted by $(t_{ij}, \boldsymbol{X}_{ij}, Y_{ij})$, where $\boldsymbol{X}_{ij} = (X_{ij1}, \dots, X_{ijk})^T$.

Statistical models and estimation methods for this type of data have been mostly concentrated on parametric approaches based on linear and nonlinear regression models, such as Diggle *et al.* (1994), Davidian and Giltinan (1995), Vonesh and Chinchilli (1997),

* Partial support for the the first author was provided by grant, R01 DA10184-01, from the National Institute on Drug Abuse. This research was carried out when the first author was visiting the Division of Epidemiology, Statistics and Prevention Research, the National Institute of Child Health and Human Development.

among others, or nonparametric models of $(t, Y(t))$ without considering the effects of covariates $X_1(t), \ldots, X_k(t)$, such as Hart and Wehrly (1986), Altman (1990) and Rice and Silverman (1991). As a mixture of parametric and nonparametric models, a semiparametric approach based on a partially linear model has been studied by Zeger and Diggle (1994) and Moyeed and Diggle (1994). However, for many situations, the existing parametric and nonparametric approaches may be either too restrictive to accommodate the unknown shapes of the curves or lacking the specific structures of being biologically interpretable. On the other hand, a general multivariate nonparametric regression model would be too complicated to be biologically interpretable.

As a useful compromise to retain meaningful biological interpretations and flexible nonparametric structures, Hoover *et al.* (1998) considered the varying coefficient model

$$Y(t) = \beta_0(t) + \boldsymbol{X}^T(t)\beta(t) + \epsilon(t), \tag{1.1}$$

where $\beta(t) = (\beta_1(t), \ldots, \beta_k(t))^T$, $\beta_l(t)$, $l = 0, \ldots, k$, are continuous curves of t, $\epsilon(t)$ is a mean zero stochastic process with $E[\epsilon^2(t)] < \infty$ for all t and $\boldsymbol{X}(t)$ and $\epsilon(t)$ are independent, and proposed a class of linear smoothing estimators of $\beta(t)$ based on local ordinary least squares or penalized least squares criteria. When the data are cross-sectional, i.e. $n_i \equiv 1$ for all $i = 1, \ldots, n$, (1.1) reduces to a special case of the models described by Hastie and Tibshirani (1993).

The linear smoothers of Hoover *et al.* (1998), such as their ordinary least squares kernel estimator, suffer two main drawbacks. First, in many situations, the corresponding estimators may be subject to large biases when the covariates are time-dependent so that no adequate estimators of $\beta(t)$ can be obtained for any choice of smoothing parameters (see Section 4). Second, since the ordinary least squares kernel and local polynomial estimators rely on only one set of bandwidth and kernel function, they can not adjust for different smoothing needs of $\beta_r(t)$, $r = 0, \ldots, k$, when they belong to different smoothness families.

A motivating example here is the Alabama Small for Gestational Age Cohort Study (Alabama SGAC Study) conducted by the National Institute of Child Health and Human Development. A main objective for this study is to determine the effects of placental development and other risk factors on fetal development, such as the growth of abdominal circumference, throughout pregnancy. Since no specific relationship between fetal growth, gestational age and other available risk factors has been rigorously justified, the nonparametric model (1.1) is clearly more adequate than parametric models for an initial data exploration. However, because of the above two drawbacks, the ordinary least squares kernel method gives obviously biased estimators which do not have meaningful biological interpretations.

As a remedy, we consider the following equivalent form of (1.1),

$$Y(t) = \beta_0^*(t) + \boldsymbol{Z}^T(t)\beta(t) + \epsilon(t), \tag{1.2}$$

where $\boldsymbol{Z}(t) = (Z_1(t), \ldots, Z_k(t))^T$, $Z_l(t) = X_l(t) - \mu_l(t)$, $\beta_0^*(t) = \beta_0(t) + \sum_{l=1}^k \mu_l(t)\beta_l(t)$ and $\mu_l(t) = E[X_l(t) \mid t]$, and propose a new class of kernel estimators of $\beta_0^*(t)$, $\beta_0(t)$ and $\beta(t)$. Our estimators are constructed based on a two-step procedure which first obtains estimators $\hat{\mu}_l(t_{ij})$ of $\mu_l(t_{ij})$ and then estimates $\beta_0^*(t)$ and $\beta(t)$ based on some local least squares criteria and the centered covariates $(X_{ijl} - \hat{\mu}_l(t_{ij}))$. As a simple generalization, we also incorporate multiple bandwidths and kernels into the estimators, so that adequate smoothing may be provided for all $\beta_l(t)$. The results of this paper have two main features. First, in many situations, particularly when $\mu_l(t)$, for some $l = 1, \ldots, k$, have large slopes at t, our two-step kernel estimators are asymptotically superior over the ordinary least squares kernel estimators in the sense that the two-step kernel estimators have smaller asymptotic mean squared errors. Second, through Monte Carlo simulations and an application to the Alabama SGAC Study, we show that the two-step kernel estimators are often more reliable than the ordinary least squares kernel estimators in practice. We only consider in this paper the practical and theoretical properties of the two-step kernel estimators in order to provide useful insights to the two-step method. Similar approaches can be generalized to two-step procedures based on local polynomials and smoothing splines. But, under the current complex longitudinal design, asymptotic properties of local polynomials and smoothing splines require substantial further development.

For the rest of the paper, we give the expressions of the ordinary least squares and the two-step kernel procedures in Section 2. The finite sample advantages of the two-step kernel method are illustrated through simulation results in Section 3. The application of the two-step kernel method to the Alabama SGAC Study is shown in Section 4. Asymptotic properties of the two-step kernel estimators and their comparisons with that of the ordinary least squares kernel estimators are developed in Section 5. Proofs of the main results are deferred to Section 6.

2. Smoothing methods

2.1 *Least Squares*

2.1.1 *Ordinary least squares*

Suppose that the $(k+1)\times(k+1)$ matrix $E[(1,\boldsymbol{X}^T(t))^T(1,\boldsymbol{X}^T(t))]$ is invertible for every $t \in R$. The ordinary least squares kernel estimators $\hat\beta_0(t;K,h)$ and $\hat\beta(t;K,h)$ of $\beta_0(t)$ and $\beta(t)$, respectively, are obtained by minimizing

$$L(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}[Y_{ij}-b_0(t)-\boldsymbol{X}_{ij}^T b(t)]^2 w_i h^{-1} K\left(\frac{t-t_{ij}}{h}\right), \tag{2.1}$$

with respect to $b_0(t)$ and $b(t)=(b_1(t),\ldots,b_k(t))^T$, where w_i are non-negative weights whose usual choices include $w_i \equiv (\sum_{i=1}^n n_i)^{-1}$ or $w_i \equiv (nn_i)^{-1}$, $K(\cdot)$ is a Borel-measurable kernel function and h is a positive bandwidth. Let

$$\boldsymbol{X}_{*i}=\begin{pmatrix}1 & X_{i11} & \cdots & X_{i1k}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & X_{in_i1} & \cdots & X_{in_ik}\end{pmatrix}, \quad \boldsymbol{Y}_i=\begin{pmatrix}Y_{i1}\\ \vdots\\ Y_{in_i}\end{pmatrix}$$

and

$$\boldsymbol{K}_i(t;h)=\mathrm{diag}\left(w_i h^{-1}K\left(\frac{t-t_{i1}}{h}\right),\ldots,w_i h^{-1}K\left(\frac{t-t_{in_i}}{h}\right)\right).$$

If $(\sum_{i=1}^n \boldsymbol{X}_{*i}^T\boldsymbol{K}_i(t;h)\boldsymbol{X}_{*i})$ is invertible, then the unique minimizer of (2.1) is given by

$$(\hat\beta_0(t;K,h),\hat\beta^T(t;K,h))^T=\left[\sum_{i=1}^n \boldsymbol{X}_{*i}^T\boldsymbol{K}_i(t;h)\boldsymbol{X}_{*i}\right]^{-1}\left[\sum_{i=1}^n \boldsymbol{X}_{*i}^T\boldsymbol{K}_i(t;h)\boldsymbol{Y}_i\right]. \tag{2.2}$$

When $w_i=(\sum_{i=1}^n n_i)^{-1}$, (2.2) is reduced to the ordinary least squares kernel estimator of Hoover *et al.* (1998).

2.1.2 *Two-step least squares*

By centering all the covariates $X_l(t)$ around their corresponding conditional means $\mu_l(t)$, (1.2) is equivalent to (1.1) and $\beta_0^*(t)$ now represents the baseline effect or the mean effect of t on $Y(t)$ when the covariates are equal to their conditional means. Given the relationship between $\beta_0^*(t)$, $\beta_0(t)$, $\beta_l(t)$ and $\mu_l(t)$, it is also natural to estimate the coefficient curves based on (1.2).

Following this approach, one would have to first estimate $\mu_l(t)$ from the data and then estimate $\beta_0^*(t)$ and $\beta(t)$ based on some local least squares criteria. Let $(\Gamma_l(\cdot),\gamma_l)$, $l=1,\ldots,k$, be k sets of Borel-measurable kernel functions and positive bandwidths. Then a kernel estimator $\hat{\mu}_l(t;\Gamma_l,\gamma_l)$ of $\mu_l(t)$ can be obtained by minimizing

$$\ell_l(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\left[X_{ijl}-a_l(t)\right]^2 w_i\gamma_l^{-1}\Gamma_l\left(\frac{t-t_{ij}}{\gamma_l}\right)\right\} \tag{2.3}$$

with respect to $a_l(t)$. It is straightforward to see that the minimizer $\hat{\mu}_l(t;\Gamma_l,\gamma_l)$ of (2.3) is a Nadaraya-Watson type estimator (see Nadaraya (1964), Watson (1964), and Härdle (1990), Chapter 3) given by

$$\hat{\mu}_l\left(t;\Gamma_l,\gamma_l\right)=\frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{X_{ijl}w_i\Gamma_l\left[\left(t-t_{ij}\right)/\gamma_l\right]\right\}}{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{w_i\Gamma_l\left[\left(t-t_{ij}\right)/\gamma_l\right]\right\}}. \tag{2.4}$$

Let $Z_{ijl}=X_{ijl}-\mu_l(t_{ij})$, $\hat{Z}_{ijl}=X_{ijl}-\hat{\mu}_l(t_{ij};\Gamma_l,\gamma_l)$ for $l=1,\ldots,k$, $\boldsymbol{Z}_{ij}=(Z_{ij1},\ldots,Z_{ijk})^T$, $\hat{\boldsymbol{Z}}_{ij}=(\hat{Z}_{ij1},\ldots,\hat{Z}_{ijk})^T$ and

$$\hat{\boldsymbol{Z}}_{*i}=\begin{pmatrix}1 & \hat{Z}_{i11} & \cdots & \hat{Z}_{i1k}\\ \vdots & \vdots & \vdots & \vdots\\ 1 & \hat{Z}_{in_i1} & \cdots & \hat{Z}_{in_ik}\end{pmatrix}.$$

Substituting $\boldsymbol{X}_{ij}$ of (2.1) by $\hat{\boldsymbol{Z}}_{ij}$, one can obtain two-step kernel estimators $\tilde{\beta}_0^*(t;K,h)$ and $\tilde{\beta}(t;K,h)$ of $\beta_0^*(t)$ and $\beta(t)$, respectively, by minimizing

$$\mathcal{L}(t;K,h)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{[Y_{ij}-b_0(t)-\hat{\boldsymbol{Z}}_{ij}^Tb(t)]^2w_ih^{-1}K\left(\frac{t-t_{ij}}{h}\right)\right\} \tag{2.5}$$

with respect to $b_0(t)$ and $b(t)=(b_1(t),\ldots,b_k(t))^T$. When $(\sum_{i=1}^n\hat{\boldsymbol{Z}}_{*i}^T\boldsymbol{K}_i(t;h)\hat{\boldsymbol{Z}}_{*i})$ is invertible, $\tilde{\beta}_0^*(t;K,h)$ and $\tilde{\beta}(t;K,h)$ are uniquely given by

$$\left(\tilde{\beta}_0^*(t;K,h),\tilde{\beta}^T(t;K,h)\right)^T=\left[\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_{*i}^T\boldsymbol{K}_i(t;h)\hat{\boldsymbol{Z}}_{*i}\right]^{-1}\left[\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_{*i}^T\boldsymbol{K}_i(t;h)\boldsymbol{Y}_i\right]. \tag{2.6}$$

Based on $\tilde{\beta}_0^*(t;K,h)$, $\tilde{\beta}(t;K,h)$ and $\hat{\mu}_l(t;\Gamma_l,\gamma_l)$, one can estimate $\beta_0(t)$ of (1.1) by

$$\tilde{\beta}_0(t;K,h)=\tilde{\beta}_0^*(t;K,h)-\left\{\sum_{l=1}^{k}\left[\hat{\mu}_l\left(t;\Gamma_l,\gamma_l\right)\tilde{\beta}_l(t;K,h)\right]\right\}, \tag{2.7}$$

where $\tilde{\beta}_l(t;K,h)$ is the l-th component of $\tilde{\beta}(t;K,h)$. Here, the adequacy of $\tilde{\beta}_0(t;K,h)$ depends on (Γ_l,γ_l) as well as (K,h).

