Statistica Sinica **10**(2000), 433-456

KERNEL SMOOTHING ON VARYING COEFFICIENT MODELS WITH LONGITUDINAL DEPENDENT VARIABLE

Colin O. Wu and Chin-Tsang Chiang

The Johns Hopkins University and Tunghai University

Abstract: This paper considers a nonparametric varying coefficient regression model with longitudinal dependent variable and cross-sectional covariates. The relationship between the dependent variable and the covariates is assumed to be linear at a specific time point, but the coefficients are allowed to change over time. Two kernel estimators based on componentwise local least squares criteria are proposed to estimate the time varying coefficients. A cross-validation criterion and a bootstrap procedure are used for selecting data-driven bandwidths and constructing confidence intervals, respectively. The theoretical properties of our estimators are developed through their asymptotic mean squared errors and mean integrated squared errors. The finite sample properties of our procedures are investigated through a simulation study. Applications of our procedures are illustrated through an epidemiological example of predicting the effects of cigarette smoking, pre-HIV infection CD4 cell percentage and age at HIV infection on the depletion of CD4 cell percentage among HIV infected persons.

Key words and phrases: Bandwidth selection, bootstrap, CD4 cell percent, HIV, local least squares, nonparametric regression, varying coefficient model.

1. Introduction

In a longitudinal study, observations are usually obtained from n independently selected subjects each repeatedly measured over a set of distinct time points. Interest of the study is often focused on evaluating the effects of time t and a set of covariates $X^{(l)}(t)$, $l = 1, \ldots, k$, which may or may not depend on t, on a time dependent outcome variable $Y(t)$. Let t_{ij} be the time of the jth measurement of the ith subject and $X_{ij}^{(l)}$, $l = 1, \ldots, k$, and Y_{ij} be the ith subject's observed covariates and outcomes at time t_{ij}. The longitudinal observations are given by $\{(t_{ij}, Y_{ij}, \mathbf{X}_{ij}^T); 1 \le i \le n, 1 \le j \le n_i\}$, where $\mathbf{X}_{ij} = (1, X_{ij}^{(1)}, \ldots, X_{ij}^{(k)})^T$ and n_i is the number of repeated measurements of the ith subject. Although the measurements are independent between different subjects, they are likely to be correlated within each subject.

Longitudinal data are common in medical and epidemiological studies. For example, in long-term follow-up cohort studies and clinical trials, patients' health status, such as CD4 (also known as T-helper lymphocytes) cell levels in Human Immunodeficiency Virus (HIV) infected persons, and other risk factors are often repeatedly measured over time (cf. Kaslow, Ostrow, Detels, Phair, Polk and Rinaldo (1987)).

Under multivariate linear and generalized linear regression models, estimation and inferences with longitudinal observations have been extensively studied by Pantula and Pollock (1985), Ware (1985), Liang and Zeger (1986), Jones (1987), Diggle (1988), Jones and Ackerson (1990), Jones and Boadi-Boteng (1991), Diggle, Liang and Zeger (1994), among others. Theory and methods with nonlinear models can be found in Davidian and Giltinan (1995) and Vonesh and Chinchilli (1997). Although parametric models and the corresponding estimation methods, such as weighted least squares, quasi-likelihoods and generalized estimation equations, have been used in numerous applications, they are less successful when there are no meaningful parametric forms available for the scientific problem being considered or the models are incorrectly specified. Existing nonparametric methods, such as Hart and Wehrly (1986), Altman (1990) and Rice and Silverman (1991), relaxed the parametric relationship between $Y(t)$ and t to a flexible smooth curve, but suffered the drawbacks of not including the effects of $X^{(l)}(t)$, $l = 1, \ldots, k$, into the model.

To incorporate the effects of both time and covariates, Zeger and Diggle (1994) considered estimation and inferences of the semiparametric partially linear model

$$Y(t) = \beta_0(t) + \sum_{l=1}^{k} X^{(l)}(t)\beta_l + \epsilon(t),$$

where $\beta_0(t)$ is a smooth function of t, β_l, $l = 1, \ldots, k$, are unknown parameters on the real line, $\epsilon(t)$ is a mean zero stochastic process and $X^{(l)}$ and $\epsilon(\cdot)$ are independent. However, for many situations, the effects of $X^{(l)}(t)$ on $Y(t)$ may not be constant over time, so that describing them through β_l appears to be unrealistic. Generally, one may model the relationship between $Y(t)$ and $(t, X^{(1)}(t), \ldots, X^{(k)}(t))$ through an unknown multivariate smooth function. But, in real applications, particularly when k is large, for example $k \geq 3$, multivariate smoothing methods may require unrealistically large sample sizes and often produce results that are biologically difficult to interpret.