Remark 2.1. In both (1.1) and (1.2), $\beta_l(t)$ with $l \geq 1$ can be interpreted as the average change of $Y(t)$ at time t that is caused by the unit change of $X_l(t)$. Although $\beta_0(t)$ may be thought as certain "baseline" curve, it does not have a meaningful physical interpretation when the values of $X_l(t)$, for some $l = 1, \ldots, k$, can not be zero. On the other hand, $\beta_0^*(t)$ can always be physically interpreted. Thus, for many practical situations, it is frequently more advantageous to consider (1.2) and the estimation of $\beta_0^*(t)$ and $\beta(t)$ instead of the estimation based on (1.1).

2.1.3 *Modified estimators*

The estimators given in (2.6) rely on a single set of kernel and bandwidth, i.e. (K, h), to estimate $\beta_0^*(t)$ and all k curves in $\beta(t)$. In most practical situations, particularly when $\beta_0^*(t)$, $\beta_1(t), \ldots, \beta_k(t)$ belong to different smoothness families, a single (K, h) may not simultaneously provide adequate smoothing for all the estimated curves. As a natural generalization of (2.6), $(\beta_0^*(t), \beta^T(t))^T$ can also be estimated by

$$\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h}) = (\tilde{\beta}_0^*(t; K_0, h_0), \tilde{\beta}_1(t; K_1, h_1), \ldots, \tilde{\beta}_k(t; K_k, h_k))^T, \tag{2.8}$$

where $\boldsymbol{K} = (K_0, \ldots, K_k)$ and $\boldsymbol{h} = (h_0, \ldots, h_k)$ are sequences of kernels and bandwidths. Similarly, by generalizing (2.2), $(\beta_0(t), \beta^T(t))^T$ can also be estimated by

$$\hat{\theta}(t; \boldsymbol{K}, \boldsymbol{h}) = (\hat{\beta}_0(t; K_0, h_0), \hat{\beta}_1(t; K_1, h_1), \ldots, \hat{\beta}_k(t; K_k, h_k))^T, \tag{2.9}$$

where $\hat{\beta}_l(t; K, h)$ is the l-th component of $\hat{\beta}(t; K, h)$.

Remark 2.2. Similar to the approaches of Hoover *et al.* (1998), two-step local polynomial estimators can be obtained by substituting $b_0(t)$ and $b_l(t)$ of (2.5) with their corresponding Taylor expansion terms. Although the current complex model and data structures are quite different from the traditional nonparametric regression with independent cross-sectional data, it is conceivable that certain boundary advantages of the local polynomials over the kernel estimators may still remain. A comprehensive study of the practical and asymptotic properties of two-step local polynomials requires substantial theoretical development, hence is omitted from this paper.

2.2 *Bandwidth choices*

Similar to kernel smoothing with independent cross-sectional data, the shapes of kernels are usually less important than the sizes of bandwidths. Although subjective bandwidths may be selected by examining the fitted curves, it is helpful in practice to have a procedure that suggests appropriate automatic bandwidths based on the available data. Because of the possible intra-correlations within each subject, the usual data-driven bandwidth selection methods that are suitable for the independent cross-sectional data may not be directly applied to the current longitudinal data. Here an appropriate bandwidth choice may be associated with the structures of the intra-correlations; see, for example, Altman (1990). But, for most applications, the intra-correlation structures are completely unknown. Then a simple and natural automatic bandwidth procedure is the "leave one subject out" cross-validation suggested by Rice and Silverman (1991).

Applying the procedure of Rice and Silverman (1991) to the two-step smoothing estimators, we would have to first compute the cross-validated values for γ_l. Let $\hat{\mu}_l^{(-i)}(t; \Gamma_l, \gamma_l)$ be the estimator of $\mu_l(t)$ using (2.4) with the observations of the i-th subject left out. Suppose that we would like to select γ_l to minimize the the average prediction squared error of $\hat{\mu}_l(t;\Gamma_l,\gamma_l)$. The cross-validation bandwidth $\gamma_{l,\text{cv}}$ then minimizes

$$\text{CV}(\gamma_l) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{w_i[X_{ijl} - \hat{\mu}_l^{(-i)}(t_{ij};\Gamma_l,\gamma_l)]^2\} \tag{2.10}$$

with respect to γ_l.

Let $\tilde{\theta}^{(-i)}(t;\boldsymbol{K},\boldsymbol{h})$ be the estimator of $(\beta_0^*(t),\beta^T(t))^T$ based on (2.8), $(\gamma_{1,\text{cv}},\ldots,\gamma_{k,\text{cv}})$ and the remaining data with the observations of the i-th subject left out. We define

$$\text{CV}(\boldsymbol{h}) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{w_i[Y_{ij} - (1,\hat{\boldsymbol{Z}}_{ij}^T)\tilde{\theta}^{(-i)}(t_{ij};\boldsymbol{K},\boldsymbol{h})]^2\} \tag{2.11}$$

to be the cross-validation score for $\tilde{\theta}^{(-i)}(t;\boldsymbol{K},\boldsymbol{h})$. Then the cross-validation bandwidth vector $\boldsymbol{h}_{\text{cv}} = (h_{0,\text{cv}},\ldots,h_{k,\text{cv}})$ is a minimizer of $\text{CV}(\boldsymbol{h})$.

Remark 2.3. By minimizing (2.11), we approximately minimize the average prediction squared error

$$\text{APSE}(\tilde{\theta}(\cdot;\boldsymbol{K},\boldsymbol{h})) = \sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i E\left\{\left[Y_{ij}^* - \tilde{\beta}_0^*(t_{ij};K_0,h_0) - \sum_{l=1}^{k}(Z_{ijl}\tilde{\beta}_l(t_{ij};K_l,h_l))\right]^2\right\}$$

with respect to $\boldsymbol{h} = (h_0,\ldots,h_k)^T$, where Y_{ij}^* is a new observation at $(t_{ij},\boldsymbol{X}_{ij})$. To give a heuristic reason for this, we can consider the following decomposition

$$\begin{aligned}\text{CV}(\boldsymbol{h}) = & \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{w_i(Y_{ij} - \beta_0^*(t_{ij}) - \hat{\boldsymbol{Z}}_{ij}^T\beta(t_{ij}))^2\} \\ & + \sum_{i=1}^{n}\sum_{j=1}^{n_i}\{w_i[(1,\hat{\boldsymbol{Z}}_{ij}^T)((\beta_0^*(t_{ij}),\beta^T(t_{ij}))^T - \tilde{\theta}^{(-i)}(t_{ij};\boldsymbol{K},\boldsymbol{h}))]^2\} \\ & + 2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\{w_i(Y_{ij} - \beta_0^*(t_{ij}) - \hat{\boldsymbol{Z}}_{ij}^T\beta(t_{ij})) \\ & \qquad \times [(1,\hat{\boldsymbol{Z}}_{ij}^T)((\beta_0^*(t_{ij}),\beta^T(t_{ij}))^T - \tilde{\theta}^{(-i)}(t_{ij};\boldsymbol{K},\boldsymbol{h}))]\}.\end{aligned} \tag{2.12}$$

The first term of the right side of (2.12) does not depend on the bandwidths. By the definitions of $\tilde{\theta}^{(-i)}(t;\boldsymbol{K},\boldsymbol{h})$ and $\hat{\boldsymbol{Z}}_{ij}$, we can show that the expectation of the third term of the right side of (2.12) is negligible when n is sufficiently large. Thus, $\boldsymbol{h}_{\text{cv}}$ approximately minimizes the second term of the right side of (2.12). By a straightforward comparison between $\text{APSE}(\tilde{\theta}(\cdot;\boldsymbol{K},\boldsymbol{h}))$ and the second term of the right side of (2.12), we can show that $\boldsymbol{h}_{\text{cv}}$ also approximately minimizes $\text{APSE}(\tilde{\theta}(\cdot;\boldsymbol{K},\boldsymbol{h}))$. Without the presence of the covariates $\boldsymbol{X}(t)$, the consistency of a similar "leave-one-subject-out" cross-validation bandwidth has been established by Hart and Wehrly (1993) for the estimation of $E[Y(t) \mid t]$. But rigorous theoretical properties and efficient algorithms of $\boldsymbol{h}_{\text{cv}}$ under the current varying coefficient models have not been developed and deserve further study.

Remark 2.4. A systematic and rigorous search of $\boldsymbol{h}_{\text{cv}}$ may be time consuming and may require sophisticated optimization software. However, instead of searching for the global minima, it is frequently reasonable in practice to use those bandwidths that approximately minimize the cross-validation scores. Such approximate cross-validation bandwidths will be used in the simulation study of Section 3 and again in the Alabama SGAC Study of Section 4.

3. Monte Carlo simulations

We consider model (1.1) with $k = 2$ and $\beta_0(t) = 3.5 + 6.5\sin(t\pi/60)$,

$$\beta_1(t) = 2.5 - 0.0074\left(\frac{30-t}{10}\right)^3 \quad \text{and} \quad \beta_2(t) = -0.2 - 1.6\cos\left(\frac{(t-30)\pi}{60}\right).$$

For each t, $X_1(t)$ is a time-dependent random variable from the Gaussian distribution with mean $3\exp[t/30]$ and variance 1, and X_2 is a time-independent Bernoulli random variable with probability 0.5 for being 1 or 0. Each longitudinal sample has 400 randomly generated subjects. The cross-sectional observations X_{i2}, $i = 1, \ldots, 400$, are independently generated from the distribution of X_2. Each subject has a probability of 0.4 to be observed at time points $0, 1, 2, \ldots, 30$. The observed time points form t_{ij}, $i = 1, \ldots, 400$, $j = 1, \ldots, n_i$, which are unequally spaced. Based on each t_{ij}, independent samples X_{ij1} and $\epsilon_{ij} = \epsilon_i(t_{ij})$ are generated, respectively, from the distribution of $X_1(t)$ and the mean zero Gaussian stationary process with covariance

$$\text{cov}(\epsilon_{i_1}(t), \epsilon_{i_2}(s)) = \begin{cases} 0.0625\exp(-|t-s|), & \text{if } \ i_1 = i_2, \\ 0, & \text{if } \ i_1 \neq i_2. \end{cases}$$

The simulated outcome values Y_{ij} are obtained by substituting t_{ij}, X_{ij1}, X_{i2}, ϵ_{ij} and the above $(\beta_0(t_{ij}), \beta_1(t_{ij}), \beta_2(t_{ij}))$ into (1.1).

This simulation process was replicated 200 times. For each simulated longitudinal sample, kernel estimators of $\beta_0(t)$, $\beta_1(t)$ and $\beta_2(t)$ were computed using both the ordinary least squares and the two-step methods. Since different kernel choices, such as the Epanechnikov kernel (see Härdle (1990), Chapter 3) and the standard Gaussian kernel, gave very similar results, the estimated values presented here are only for the standard Gaussian kernel. Following the bandwidth procedure of Section 2.2, we computed the cross-validation bandwidths for each simulated dataset. For the purpose of illustration and comparison, we also used several other bandwidths which had cross-validation scores very close to those given by the cross-validation bandwidths. For simulations with the ordinary least squares kernel estimators, the cross-validated bandwidths $h_{0,\text{cv}}$ and $h_{1,\text{cv}}$ were around 0.3 and the values of $h_{2,\text{cv}}$ were in the range of 0.4 to 0.6. For simulations with the two-step kernel estimators, the cross-validated bandwidths $\gamma_{1,\text{cv}}$ for the estimation of $\mu_1(t)$ were around 0.7, $h_{0,\text{cv}}$ and $h_{1,\text{cv}}$ were around 0.4, and $h_{2,\text{cv}}$ were in the range of 0.4 to 0.6.

Figures 1a and 2a show the real $\beta_0(t)$ and $\beta_1(t)$ in solid curves, and the averages over 200 simulations of the ordinary least squares kernel estimators computed using the cross-validated bandwidths and $(h_0, h_1, h_2) = (0.5, 0.5, 0.7)$ in dashed and dotted curves, respectively. Similarly, in Figs. 1b and 2b, the solid curves represent the real $\beta_0(t)$ and $\beta_1(t)$, while the dashed and dotted curves show their corresponding averages over 200 simulations of the two-step kernel estimators computed using the cross-validated bandwidths and $(\gamma_1, h_0, h_1, h_2) = (0.7, 0.6, 0.6, 0.8)$, respectively. To give an indication of

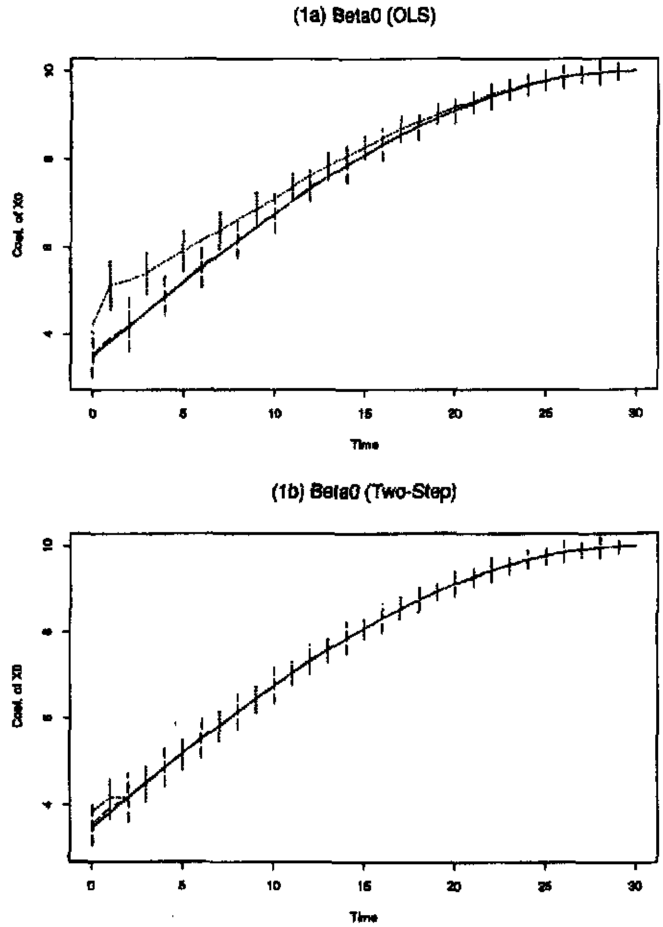

Fig. 1. The solid curve gives the actual $\beta_0(t)$ and the vertical bars give the ± 2 standard errors of the estimates at the corresponding time points. (1a): the averages of the ordinary least squares (OLS) kernel estimators of $\beta_0(t)$ base on the standard Gaussian kernel, the cross-validated bandwidths (dashed curves) and the subjective bandwidths $(h_0, h_1, h_2)= (0.5, 0.5, 0.7)$ (dotted curves); (1b): the averages of the two-step kernel estimators of $\beta_0(t)$ based on the standard Gaussian kernel, the cross-validated bandwidths (dashed curves) and the subjective bandwidths $(\gamma_1, h_0, h_1, h_2)= (0.7, 0.6, 0.6, 0.8)$ (dotted curves).

the variability of these estimates, the vertical bars in the figures represent the pointwise ± 2 standard errors of the 200 simulation estimates at the corresponding time points. Both the ordinary least squares and the two-step kernel methods give estimates very close to the true $\beta_2(t)$ curve, hence, their plots are omitted. The cross-validation bandwidths appear to give slightly undersmoothed estimators for both the ordinary least squares and the two-step kernel estimators. These figures also show that a slight increase of the bandwidths causes significant upward shifts of the ordinary least squares kernel estimators, but the same increase of the bandwidths only affects the smoothness of the two-step kernel estimators.

Although nonparametric inference procedures have yet not been systematically investigated for the current model and data structure, a potentially useful bootstrap method, suggested by Hoover *et al.* (1998), is to resample with replacement the entire repeated measurements of subjects. To assess the validity of the estimators, we computed the 0.95 pointwise bootstrap percentile confidence intervals for $\beta_1(t)$ and $\beta_2(t)$

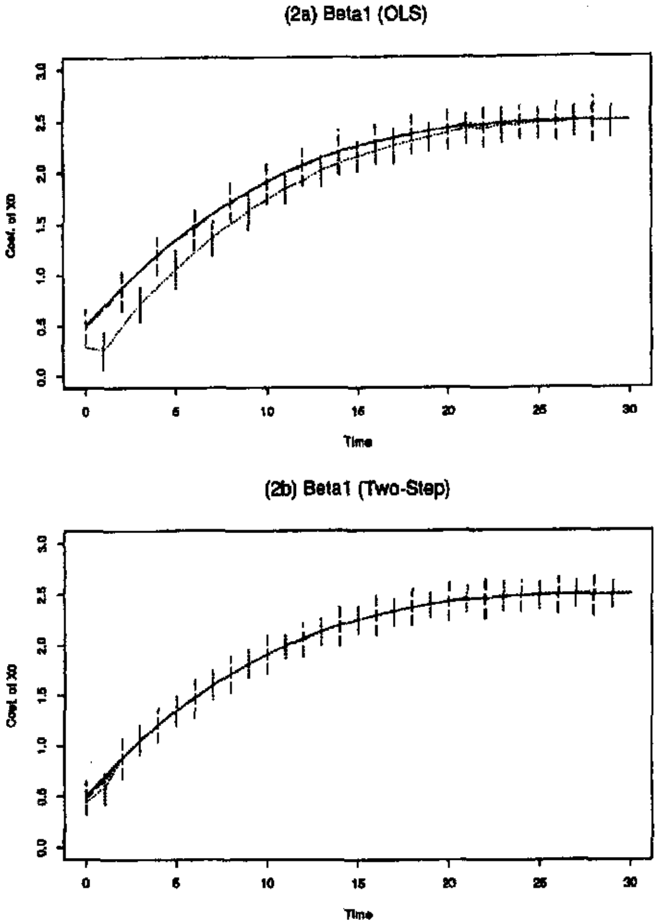

Fig. 2. The solid curve gives the actual $\beta_1(t)$ and the vertical bars give the ± 2 standard errors of the estimates at the corresponding time points. (2a): the averages of the ordinary least squares (OLS) kernel estimators of $\beta_1(t)$ base on the standard Gaussian kernel, the cross-validated bandwidths (dashed curves) and the subjective bandwidths $(h_0, h_1, h_2) = (0.5, 0.5, 0.7)$ (dotted curves); (2b): the averages of the two-step kernel estimators of $\beta_1(t)$ based on the standard Gaussian kernel, the cross-validated bandwidths (dashed curves) and the subjective bandwidths $(\gamma_1, h_0, h_1, h_2) = (0.7, 0.6, 0.6, 0.8)$ (dotted curves).

Table 1. The estimated coverage probabilities of the 0.95 bootstrap pointwise confidence intervals for $\beta_1(t)$ and $\beta_2(t)$ based on the ordinary least squares kernel estimators with the standard Gaussian kernel, $\boldsymbol{h}_{cv}$ and $\boldsymbol{h} = (0.5, 0.5, 0.7)$.

Time point	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
$\hat{\beta}_1(t; K, \boldsymbol{h}_{cv})$	0.95	0.96	0.94	0.93	0.94	0.95	0.92	0.92	0.94
$\hat{\beta}_1(t; K, \boldsymbol{h})$	0.04	0.15	0.37	0.54	0.73	0.85	0.85	0.90	0.92
$\hat{\beta}_2(t; K, \boldsymbol{h}_{cv})$	0.92	0.92	0.92	0.95	0.94	0.93	0.94	0.94	0.93
$\hat{\beta}_2(t; K, \boldsymbol{h})$	0.96	0.91	0.91	0.93	0.94	0.92	0.92	0.93	0.91

at nine time points. These confidence intervals were computed based on 200 bootstrap replications. Based on the 200 simulations, Table 1 shows the estimated coverage probabilities of the ordinary least squares kernel estimators with the cross-validated bandwidths and $(h_0, h_1, h_2) = (0.5, 0.5, 0.7)$. Similarly, Table 2 shows the coverage

Table 2. The estimated coverage probabilities of the 0.95 bootstrap pointwise confidence intervals for $\beta_1(t)$ and $\beta_2(t)$ based on the two-step kernel estimators with the standard Gaussian kernel, $\boldsymbol{h}_{\rm cv}$ and $\boldsymbol{h} = (0.7, 0.6, 0.6, 0.8)$.

Time point	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
$\tilde\beta_1(t; K, \boldsymbol{h}_{\rm cv})$	0.95	0.96	0.93	0.92	0.94	0.95	0.92	0.92	0.93
$\tilde\beta_1(t; K, \boldsymbol{h})$	0.95	0.94	0.92	0.93	0.91	0.93	0.91	0.92	0.94
$\tilde\beta_2(t; K, \boldsymbol{h}_{\rm cv})$	0.94	0.91	0.92	0.97	0.93	0.93	0.94	0.92	0.93
$\tilde\beta_2(t; K, \boldsymbol{h})$	0.94	0.90	0.93	0.94	0.92	0.93	0.92	0.95	0.93

probabilities of the two-step kernel estimators with the cross-validated bandwidths and $(\gamma_1, h_0, h_1, h_2) = (0.7, 0.6, 0.6, 0.8)$. Although the coverage probabilities of the ordinary least squares kernel estimators appear to be acceptable for the cross-validation bandwidths, they are very sensitive to the bandwidth choices and a slight change in bandwidth choices may lead to totally unacceptable coverage probabilities; see, for example, the coverage probabilities of $\hat\beta_1(t; K, \boldsymbol{h})$ in Table 1. On the other hand, the coverage probabilities of the two-step kernel estimators are more stable and generally acceptable. Similar conclusions can also be obtained for the estimation of $\beta_0(t)$.