A promising alternative is to consider regression models that are more flexible than the classical parametric or semiparametric models and also have specific structures which can be easily interpreted in real applications. For this purpose, Hoover, Rice, Wu and Yang (1998) considered the varying coefficient model

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

where $\mathbf{X}(t) = (1, X^{(1)}(t), \ldots, X^{(k)}(t))^T$, $\beta(t) = (\beta_0(t), \ldots, \beta_k(t))^T$, $\beta_l(t)$, $l = 0, \ldots, k$, are smooth functions of t, $\epsilon(t)$ is a mean zero stochastic process and $\mathbf{X}(\cdot)$ and $\epsilon(\cdot)$ are independent. These authors proposed, among other methods, a kernel estimator $\widehat{\beta}(t)$ of $\beta(t)$ which minimizes the ordinary local least squares criterion:

$$\ell_N(t) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \frac{1}{N} \left(Y_{ij} - \mathbf{X}_{ij}^T b(t) \right)^2 K\left(\frac{t - t_{ij}}{h} \right) \right\}, \tag{1.2}$$

with respect to $b(t) = (b_0(t), \ldots, b_k(t))$, where $N = \sum_{i=1}^{n} n_i$, $K(\cdot)$ is a Borel measurable kernel function and h is a positive bandwidth. When the observations are cross-sectional, i.e. each subject is only measured once, (1.1) is a special case of the models considered by Hastie and Tibshirani (1993).

For many longitudinal studies, the covariate variables are independent of t, so that their observations are cross-sectional. Then an important special case of (1.1) is

$$Y(t) = \mathbf{X}^T \beta(t) + \epsilon(t), \tag{1.3}$$

where $\mathbf{X} = (1, X^{(1)}, \ldots, X^{(k)})^T$, $X^{(l)}$, $l = 1, \ldots, k$, are time independent covariates and $\beta(t)$ and $\epsilon(t)$ are defined in (1.1). The longitudinal observations are now given by $\{(t_{ij}, Y_{ij}, \mathbf{X}_i^T);\ 1 \le i \le n, 1 \le j \le n_i\}$, where $\mathbf{X}_i = (1, X_i^{(1)}, \ldots, X_i^{(k)})^T$. The HIV/Growth example studied by Hoover, et al. (1998) actually belongs to this special case.

We consider the estimation of $\beta(t)$ based on (1.3) and $\{(t_{ij}, Y_{ij}, \mathbf{X}_i^T);\ 1 \le i \le n, 1 \le j \le n_i\}$. Using the special feature that the $\mathbf{X}_i$ are cross-sectional, we propose two new kernel estimators based on componentwise local least squares criteria. Our estimators have two major advantages over the ordinary local least squares approach of (1.2). First, since (1.2) relies on only one set of bandwidth and kernel function to simultaneously estimate all $(k+1)$ curves of $\beta(t)$, $\widehat{\beta}(t)$ may not be able to provide adequate smoothing for some components of $\beta(t)$ when $\beta_l(t)$, $l = 1, \ldots, k$, belong to different smoothness families. Our estimators allow for $(k+1)$ different sets of bandwidths and kernel functions and are capable of providing separate smoothing for each component of $\beta(t)$. Second, since (1.2) assigns equal weight to each measurement point $(t_{ij}, Y_{ij}, \mathbf{X}_i^T)$, $\widehat{\beta}(t)$ may be inconsistent when $\max_{1 \le i \le n}(n_i/N)$ does not converge to zero as $n \to \infty$ (cf. Hoover, et al. (1998)). Because the subjects are assumed to be independent of each other, our estimators use two weighting schemes: (i) equal weight for each subject; (ii) equal weight for each measurement. The first weighting scheme always gives consistent estimators regardless of how the n_i are chosen. These weighting schemes are identical when all subjects have the same number of repeated measurements and give different asymptotic results when the numbers of repeated measurements are different per subject. None of the weighting schemes uniformly dominates the other theoretically, or in applications. A general weighted local least squares approach may require the knowledge of correlation structures of the data, which are usually unknown in practice. Thus, our estimation approach differs from that of Hoover, et al. (1998), both in smoothing techniques and weighting schemes. The asymptotic results of this paper show that the convergence rates of our estimators are at least as good as, and frequently better than, that of $\widehat{\beta}(t)$. In addition to the estimation method, we also propose a cross-validation procedure for selecting data-driven bandwidths and a "resampling-subject" bootstrap procedure for constructing pointwise confidence intervals. Theoretical properties of these bandwidth and confidence procedures have not yet been developed. However, through a Monte Carlo simulation and an epidemiological example of CD4 cell levels in HIV infected persons, we show that our procedures are useful in nonparametric longitudinal analysis and implementable in practice.