4. Application to the Alabama SGAC study

The data are from a prospective study of risk factors for intrauterine growth retardation. All 1475 women enrolled in the study were scheduled to have their fetal anthropometry measurements carried out by ultrasound at approximately 17, 25, 31 and 36 weeks of gestation. However, this schedule was not closely followed. The actual visits were scattered between 12 and 43 weeks of gestation, which led to unbalanced repeated measurements. Since normal fetal growth is naturally thought to be associated with proper childhood development, a main objective of this study is to assess the role of the risk factors or covariates that might affect fetal growth. Although many maternal behavioral risk factors, such as smoking, alcohol use and drug abuse, etc., may have significant influence on the fetal development, for the purpose of demonstration, our analysis here is focused on the effects of the placental development, measured by placental thickness over time, on the development of fetal abdominal circumference. Since mother's height is also an obvious factor that may be positively correlated with fetal size, this covariate is also included in the analysis. Biomedical and epidemiological implications of the statistical results here are also interesting but deserve further study and need to be independently verified from other data. So these implications will not be addressed in this paper.

Let $Y(t)$ be the fetal abdominal circumference at t weeks of gestation, $X_1(t)$ be the placental thickness at t and X_2 be the mother's height. Under model (1.2), $\beta_0^*(t)$ represents the mean curve of fetal abdominal circumference when the woman has average placental thickness at t weeks of gestation and is of average height, and $\beta_1(t)$ and $\beta_2(t)$ represent the unit change in fetal abdominal circumference associated with the unit change in placental thickness at t and the unit change in mother's height, respectively.

Fitting model (1.2) to the data, Figs. 3a and 4a show the two-step kernel estimators of $\beta_0^*(t)$ and $\beta_1(t)$ based on the standard Gaussian kernel and the cross-validated bandwidths $\gamma_{1,\rm cv} = 1.5$, $h_{l,\rm cv} = 0.3$, $l = 0, 1, 2$. The dashed curves show the 95% bootstrap percentile pointwise confidence intervals computed by resampling the subjects with replacement and 200 bootstrap replications. Because the cross-validation score corresponding to the bandwidths $\gamma_1 = 1.5$, $h_0 = 1.0$, $h_1 = 2.0$ and $h_2 = 1.0$ is very close

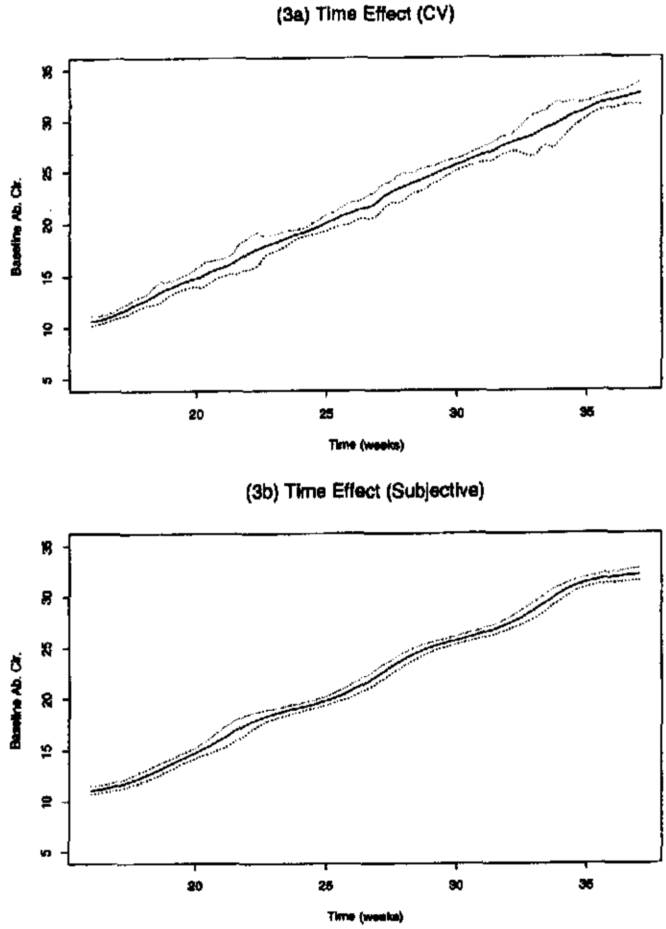

Fig. 3. The solid curve gives the two-step kernel estimator of the baseline time effect on the fetal abdominal circumferences. The dashed curves give the corresponding 0.95 bootstrap pointwise confidence intervals. The cross-validated bandwidths (1.5, 0.3, 0.3, 0.3) were used in (3a). The bandwidths (1.5, 1.0, 2.0, 1.0) were used in (3b).

to the cross-validation score corresponding to $(\gamma_{1,\text{cv}}, h_{0,\text{cv}}, h_{1,\text{cv}}, h_{2,\text{cv}})$, Figs. 3b and 4b show the two-step kernel estimators of $\beta_0^*(t)$ and $\beta_1(t)$, respectively, computed based on the standard Gaussian kernel and bandwidths $\gamma_1 = 1.5$, $h_0 = 1.0$, $h_1 = 2.0$ and $h_2 = 1.0$. These estimators are smoother and perhaps have a better biological interpretation than those given in Figs. 3a and 4a. The estimates for $\beta_2(t)$ stay very close to 0 across the entire time range, hence, are omitted from the presentation.

Qualitatively, we can see from these figures that placental thickness appears to be positively associated with fetal abdominal circumference. This positive association appears to be significant for the period roughly between 22 and 27 weeks of gestation and levelling off at either the beginning or the end of pregnancy. On the other hand, mother's height shows no significant effect on the growth of fetal abdominal circumference.

We also computed the estimators of $\beta_0(t)$, $\beta_1(t)$ and $\beta_2(t)$ based on model (1.1) and the ordinary least squares kernel estimators with cross-validated and a range of other bandwidths. These estimators are less smooth than the two-step kernel estimators for the cross-validated bandwidths and have significant upward shifts when the bandwidths are slightly increased. Thus, the results based on the ordinary least squares method are not presented in this paper.

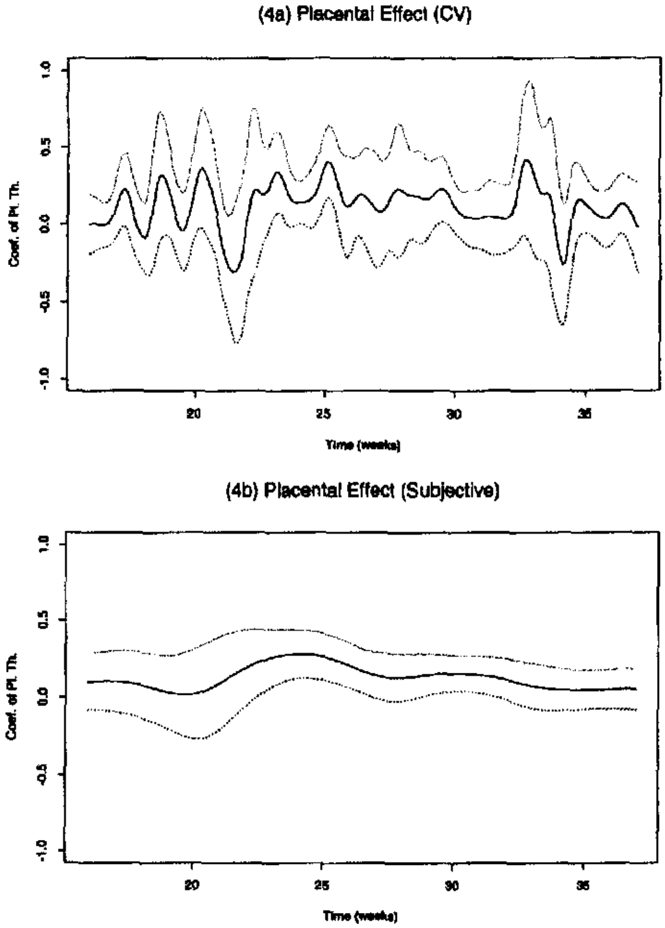

Fig. 4. The solid curve gives the two-step kernel estimator of the placental effect on the fetal abdominal circumferences. The dashed curves give the corresponding 0.95 bootstrap pointwise confidence intervals. The cross-validated bandwidths $(1.5, 0.3, 0.3, 0.3)$ were used in (4a). The bandwidths $(1.5, 1.0, 2.0, 1.0)$ were used in (4b).

5. Asymptotic properties

5.1 *Notation and assumptions*

Biomedical data frequently are unbalanced in the sense that not all subjects are observed at the same time points. We assume here that the time points t_{ij}, with $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$, are randomly drawn from an unknown cumulative distribution function $F(\cdot)$ with density $f(\cdot)$. With a slightly different set of notation, our calculations here can be modified to accommodate fixed and balanced designs.

The following regularity conditions are assumed throughout this section:

(A1) The kernel functions $K_r(\cdot)$ and $\Gamma_l(\cdot)$ with $0 \le r \le k$ and $1 \le l \le k$ are bounded, symmetric about the origin and compactly supported on the real line and satisfy $\int K_r(u)du = \int \Gamma_l(u)du = 1$.

(A2) The bandwidths h_r and γ_l and the weight functions w_i satisfy $\sum_{i=1}^n n_i w_i = 1$ and, as $n \to \infty$, $h_r \to 0$, $\gamma_l \to 0$, $n_i w_i h_r^{-2} \to 0$ and $n_i w_i \gamma_l^{-2} \to 0$ for all $r = 0, \ldots, k$ and $l = 1, \ldots, k$.

(A3) For $r = 1, \ldots, k$, $X_r(t)$ are compactly supported stochastic processes on the real line and are independent of $\epsilon(t)$. Define

$$\sigma^2(s) = E[\epsilon^2(s)], \quad \tau(s_1, s_2) = E[\epsilon(s_1)\epsilon(s_2)], \quad \tau(s) = \lim_{c\to 0} E[\epsilon(s)\epsilon(s+c)],$$
$$\rho_{r,l}(s) = \mathrm{cov}(X_r(s), X_l(s)), \quad \rho_{r,l}(s_1, s_2) = \mathrm{cov}(X_r(s_1), X_l(s_2))$$

and

$$\rho_{r,l}(s,s) = \lim_{c\to 0} \mathrm{cov}(X_r(s), X_l(s+c)).$$

For all r, $l = 0, \ldots, k$ and $p = 1, \ldots, k$, $\beta_r(s)$, $\rho_{r,l}(s)$, $f(s)$ and $\mu_p(s)$ are continuous at t and belong to the same smoothness family; for example, $\beta_r(s)$, $\rho_{r,l}(s)$, $f(s)$ and $\mu_p(s)$ are Lipschitz continuous with order $\alpha > 0$ at t.

Some of the above regularity conditions, such as the smoothness assumptions of $\beta_r(s)$, $\rho_{r,l}(s)$, $f(s)$ and $\mu_p(s)$ and the compactness of the supports of $X_r(t)$, $K_r(\cdot)$ and $\Gamma_l(\cdot)$, are merely assumed for mathematical simplicity and can be weakened if necessary. The application potential of $\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ may not be restricted by these technical conditions.