For the rest of the paper, we summarize the estimation, bandwidth selection and bootstrap confidence procedures in Section 2. The application of model (1.3) and our procedures to the CD4/HIV example is given in Section 3. Section 4 presents the results of the simulation study. The asymptotic representations of the mean squared errors and the mean integrated squared errors of our estimators are derived in Section 5. Finally, proofs of the asymptotic results are sketched in appendices.

2. Estimation Methods

2.1. Componentwise kernel estimators

For most epidemiological studies, the subjects are randomly selected, so that the observed covariates $\mathbf{X}_i$ are random. We assume throughout this paper that the covariate vector $\mathbf{X}$ of (1.3) is random, $\mathbf{X}$ and $\epsilon(\cdot)$ are independent and the $(k+1)\times(k+1)$ matrix $E(\mathbf{X}\mathbf{X}^T)$ is invertible. Let $E^{-1}_{\mathbf{X}\mathbf{X}^T}$ be the inverse of $E(\mathbf{X}\mathbf{X}^T)$. Multiplying both sides of (1.3) by $\mathbf{X}$ and then taking expectations, $\beta(t)$ can be expressed as

$$\beta(t) = \left(E^{-1}_{\mathbf{X}\mathbf{X}^T}\right) E\left[\mathbf{X}Y(t)\right]. \tag{2.1}$$

Let e_{rl} be the (r,l)th element of $E^{-1}_{\mathbf{X}\mathbf{X}^T}$. Then (2.1) implies that, for $r = 0,\ldots,k$,

$$\beta_r(t) = E\Big[\Big(\sum_{l=0}^{k} e_{rl} X^{(l)}\Big) Y(t)\Big]. \tag{2.2}$$

Since $E(\mathbf{X}\mathbf{X}^T)$ is time-independent, it can be simply estimated by the sample mean $\widehat{E}_{\mathbf{X}\mathbf{X}^T} = n^{-1}\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)$. Suppose that $\widehat{E}_{\mathbf{X}\mathbf{X}^T}$ is invertible. A natural estimator of e_{rl} is $\widehat{e}_{rl}$ where $\widehat{e}_{rl}$ denotes the (r,l)th element of $(\widehat{E}_{\mathbf{X}\mathbf{X}^T})^{-1}$.

Substituting e_{rl} in (2.2) by $\widehat{e}_{rl}$, we can first approximate $(\sum_{l=0}^{k} e_{rl}X^{(l)})Y(t)$ by $(\sum_{l=0}^{k} \widehat{e}_{rl}X^{(l)})Y(t)$ and then estimate $E[(\sum_{l=0}^{k} \widehat{e}_{rl}X^{(l)})Y(t)]$ by minimizing

$$L_r(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\Big(\frac{1}{nn_i}\Big)\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij} - b_r(t)\Big]^2 K_r\Big(\frac{t-t_{ij}}{h_r}\Big)\Big\} \tag{2.3}$$

with respect to $b_r(t)$, where $K_r(\cdot)$ is a Borel measurable kernel function and h_r is a positive bandwidth. The solution of (2.3) leads to the estimator $\widetilde{\beta}(t) = (\widetilde{\beta}_0(t), \ldots, \widetilde{\beta}_k(t))^T$, where $\beta_r(t)$ is estimated by

$$\widetilde{\beta}_r(t) = \frac{\sum_{i=1}^{n}\Big\{n_i^{-1}\sum_{j=1}^{n_i}\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}K_r\left(\left(t-t_{ij}\right)/h_r\right)\Big]\Big\}}{\sum_{i=1}^{n}\Big\{n_i^{-1}\sum_{j=1}^{n_i}\left[K_r\left(\left(t-t_{ij}\right)/h_r\right)\right]\Big\}}. \tag{2.4}$$

Note that (2.3) assigns weight $(nn_i)^{-1}$ to each measurement of the ith subject. These weights are different when the numbers of repeated measurements are different per subject.

An alternative local least squares approach is to minimize

$$L_r^*(t) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\frac{1}{N}\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij} - b_r(t)\Big]^2 K_r\Big(\frac{t-t_{ij}}{h_r}\Big)\Big\} \tag{2.5}$$

with respect to $b_r(t)$. This leads to a kernel estimator $\widetilde{\beta}^*(t) = (\widetilde{\beta}_0^*(t), \ldots, \widetilde{\beta}_k^*(t))^T$, where

$$\widetilde{\beta}_r^*(t) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}K_r\left(\left(t-t_{ij}\right)/h_r\right)\Big]}{\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[K_r\left(\left(t-t_{ij}\right)/h_r\right)\right]} \tag{2.6}$$

estimates $\beta_r(t)$. Similar to (1.2), (2.5) assigns the uniform weight N^{-1} to all the measurement points, so that $\widetilde{\beta}^*$ is more influenced by those subjects with large numbers of repeated measurements.