Similar to the situations with independent cross-sectional data, the moments of the components of $\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ may not exist, e.g. Rosenblatt (1969). To avoid this minor technical inconvenience, we consider the following approximation

$$\Delta(t; \boldsymbol{K}, \boldsymbol{h}) = (I + \boldsymbol{o}_p(1))\,[\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h}) - (\beta_0^*(t), \beta^T(t))^T] = \sum_{v=0}^{k} e_{v+1} \Delta(t; K_v, h_v), \tag{5.1}$$

whose proof is given in Section 6, where I is the $(k+1) \times (k+1)$ identity matrix, $\boldsymbol{o}_p(1)$ is the $(k+1) \times (k+1)$ matrix whose elements converge to zero in probability as $n \to \infty$, e_{v+1} is the $(k+1) \times (k+1)$ matrix with 1 at its $(v+1, v+1)$-th entry and 0 elsewhere,

$$\Delta(t; K, h) = (f(t))^{-1} E^{-1}[\boldsymbol{Z}_*(t) \boldsymbol{Z}_*^T(t)] R(t; K, h), \tag{5.2}$$
$$\boldsymbol{Z}_*(t) = (1, \boldsymbol{Z}^T(t))^T, \quad R(t; K, h) = (R_0(t; K, h), \ldots, R_k(t; K, h))^T,$$

$$R_r(t; K, h) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \left[Y_{ij} - \beta_0(t) - \sum_{l=1}^{k} \left(\left(Z_{ijl} - \frac{\delta_l(t_{ij})}{f(t_{ij})} \right) \beta_l(t) \right) \right] \times \left(Z_{ijr} - \frac{\delta_r(t_{ij})}{f(t_{ij})} \right) \frac{w_i}{h} K\left(\frac{t - t_{ij}}{h} \right) \right\} \tag{5.3}$$

and, for $l = 1, \ldots, k$,

$$\delta_l(t) = \sum_{i'=1}^{n} \sum_{j'=1}^{n_{i'}} [X_{i'j'l} - \mu_l(t)] \frac{w_{i'}}{\gamma_l} \Gamma_l \left(\frac{t - t_{i'j'}}{\gamma_l} \right). \tag{5.4}$$

For convenience, we define $Z_{ij0} = X_{ij0} = 1$ and $\hat{\mu}_0(t) = \mu_0(t) = \delta_0(t) = 0$.

Because a vector of smooth curves are estimated, several risk configurations of $\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ may be consider under the current context. Here, we consider a natural mean squared risk of $\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ defined by

$$\begin{aligned} \mathrm{MSE}_{\boldsymbol{p}}\{\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})\} &= E[\Delta^T(t; \boldsymbol{K}, \boldsymbol{h}) \boldsymbol{p} \Delta(t; \boldsymbol{K}, \boldsymbol{h})] \\ &= \sum_{l=0}^{k} p_l \{E^2[\Delta_l(t; K_l, h_l)] + \mathrm{var}[\Delta_l(t; K_l, h_l)]\}, \end{aligned} \tag{5.5}$$

where $\boldsymbol{p} = \mathrm{diag}(p_0, \ldots, p_k)$, p_l are non-negative known constants and $\Delta_l(t; K, h)$ is the $(l+1)$-th element of $\Delta(t; K, h)$.

By the same method in the derivation of (5.1), it is straightforward to show that

$$\Lambda(t; \boldsymbol{K}, \boldsymbol{h}) = (I + \boldsymbol{o}_p(1))[\hat{\theta}(t; \boldsymbol{K}, \boldsymbol{h}) - (\beta_0(t), \beta^T(t))^T] = \sum_{v=0}^{k} e_{v+1}\Lambda(t; K_v, h_v), \tag{5.6}$$

where I, $\boldsymbol{o}_p(1)$ and e_{v+1} are defined in (5.1),

$$\Lambda(t; K, h) = (f(t))^{-1} E^{-1}[\boldsymbol{X}_*(t)\boldsymbol{X}_*^T(t)]Q(t; K, h), \tag{5.7}$$

$\boldsymbol{X}_*(t) = (1, \boldsymbol{X}^T(t))^T$, and $Q(t; K, h)$ is a $k+1$ column vector given by

$$Q(t; K, h) = \sum_{i=1}^{n}\{w_i \boldsymbol{X}_{*i}^T \boldsymbol{K}_i(t; h)[\boldsymbol{Y}_i - \boldsymbol{X}_{*i}(\beta_0(t), \beta^T(t))^T]\}. \tag{5.8}$$

The mean squared risk of $\hat{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ for the estimation of $(\beta_0(t), \beta^T(t))^T$ can be defined by

$$\begin{aligned} \mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t; K, h)] &= E[\Lambda^T(t; \boldsymbol{K}, \boldsymbol{h})\boldsymbol{p}\Lambda(t; \boldsymbol{K}, \boldsymbol{h})] \\ &= \sum_{l=0}^{k} p_l[(E\Lambda_l(t; K_l, h_l))^2 + \mathrm{var}(\Lambda_l(t; K_l, h_l))], \end{aligned} \tag{5.9}$$

where $\Lambda_l(t; K, h)$ is the $(l+1)$-th element of $\Lambda(t; K, h)$.

5.2 *Risk representations of $\tilde{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$ and $\hat{\theta}(t; \boldsymbol{K}, \boldsymbol{h})$*

We now summarize the main asymptotic results of this section. The proofs of these results are given in Section 6.

Let $\tilde{\theta}(t; K, h)$ and $\hat{\theta}(t; K, h)$ be the estimators defined in (2.8) and (2.9), respectively, with $(\boldsymbol{K}, \boldsymbol{h})$ replaced by a single set of kernel and bandwidth (K, h). By (5.5), the mean squared risk of $\tilde{\theta}(t; K, h)$ is

$$\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t; K, h)] = \sum_{r=0}^{k}\sum_{p=0}^{k}\{M_{r,p}^*(t; \boldsymbol{p})E[R_r(t; K, h)R_p(t; K, h)]\}, \tag{5.10}$$

where $M_{r,p}^*(t; \boldsymbol{p})$ is the $(r+1, p+1)$-th element of the $(k+1) \times (k+1)$ matrix

$$M^*(t; \boldsymbol{p}) = (f(t))^{-2}\{E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)]\}^T \boldsymbol{p}\{E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)]\}.$$

Similarly, (5.9) implies that the mean squared risk of $\hat{\theta}(t; K, h)$ is

$$\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t; K, h)] = \sum_{r=0}^{k}\sum_{p=0}^{k}\{M_{r,p}(t; \boldsymbol{p})E[Q_r(t; K, h)Q_p(t; K, h)]\}, \tag{5.11}$$

where $M_{r,p}(t; \boldsymbol{p})$ is the $(r+1, p+1)$-th element of

$$M(t; \boldsymbol{p}) = (f(t))^{-2}\{E^{-1}[\boldsymbol{X}_*(t)\boldsymbol{X}_*^T(t)]\}^T \boldsymbol{p}\{E^{-1}[\boldsymbol{X}_*(t)\boldsymbol{X}_*^T(t)]\}.$$

For the mean squared risks of the more general estimators $\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})$ and $\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})$, we have

$$\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})] = \sum_{l=0}^{k} p_l\{\mathrm{MSE}_{u_{(l+1)}}[\tilde{\theta}(t;K_l,h_l)]\}$$

and

$$\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})] = \sum_{l=0}^{k} p_l\{\mathrm{MSE}_{u_{(l+1)}}[\hat{\theta}(t;K_l,h_l)]\},$$

where u_r is the $(k+1)\times(k+1)$ diagonal matrix with 1 at the r-th diagonal place and 0 elsewhere. Thus, the asymptotic representations of $\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})]$ and $\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})]$ are expressed through the asymptotic covariances of $R_r(t;K,h)$ and $Q_r(t;K,h)$.

For $r,p=0,\ldots,k$, let $II^*_{r,p}(t)=\rho_{r,p}(t)\sigma^2(t)f(t)[\int K^2(u)du]$, $III^*_{r,p}(t)=\rho_{r,p}(t,t)\tau(t)(f(t))^2$ and $II_{r,p}(t)$ and $III_{r,p}(t)$ be defined as $II^*_{r,p}(t)$ and $III^*_{r,p}(t)$ with $\rho_{r,p}(t)$ and $\rho_{r,p}(t,t)$ replaced by $\rho_{r,p}(t)-\mu_r(t)\mu_p(t)$ and $\rho_{r,p}(t,t)-\mu_r(t)\mu_p(t)$, respectively.

THEOREM 5.1. *Suppose that assumptions* (A1), (A2) *and* (A3) *are satisfied, t is an interior point of the support of $f(\cdot)$ and, for all $l=1,\ldots,k$, $\gamma_l/h=O(1)$ as $n\to\infty$. Then*

$$E[R_0(t;K,h)] = \int[\beta_0(t-hu)-\beta_0(t)]K(u)f(t-hu)du + \sum_{l=1}^{k}\left\{\beta_l(t)\int[\mu_l(t-\gamma_l u)-\mu_l(t)]\Gamma_l(u)f(t-\gamma_l u)du\right\}, \tag{5.12}$$

for $1\le r\le k$,

$$E[R_r(t;K,h)] = \sum_{l=1}^{k}\left\{\int \rho_{r,l}(t-hu)[\beta_l(t-hu)-\beta_l(t)]K(u)f(t-hu)du\right\} + \left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right)\sum_{l=1}^{k}[\beta_l(t)\rho_{r,l}(t,t)f(t)] + o\left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right), \tag{5.13}$$

and, for all $r,p=0,\ldots,k$,

$$\begin{aligned} &E[R_p(t;K,h)R_r(t;K,h)] \\ &= [E(R_p(t;K,h))][E(R_r(t;K,h))] + \left(\sum_{i=1}^{n} n_i w_i^2 h^{-1}\right) II^*_{r,p}(t) \\ &+ \left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right) III^*_{r,p}(t) + o\left(\sum_{i=1}^{n} n_i w_i^2 h^{-1}\right) \\ &+ o\left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right). \end{aligned} \tag{5.14}$$

THEOREM 5.2. *Suppose that the assumptions of Lemma* 5.1 *are satisfied. Let* $Q_r(\cdot)$ *be the* $(r+1)$*-th element of* $Q(\cdot)$. *Then, for all* r, $p = 0, \ldots, k$,

$$(5.15)\qquad E[Q_r(t;K,h)] = \sum_{l=0}^{k} \int (\rho_{r,l}(t-hu) - \mu_r(t-hu)\mu_l(t-hu)) \times (\beta_l(t-hu) - \beta_l(t)) f(t-hu) K(u) du$$

and $E[Q_r(t;K,h)Q_p(t;K,h)]$ *equals the right side of* (5.14) *with* $R_r(t;K,h)$, $R_p(t;K,h)$, $II^*_{r,p}(t)$ *and* $III^*_{r,p}(t)$ *replaced by* $Q_r(t;K,h)$, $Q_p(t;K,h)$, $II_{r,p}(t,t)$ *and* $III_{r,p}(t)$, *respectively.*

Remark 5.1. Define an estimator to be consistent if, for all choices of $\boldsymbol{p}$, its corresponding mean squared risks (5.5) or (5.9) converge to zero as $n \to \infty$. Since assumption (A2) implies that $\sum_{i=1}^{n}(n_i w_i^2 h^{-1}) \to 0$ and $\sum_{i=1}^{n}[n_i(n_i-1)w_i^2] \to 0$ as $n \to \infty$, it is easy to derive from Theorems 5.1 and 5.2 that, under assumptions (A1) through (A3), $\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})$ and $\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})$ are consistent estimators of $(\beta_0^*(t), \beta^T(t))^T$ and $(\beta_0(t), \beta^T(t))^T$, respectively.

The general asymptotic expressions of Theorems 5.1 and 5.2 are applicable, for example, for the families of Lipschitz continuous $\beta_r(t)$, $\rho_{r,l}(t)$, $f(t)$ and $\mu_p(t)$ for r, $l = 0, \ldots, k$ and $p = 1, \ldots, k$. However, when further smoothness conditions are considered, the asymptotic expressions of $\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})$ and $\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})$ can be further specified. To see the asymptotic risks for twice differentiable families, let $a'(t)$ and $a''(t)$ denote the first and second derivatives of $a(t)$ at t,

$$I_0^*(t) = \sum_{l=1}^{k} \left\{ \beta_l(t) \left(\frac{\gamma_l}{h}\right)^2 \left(\frac{1}{2}\mu_l''(t)f(t) + \mu_l'(t)f'(t)\right) \left(\int u^2 \Gamma_l(u) du\right) \right\} + \left(\frac{1}{2}\beta_0''(t)f(t) + \beta'(t)f'(t)\right)\left(\int u^2 K(u)du\right) + o(1),$$

$$I_r^*(t) = \sum_{l=1}^{k} \left\{ \left[\frac{1}{2}\beta_l''(t)\rho_{r,l}(t)f(t) + \beta_l'(t)\rho_{r,l}'(t)f(t) + \beta_l'(t)\rho_{r,l}(t)f'(t)\right] \times \left(\int u^2 K(u)du\right)\right\} + o(1), \quad \text{for} \quad r = 1, \ldots, k,$$

$$I_q(t) = \sum_{l=0}^{k} \left\{ \left[\frac{1}{2}\beta_l''(t)f(t)\left(\rho_{q,l}(t) + \mu_q(t)\mu_l(t)\right) + \beta_l'(t)f'(t)\left(\rho_{q,l}(t) + \mu_q(t)\mu_l(t)\right) + \beta_l'(t)f(t)\left(\rho_{q,l}'(t) + \mu_q'(t)\mu_l(t) + \mu_l'(t)\mu_q(t)\right)\right] \left(\int u^2 K(u)du\right)\right\} + o(1),$$

for $q = 0, \ldots, k$.

COROLLARY 5.1. *Suppose that the assumptions of Theorems* 5.1 *and* 5.2 *are satisfied and* $\beta_r(t)$, $\rho_{r,l}(t)$, $f(t)$ *and* $\mu_p(t)$ *for all* r, $l = 0, \ldots, k$ *and* $p = 1, \ldots, k$ *are twice continuously differentiable at* t. *The mean squared risk of* $\tilde{\theta}(t;K,h)$ *is*

$$(5.16)\quad \mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;K,h)] = \sum_{r=0}^{k}\sum_{p=0}^{k}\left\{ M_{r,p}^*(t;\boldsymbol{p}) \left[h^4 I_r^*(t) I_p^*(t) + \left(\sum_{i=1}^{n} n_i w_i^2 h^{-1}\right) II_{r,p}^*(t) + \left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right) III_{r,p}^*(t)\right](1+o(1))\right\}.$$

Similarly, the mean squared risk, $\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t;K,h)]$, *of* $\hat{\theta}(t;K,h)$ *is given by* (5.16) *with* $M^*_{r,p}(t;\boldsymbol{p})$, $I^*_r(t)$, $II^*_{r,p}(t)$ *and* $III^*_{r,p}(t)$ *replaced by* $M_{r,p}(t;\boldsymbol{p})$, $I_r(t)$, $II_{r,p}(t)$ *and* $III_{r,p}(t)$, *respectively.*

PROOF. By the Taylor expansions of $\beta_r(t)$, $\rho_{r,l}(t)$, $f(t)$ and $\mu_p(t)$, $I^*_r(t)$ and $I_r(t)$ follow from (5.13) and (5.15), respectively, and assumption (A2). Then the mean squared risks are direct consequences of Theorems 5.1 and 5.2. □

Remark 5.2. Notice that, since $\tilde{\beta}^*_0(t;K_0,h_0)$ and $\hat{\beta}_0(t;K_0,h_0)$ are estimating different curves, $\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;\boldsymbol{K},\boldsymbol{h})]$ may not be compared to $\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t;\boldsymbol{K},\boldsymbol{h})]$ unless $p_0=0$. The asymptotic risks of $\tilde{\beta}_l(t;K_l,h_l)$ and $\hat{\beta}_l(t;K_l,h_l)$, $l=1,\ldots,k$, can be directly compared by taking $p_l>0$ and all other entries of $\boldsymbol{p}$ to be zero. When the estimation of $\beta_0(t)$ is of interest, one may compare the asymptotic risk of $\tilde{\beta}_0(t;K_0,h_0)$ of (2.7) with that of $\hat{\beta}_0(t;K_0,h_0)$. Here, the asymptotic mean squared risk of $\tilde{\beta}_0(t;K_0,h_0)$ can be derived by straightforward but tedious computations similar to those used in the derivations of Theorems 5.1 and 5.2 and Corollary 5.1. For clarity and simplicity, the exact asymptotic risk representations of $\tilde{\beta}_0(t;K_0,h_0)$ are omitted from this paper.

Remark 5.3. The rates of $\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;K,h)]$ and $\mathrm{MSE}_{\boldsymbol{p}}[\hat{\theta}(t;K,h)]$ converging to zero also strongly depend on the choices of w_i, $i=1,\ldots,n$. If $w_i=1/(nn_i)$, (5.16) is reduced to

$$\text{(5.17)}\quad \mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;K,h)]=\sum_{r=0}^{k}\sum_{p=0}^{k}\left\{M^*_{r,p}(t;\boldsymbol{p})\left[h^4I^*_r(t)I^*_p(t)+h^{-1}\left(\sum_{i=1}^{n}\frac{1}{n^2n_i}\right)II^*_{r,p}(t)+\left(n^{-1}-n^{-2}\sum_{i=1}^{n}\frac{1}{n_i}\right)III^*_{r,p}(t)\right](1+o(1))\right\}.$$

When n_i, $i=1,\ldots,n$, are bounded, the best convergence rate of (5.17) is $n^{4/5}$, i.e. $\mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;K,h)]=O(n^{-4/5})$, which is attained by taking $h=O(n^{-1/5})$. If $w_i=1/\sum_{i=1}^{n}n_i$, (5.16) is reduced to

$$\text{(5.18)}\quad \mathrm{MSE}_{\boldsymbol{p}}[\tilde{\theta}(t;K,h)]=\sum_{r=0}^{k}\sum_{p=0}^{k}\left\{M^*_{r,p}(t;\boldsymbol{p})\left[h^4I^*_r(t)I^*_p(t)+h^{-1}\left(\sum_{i=1}^{n}n_i\right)^{-1}II^*_{r,p}(t)+\left(\frac{\sum_{i=1}^{n}n_i^2}{(\sum_{i=1}^{n}n_i)^2}-\frac{1}{\sum_{i=1}^{n}n_i}\right)III^*_{r,p}(t)\right](1+o(1))\right\}.$$

Again, when n_i, $i=1,\ldots,n$, are bounded, the best convergence rate of (5.18) is $n^{4/5}$ which is also attained by $h=O(n^{-1/5})$. It is interesting to note that, if $\max_{1\le i\le n}(n_i/\sum_{i=1}^{n}n_i)$ does not converge to zero as $n\to\infty$, then the assumption of $\lim_{n\to\infty}n_iw_ih^{-2}=0$ of (A2) fails and $(\sum_{i=1}^{n}n_i^2/(\sum_{i=1}^{n}n_i)^2)$ does not converge to zero as $n\to\infty$ (see, Theorem 1 of Hoover *et al.* (1998)), hence $\tilde{\theta}(t;K,h)$ is inconsistent.

5.3 *Comparison of two smoothing methods*

Because (5.10) and (5.11) involve nonlinear transformations of $E[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)]$ and $E[\boldsymbol{X}_*(t)\ \boldsymbol{X}_*^T(t)]$ which are difficult to express explicitly when $k \geq 2$, it is difficult to give a general comparison between the asymptotic risks of Theorem 5.1 and Theorem 5.2. We consider here the special case of $k = 1$, i.e. $\boldsymbol{X}_*(t) = (1, X(t))^T$ with $X(t)$ on the real line, and compare the asymptotic risks of $\tilde{\beta}_1(t; K, h)$ and $\hat{\beta}_1(t; K, h)$. For simplicity, we assume that $\rho_{1,1}(t) = \rho_{1,1}(t, t)$.

Since only the estimation of $\beta_1(t)$ is of interest, we take $\boldsymbol{p} = \mathrm{diag}(0, 1)$. Direct calculation shows that $E[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)] = \mathrm{diag}(1, \rho_{1,1}(t))$ and $E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)] = \mathrm{diag}(1, \rho_{1,1}^{-1}(t))$. Then, by (5.10), the mean squared risk of $\tilde{\beta}_1(t; K, h)$ is

$$\mathrm{MSE}[\tilde{\beta}_1(t; K, h)] = [f(t)\rho_{1,1}(t)]^{-2} E[R_1^2(t; K, h)],$$

where

$$(5.19)\quad E[R_1^2(t; K, h)] = E^2[R_1(t; K, h)] + \left(\sum_{i=1}^n n_i w_i^2 h^{-1}\right) II_{1,1}^*(t) + \left(\sum_{i=1}^n n_i(n_i - 1)w_i^2\right) III_{1,1}^*(t) + o\left(\sum_{n=1}^n n_i w_i^2 h^{-1}\right) + o\left(\sum_{i=1}^n n_i(n_i - 1)w_i^2\right)$$

and

$$(5.20)\quad E[R_1(t; K, h)] = \int [\beta_1(t - hu) - \beta_1(t)]\rho_{1,1}(t - hu)K(u)f(t - hu)du + \left(\sum_{i=1}^n n_i(n_i - 1)w_i^2\right) \beta_1(t)\rho_{1,1}(t)f(t) + o\left(\sum_{i=1}^n n_i(n_i - 1)w_i^2\right).$$

It can be shown by similar calculations that the mean squared risk of $\hat{\beta}_1(t; K, h)$ is

$$\mathrm{MSE}[\hat{\beta}_1(t; K, h)] = [f(t)\rho_{1,1}(t)]^{-2} E[(Q_1(t; K, h) - \mu_1(t)Q_0(t; K, h))^2],$$

where $E[(Q_1(t; K, h) - \mu_1(t)Q_0(t; K, h))^2]$ equals the right side of (5.19) with $E^2[R_1(t; K, h)]$ replaced by $E^2[Q_1(t; K, h) - \mu_1(t)Q_0(t; K, h)]$ and

$$(5.21)\quad E[Q_1(t; K, h) - \mu_1(t)Q_0(t; K, h)] = \int [\beta_1(t - hu) - \beta_1(t)]\, \rho_{1,1}(t - hu)f(t - hu)K(u)du + \int [\beta_1(t - hu) - \beta_1(t)][\mu_1(t - hu) - \mu_1(t)]\mu_1(t - hu)f(t - hu)K(u)du + \int [\beta_0(t - hu) - \beta_0(t)][\mu_1(t - hu) - \mu_1(t)]f(t - hu)K(u)du.$$

Thus, *it suffices to compare (5.20) with (5.21).*

The major difference between (5.20) and (5.21) is that the convergence rate of (5.21) not only depends on the smoothness of $\beta_1(t)$ but also is affected by the smoothness of $\beta_0(t)$ and $\mu_1(t)$. If $w_i = 1/(nn_i)$ or $1/(\sum_{i=1}^n n_i)$, $\beta_1(t)$ and $\rho_{1,1}(t)$ are twice differentiable but $\beta_0(t)$ and $\mu_1(t)$ are only assumed to be continuous at t, then $\tilde{\beta}_1(t;K,h)$ will have a better convergence rate than that of $\hat{\beta}_1(t;K,h)$ in the sense that $\mathrm{MSE}[\tilde{\beta}_1(t;K,h)]$ will be dominated by $\mathrm{MSE}[\hat{\beta}_1(t;K,h)]$. If $w_i = 1/(nn_i)$ or $1/(\sum_{i=1}^n n_i)$ and $\beta_1(t)$, $\rho_{1,1}(t)$, $\beta_0(t)$ and $\mu_1(t)$ are all twice continuously differentiable at t, then the asymptotic mean squared risk of $\tilde{\beta}_1(t;K,h)$ will be smaller than that of $\hat{\beta}_1(t;K,h)$ when $\beta_0(t)$ and/or $\mu_1(t)$ have large first derivatives at t. Comparisons under other smoothness conditions can be similarly discussed.

6. Proofs

We start with a useful technical lemma. For simplicity, we denote throughout this section that $\hat{\mu}_l(t) = \hat{\mu}_l(t;\Gamma_l,\gamma_l)$.

LEMMA 6.1. *If assumptions* (A1), (A2) *and* (A3) *are satisfied and* t *is an interior point of the support of* $f(\cdot)$, *then, for* $\delta_l(t)$ *and* $\hat{\mu}_l(t)$ *defined in* (5.4) *and* (2.4), *respectively, and* $l = 1,\dots,k$,

$$
\begin{aligned}
(6.1)\qquad E[\delta_l^2(t)] &= \left[\int (\mu_l(t-\gamma_l u) - \mu_l(t))\Gamma_l(u) f(t-\gamma_l u)\,du\right]^2 (1+o(1)) \\
&\quad + \left(\sum_{i=1}^n n_i w_i^2 \gamma_l^{-1}\right) \rho_{l,l}(t) f(t) \left[\int \Gamma_l^2(u)\,du\right] (1+o(1)) \\
&\quad + \left[\sum_{i=1}^n (n_i(n_i-1)w_i^2)\right] \rho_{l,l}(t,t) f^2(t)(1+o(1))
\end{aligned}
$$

and

$$
(6.2)\qquad [\hat{\mu}_l(t) - \mu_l(t)](1+o_p(1)) = f^{-1}(t)\delta_l(t).
$$

PROOF. By (5.4), we have the straightforward expansion

$$
\begin{aligned}
(6.3)\qquad E[\delta_l^2(t)] = E\Bigg\{ &\sum_{i=1}^n \sum_{j=1}^{n_i} (X_{ijl} - \mu_l(t))^2 \left(\frac{w_i}{\gamma_l}\right)^2 \Gamma_l^2\left(\frac{t-t_{ij}}{\gamma_l}\right) \\
&+ \sum_{i=1}^n \sum_{j_1\neq j_2} (X_{ij_1l} - \mu_l(t))(X_{ij_2l} - \mu_l(t)) \left(\frac{w_i}{\gamma_l}\right)^2 \\
&\times \Gamma_l\left(\frac{t-t_{ij_1}}{\gamma_l}\right) \Gamma_l\left(\frac{t-t_{ij_2}}{\gamma_l}\right) \\
&+ \sum_{i_1\neq i_2} \sum_{j_1,j_2} \Bigg[(X_{i_1j_1l} - \mu_l(t))(X_{j_2j_2l} - \mu_l(t)) \left(\frac{w_{i_1} w_{i_2}}{\gamma_l^2}\right) \\
&\times \Gamma_l\left(\frac{t-t_{i_1j_1}}{b_l}\right) \Gamma_l\left(\frac{t-t_{i_2j_2}}{\gamma_l}\right)\Bigg]\Bigg\}.
\end{aligned}
$$

For the first term at the right side of (6.3), we have, by direct calculation and the change of variables, that

$$
E\left[\sum_{i=1}^n \sum_{j=1}^{n_i} (X_{ijl} - \mu_l(t))^2 \left(\frac{w_i}{\gamma_l}\right)^2 \Gamma_l^2\left(\frac{t-t_{ij}}{\gamma_l}\right)\right]
$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\int E[(X_{ijl}-\mu_l(t))^2 \mid t_{ij}=s]\left(\frac{w_i}{\gamma_l}\right)^2 \Gamma_l^2\left(\frac{t-s}{\gamma_l}\right) f(s)ds\right\}$$

$$= \left(\sum_{i=1}^{n} n_i w_i^2 \gamma_l^{-1}\right)\rho_{l,l}(t)f(t)\left[\int \Gamma_l^2(u)du\right](1+o(1)),$$

while, for the second term,

$$E\left[\sum_{i=1}^{n}\sum_{j_1\neq j_2}(X_{ij_1l}-\mu_l(t))(X_{ij_2l}-\mu_l(t))\left(\frac{w_i}{\gamma_l}\right)^2\Gamma_l\left(\frac{t-t_{ij_1}}{\gamma_l}\right)\Gamma_l\left(\frac{t-t_{ij_2}}{\gamma_l}\right)\right]$$

$$= \sum_{i=1}^{n}\sum_{j_1\neq j_2}\iint\left\{E[(X_{ij_1l}-\mu_l(t))(X_{ij_2l}-\mu_l(t)) \mid t_{ij_1}=s_1, t_{ij_2}=s_2]\left(\frac{w_i}{\gamma_l}\right)^2\right.$$

$$\left.\times\Gamma_l\left(\frac{t-s_1}{\gamma_l}\right)\Gamma_l\left(\frac{t-s_2}{\gamma_l}\right)f(s_1)f(s_2)\right\}ds_1ds_2$$

$$= \left[\sum_{i=1}^{n}(n_i(n_i-1)w_i^2)\right]\rho_{l,l}(t,t)f^2(t)(1+o(1)).$$

Since the subjects are independent, it can be shown by direct calculation under assumptions (A1) through (A3) that

$$E\left[\sum_{i_1\neq i_2}\sum_{j_1,j_2}(X_{i_1j_1l}-\mu_l(t))(X_{i_2j_2l}-\mu_l(t))\left(\frac{w_{i_1}w_{i_2}}{\gamma_l^2}\right)\Gamma_l\left(\frac{t-t_{i_1j_1}}{\gamma_l}\right)\Gamma_l\left(\frac{t-t_{i_2j_2}}{\gamma_l}\right)\right]$$

$$= \left[1-\sum_{i=1}^{n}(n_i^2w_i^2)\right]\left[\int(\mu_l(t-\gamma_l u)-\mu_l(t))\Gamma_l(u)f(t-\gamma_l u)du\right]^2.$$

By $\sum_{i=1}^{n} n_i w_i = 1$, we have that, as $n\to\infty$, $\sum_{i=1}^{n} n_i^2 w_i^2 \to 0$. Thus, (6.1) follows from (6.3) and the above equalities.

To show (6.2), we first notice that

$$(6.4)\qquad (\hat{\mu}_l(t)-\mu(t))(f(t))^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(\frac{w_i}{\gamma_l}\right)\Gamma_l\left(\frac{t-t_{ij}}{\gamma_l}\right)\right] = (f(t))^{-1}\delta_l(t),$$

and, by similar calculations as in the proof of (6.1),

$$(6.5)\qquad (f(t))^{-1}\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(\frac{w_i}{\gamma_l}\right)\Gamma_l\left(\frac{t-t_{ij}}{\gamma_l}\right)\right] = 1+o_p(1).$$

Then, (6.2) follows from (6.4) and (6.5). □

Proof of (5.1). By (2.6) and straightforward algebra, we have

$$(6.6)\quad [\tilde{\theta}(t;K,h)-(\beta_0^*(t),\beta^T(t))^T]\left[\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_{*i}^T\boldsymbol{K}_i(t;h)\hat{\boldsymbol{Z}}_{*i}\right]E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)](f(t))^{-1}$$

$$= (f(t))^{-1}E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)]\left\{\sum_{i=1}^{n}[\hat{\boldsymbol{Z}}_{*i}^T\boldsymbol{K}_i(t;h)(\boldsymbol{Y}_i-\hat{\boldsymbol{Z}}_{*i}(\beta_0^*(t),\beta^T(t))^T)]\right\}.$$

By Lemma 6.1, we can show that

$$\left(\sum_{i=1}^{n}\hat{\boldsymbol{Z}}_{*i}^{T}\boldsymbol{K}_i(t;h)\hat{\boldsymbol{Z}}_{*i}\right)E^{-1}[\boldsymbol{Z}_*(t)\boldsymbol{Z}_*^T(t)](f(t))^{-1}=1+o_p(1). \tag{6.7}$$

Notice that the $(r+1)$-th element of $\sum_{i=1}^{n}[\hat{\boldsymbol{Z}}_{*i}^{T}\boldsymbol{K}_i(t;h)(\boldsymbol{Y}_i-\hat{\boldsymbol{Z}}_{*i}(\beta_0^*(t),\beta^T(t))^T)]$ is

$$\hat{R}_r(t;K,h)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\hat{Z}_{ijr}\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right)\left[Y_{ij}-\beta_0^*(t)-\sum_{l=1}^{k}\hat{Z}_{ijl}\beta_l(t)\right]\right\}.$$

By (6.2) and the definition of $\hat{Z}_{ijr}$, we have $\hat{R}_r(t;K,h)=R_r(t;K,h)(1+o_p(1))$, where $R_r(t;K,h)$ is defined in (5.3). Then, (5.1) follows from (2.8), (6.6) and (6.7). □

PROOF OF THEOREM 5.1. By (5.3) and straightforward algebra, we have

$$E[R_r(t;K,h)]=\sum_{l=1}^{6}E[A_{lr}(t)], \tag{6.8}$$

where

$$A_{1r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=0}^{k}Z_{ijr}Z_{ijl}[\beta_l(t_{ij})-\beta_l(t)]\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right),$$

$$A_{2r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}Z_{ijr}\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right)\epsilon_{ij},$$

$$A_{3r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=1}^{k}(f(t_{ij}))^{-1}Z_{ijr}\delta_l(t_{ij})\beta_l(t_{ij})\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right),$$

$$A_{4r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left(\frac{w_i}{h}\right)(f(t_{ij}))^{-1}\epsilon_{ij}\delta_r(t_{ij})K\left(\frac{t-t_{ij}}{h}\right),$$

$$A_{5r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=0}^{k}(f(t_{ij}))^{-1}Z_{ijl}\delta_r(t_{ij})[\beta_l(t_{ij})-\beta_l(t)]\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right),$$

$$A_{6r}(t)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=1}^{k}(f(t_{ij}))^{-2}\delta_r(t_{ij})\delta_l(t_{ij})\beta_l(t)\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right).$$

We now compute each term of the right side of (6.8). First, consider the case of $r\ge 1$. By $Z_{ij0}=1$, $E[Z_{ijr}]=E[\epsilon_{ij}]=0$ for $r\ge 1$, the independence between X_{ijr} and ϵ_{ij} and the change of variables, we have

$$\begin{aligned}E[A_{1r}(t)]&=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=0}^{k}\int E(Z_{ijr}Z_{ijl}\mid t_{ij}=s)(\beta_l(s)-\beta_l(t))\left(\frac{w_i}{h}\right)K\left(\frac{t-s}{h}\right)f(s)ds\\&=\sum_{l=1}^{k}\int\rho_{r,l}(t-hu)(\beta_l(t-hu)-\beta_l(t))K(u)f(t-hu)du,\end{aligned}$$

$$E[A_{2r}(t)] = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{w_i \int E(Z_{ijr}\epsilon_{ij} \mid t_{ij}=s)h^{-1}K\left(\frac{t-s}{h}\right)f(s)ds\right\}$$
$$= 0,$$
$$E[A_{4r}(t)] = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{w_i \int E(\delta_r(t_{ij})\epsilon_{ij} \mid t_{ij}=s)h^{-1}K\left(\frac{t-s}{h}\right)ds\right\}$$
$$= 0.$$

Since, by assumption (A3), $\beta_r(t)$ and $\mu_p(t)$ belong to the same smoothness family, it is easy to show using the Cauchy-Schwarz's inequality that

$$|E[A_{5r}(t)]| = o(|E[A_{3r}(t)]|) \quad \text{and} \quad |E[A_{6r}(t)]| = o(|E[A_{1r}(t)]|).$$

It remains to compute $E[A_{3r}(t)]$. Since

$$E[A_{3r}(t)] = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\sum_{l=1}^{k}\left\{w_i \int E(Z_{ijr}\delta_l(t_{ij}) \mid t_{ij}=s)\beta_l(t)h^{-1}K\left(\frac{t-s}{h}\right)ds\right\}$$

and, by direct calculation,

$$E[Z_{ijr}\delta_l(t_{ij}) \mid t_{ij}=s] = \frac{w_i}{\gamma_l}c(s) + \sum_{(i',j')\neq(i,j)}\{w_{i'}[T_{r,l,1}(s)+T_{r,l,2}(s)]\},$$

where $c(s)$ is some bounded function on the real line,

$$T_{r,l,1}(s) = E\left[(X_{ijr}-\mu_r(t_{ij}))(X_{i'j'l}-\mu_l(t_{i'j'}))\gamma_l^{-1}\Gamma_l\left(\frac{t_{i'j'}-t_{ij}}{\gamma_l}\right)\,\middle|\, t_{ij}=s\right],$$
$$T_{r,l,2}(s) = E\left[(X_{ijr}-\mu_r(t_{ij}))(\mu_l(t_{i'j'})-\mu_l(t_{ij}))\gamma_l^{-1}\Gamma_l\left(\frac{t_{i'j'}-t_{ij}}{\gamma_l}\right)\,\middle|\, t_{ij}=s\right].$$

If $i \neq i'$, it follows from $E[(X_{ijr}-\mu_r(t_{ij}))(X_{i'j'l}-\mu_l(t_{i'j'}))] = 0$ that $T_{r,l,1}(s)=0$. If $i=i'$ and $j\neq j'$, it follows from direct integration and the change of variables that

$$T_{r,l,1}(s) = \int \rho_{r,l}(s, s-\gamma_l u)\Gamma_l(u)f(s-\gamma_l u)du.$$

By $E[X_{ijr}-\mu_r(t_{ij}) \mid t_{ij}, t_{i'j'}] = 0$, we also have $T_{r,l,2}(s)=0$. Thus, substituting $T_{r,l,1}(s)$ into $E(Z_{ijr}\delta_l(t_{ij}) \mid t_{ij}=s)$ and noting that, by assumption (A2), $\sum_{i=1}^{n}(n_i w_i^2/\gamma_l)$ is negligible relative to $E[A_{lr}(t)]$, we have

$$E[A_{3r}(t)] = \left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right)\sum_{l=1}^{k}\{[\beta_l(t)\rho_{r,l}(t,t)f(t)](1+o(1))\} + o(E[A_{lr}(t)]).$$

Then, (5.13) follows from substituting the above expressions of $E[A_{lr}(t)]$, $l=1,\ldots,6$, into (6.8).

When $r = 0$, we can show by similar calculations as above that

$$E[A_{10}(t)] = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{w_i \int (\beta_0(s) - \beta_0(t)) \left(\frac{w_i}{h}\right) K\left(\frac{t-s}{h}\right) f(s)ds\right\}$$
$$= \int (\beta_0(t-hu) - \beta_0(t))K(u)f(t-hu)du,$$
$$E\left[A_{30}(t)\right] = \sum_{l=1}^{k} \int E[\delta_l(t_{ij}) \mid t_{ij} = s]\beta_l(t)h^{-1}K\left(\frac{t-s}{h}\right)$$
$$= \sum_{l=1}^{k}\left\{\beta_l(t)\int [\mu_l(t-\gamma_l u) - \mu_l(t)]\Gamma_l(u)f(t-\gamma_l u)du\right\}$$

and $E[A_{l0}(t)] = 0$ for $l = 2, 4, 5, 6$. Thus, (5.12) follows.

To compute the right hand side of (5.14), we first notice that

$$(6.9)\quad R_p(t;K,h)R_r(t;K,h)$$
$$= \left[\sum_{i=1}^{n}\sum_{j=1}^{n_i} a_{ijr}(t)\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right)\right]\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i} a_{ijp}(t)\left(\frac{w_i}{h}\right)K\left(\frac{t-t_{ij}}{h}\right)\right]$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{n_i} a_{ijr}(t)a_{ijp}(t)\left(\frac{w_i}{h}\right)^2 K^2\left(\frac{t-t_{ij}}{h}\right)$$
$$+ \sum_{i=1}^{n}\sum_{j\neq j'} a_{ijr}(t)a_{ij'p}(t)\left(\frac{w_i}{h}\right)^2 K\left(\frac{t-t_{ij}}{h}\right)K\left(\frac{t-t_{ij'}}{h}\right)$$
$$+ \sum_{i\neq i'}\sum_{j,j'} a_{ijr}(t)a_{i'j'p}(t)\left(\frac{w_iw_{i'}}{h^2}\right)K\left(\frac{t-t_{ij}}{h}\right)K\left(\frac{t-t_{i'j'}}{h}\right),$$

where $a_{ijr}(t) = \{Y_{ij} - \beta_0(t) - \sum_{l=1}^{k}[(Z_{ijl} - (f(t_{ij}))^{-1}\delta_l(t_{ij}))\beta_l(t_{ij})]\}[Z_{ijr} - (f(t_{ij}))^{-1}\delta_l(t_{ij})]$. Direct integrations as those in the proof of (5.1) show that

$$(6.10)\qquad E\left[\sum_{i=1}^{n}\sum_{j=1}^{n_i} a_{ijr}(t)a_{ijp}(t)\left(\frac{w_i}{h}\right)^2 K^2\left(\frac{t-t_{ij}}{h}\right)\right]$$
$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i^2 h^{-1}\right)\rho_{r,p}(t)\sigma^2(t)f(t)\left[\int K^2(u)du\right](1+o(1)),$$

$$(6.11)\qquad \sum_{i=1}^{n}\sum_{j,j'}\left[a_{ijr}(t)a_{ij'p}(t)\left(\frac{w_i}{h}\right)^2 K\left(\frac{t-t_{ij}}{h}\right)K\left(\frac{t-t_{ij'}}{h}\right)\right]$$
$$= \left(\sum_{i=1}^{n} w_i^2(n_i-1)n_i\right)\rho_{r,p}(t,t)\tau(t)(f(t))^2,$$

and

$$(6.12)\qquad \sum_{i\neq i'}\sum_{j,j'}\left[a_{ijr}(t)a_{i'j'p}(t)\left(\frac{w_iw_{i'}}{h^2}\right)K\left(\frac{t-t_{i'j'}}{h}\right)K\left(\frac{t-t_{i'j'}}{h}\right)\right]$$
$$= [E(R_p(t;K,h))][E(R_r(t;K,h))](1+o(1)).$$

Thus, (5.14) is a direct consequence of (6.9) through (6.12).

PROOF OF THEOREM 5.2. By (5.8), the $(r+1)$-th component of $Q(t;K,h)$ is

$$Q_r(t;K,h) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\left(\frac{w_i}{h}\right) b_{ijr}(t) K\left(\frac{t-t_{ij}}{h}\right)\right\}, \tag{6.13}$$

where

$$b_{ijr}(t) = \sum_{l=0}^{k}\{X_{ijr}X_{ijl}[\beta_l(t_{ij}) - \beta_l(t)]\} + X_{ijr}\epsilon_{ij}.$$

Then, (5.15) follows from $E[b_{ijr}(t) \mid t_{ij} = s] = \sum_{l=0}^{k}\{E(X_r(s)X_l(s))[\beta_l(s) - \beta_l(t)]\}$ and

$$E[Q_r(t;K,h)] = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\left(\frac{w_i}{h}\right)\int E[b_{ijr}(t) \mid t_{ij} = s] K\left(\frac{t-s}{h}\right) f(s)ds\right\}.$$

To compute $E[Q_r(t;K,h)Q_p(t;K,h)]$, we consider the decomposition

$$Q_r(t;K,h)Q_p(t;K,h) = D_{rp1}(t) + D_{rp2}(t) + D_{rp3}(t), \tag{6.14}$$

where

$$D_{rp1}(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\left(\frac{w_i}{h}\right)^2 b_{ijr}(t)b_{ijp}(t)K^2\left(\frac{t-t_{ij}}{h}\right)\right\},$$
$$D_{rp2}(t) = \sum_{i=1}^{n}\sum_{j_1\neq j_2}\left\{\left(\frac{w_i}{h}\right)^2 b_{ij_1r}(t)b_{ij_2p}(t)K\left(\frac{t-t_{ij_1}}{h}\right)K\left(\frac{t-t_{ij_2}}{h}\right)\right\},$$
$$D_{rp2}(t) = \sum_{i_1\neq i_2}\sum_{j_1,j_2}\left\{\left(\frac{w_{i_1}w_{i_2}}{h^2}\right) b_{i_1j_1r}(t)b_{i_2j_2p}(t)K\left(\frac{t-t_{i_1j_1}}{h}\right)K\left(\frac{t-t_{i_2j_2}}{h}\right)\right\}.$$

By the definitions of $II_{r,p}(t)$ and $III_{r,p}(t)$, direct integration then shows that

$$E[D_{rp1}(t)] = \left(\sum_{i=1}^{n}\sum_{j=1}^{n_i} w_i^2 h^{-1}\right) II_{r,p}(t)(1+o(1)), \tag{6.15}$$

$$E[D_{rp2}(t)] = \left(\sum_{i=1}^{n} n_i(n_i-1)w_i^2\right) III_{r,p}(t)(1+o(1)), \tag{6.16}$$

$$E\left[D_{rp3}(t)\right] = [E(Q_r(t;K,h))][E(Q_p(t;K,h))](1+o(1)). \tag{6.17}$$

The asymptotic expression of $E[Q_r(t;K,h)Q_p(t;K,h)]$ follows from (6.14) through (6.17).

Acknowledgements

We would like to thank an associate editor and two referees for many helpful comments and suggestions.

References

Altman, N. S. (1990). Kernel smoothing of data with correlated errors, *J. Amer. Statist. Assoc.*, **85**, 749–759.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, Chapman Hall, London.

Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994). *Analysis of Longitudinal Data*, Oxford University Press, Oxford, England.

Härdle, W. (1990). Applied Nonparametric Regression, Cambridge University Press, Cambridge, England.

Hart, J. D. and Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data, *J. Amer. Statist. Assoc.*, **81**, 1080–1088.

Hart, J. D. and Wehrly, T. E. (1993). Consistency of cross-validation when the data are curves, *Stochastic Process. Appl.*, **45**, 351–361.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models, *J. Roy. Statist. Soc. Ser. B*, **55**, 757–796.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809–822.

Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semiparametric modeling of longitudinal data, *Austral. J. Statist.*, **36**, 75–93.

Nadaraya, E. A. (1964). On estimating regression, *Theory Probab. Appl.*, **10**, 186–190.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves, *J. Roy. Statist. Soc. Ser. B*, **53**, 233–243.

Rosenblatt, M. (1969). Conditional probability density and regression estimates, *Multivariate Analysis II* (ed. Krishnaiah), 25–31, Academic Press, New York.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā Series A*, **26**, 359–372.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics*, **50**, 689–699.