Remark 2.1. Contrary to (1.2), which depends on a kernel smoother to estimate $E(\mathbf{XX}^T)$, the smoothing methods of (2.3) and (2.5) rely heavily on the time independent nature of $\mathbf{X}$ and use the sample mean to estimate $E(\mathbf{XX}^T)$. Thus, different smoothing needs of $\beta_r(t)$, $r = 0, \ldots, k$, can be adjusted in (2.3) and (2.5) by selecting appropriate bandwidths h_r and kernels $K_r(\cdot)$. The choices of weighting schemes may also have profound influences on the adequacy of the estimators. It may be theoretically beneficial if $(nn_i)^{-1}$ in (2.3) or N^{-1} in (2.5) could be replaced by non-negative weights w_i, $i = 1, \ldots, n$, which depend on the intra-correlations of the data. However, the structures of the intra-correlations are often unknown in practice, so that the special choices $(nn_i)^{-1}$ and N^{-1} appear to be natural in these situations. We will see in Section 5 that neither

$\widetilde{\beta}_r(t)$ or $\widetilde{\beta}^*_r(t)$ asymptotically dominates the other uniformly. However, Monte Carlo simulation similar to those presented in Section 4 consistently indicates that $\widetilde{\beta}_r(t)$ provides better fits than $\widetilde{\beta}^*_r(t)$. Thus, for the rest of the paper, we concentrate on the properties of $\widetilde{\beta}_r(t)$, and leave the properties of $\widetilde{\beta}^*_r(t)$ to brief remarks.

Remark 2.2. Of course the idea employed in (2.3) and (2.5) can be extended to other smoothing methods. For example, a $(d-1)$-degree local polynomial estimator $\widetilde{b}_r(t)$ can be obtained by minimizing (2.3) or (2.5) with $\beta_r(t)$ replaced by a polynomial of the form $\sum_{s=0}^{d-1}(t-t_{ij})^s b_{sr}(t)$. The local polynomial estimator $\widetilde{b}(t) = (\widetilde{b}_0(t), \ldots, \widetilde{b}_k(t))^T$ then also uses $(k+1)$ different sets of bandwidths and kernel functions, while the ordinary least squares local polynomials of Hoover, et al. (1998) restrict to one bandwidth and kernel. The additional bandwidths and kernel functions used in $\widetilde{b}_r(t)$ provide more flexibility for adapting the smoothing needs of $\beta_r(t)$, $r = 0, \ldots, k$. When $\beta_r(t)$, $r = 0, \ldots, k$, are twice continuously differentiable on the real line, another approach is to estimate $\beta(t)$ by a smoothing spline estimator $\widetilde{\beta}^{(S)}(t) = (\widetilde{\beta}^{(S)}_0(t), \ldots, \widetilde{\beta}^{(S)}_k(t))^T$ such that $\widetilde{\beta}^{(S)}_r(t)$ minimizes

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\frac{1}{nn_i}\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij} - b_r(t_{ij})\Big]^2\Big\} + \lambda_r\int (b_r''(s))^2 ds$$

with respect to $b_r(t)$, where λ_r is a positive smoothing parameter. Computationally, $\widetilde{b}(t)$ and $\widetilde{\beta}^{(S)}(t)$ require more effort than $\widetilde{\beta}(t)$. The theoretical and practical properties of $\widetilde{b}(t)$ and $\widetilde{\beta}^{(S)}(t)$ have not been developed and deserve a further study.

2.2. Cross-validation bandwidth choices

It is well known in kernel regression that the selection of bandwidths is generally more important than the selection of kernel functions. In practice, under-smoothing or over-smoothing is mainly caused by inappropriate bandwidth choices, but is rarely influenced by the kernel shapes. Usual choices of kernels, such as the standard Gaussian kernel, the Epanechnikov kernel and other probability density functions, normally give satisfactory results. Bandwidths may be selected subjectively by examining the plots of the fitted curves. But finding automatic bandwidths suggested by the data is of both theoretical and practical interest.

Following the heuristic procedure suggested by Rice and Silverman (1991), we calculate the bandwidths of $\widetilde{\beta}(t)$ using the following "leave-one-subject-out" cross-validation: Let $\widetilde{\beta}_{-i}(t)$ be a kernel estimator of $\beta(t)$ computed using the data with all the repeated measurements of the ith subject left out, and define

$$\mathrm{CV}(h) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\frac{1}{nn_i}\Big(Y_{ij} - \mathbf{X}_i^T\widetilde{\beta}_{-i}(t_{ij})\Big)^2\Big\} \tag{2.7}$$

to be the cross-validation score of $h = (h_0, \ldots, h_k)^T$. The cross-validation bandwidth vector $h_{cv} = (h_{0,cv}, \ldots, h_{k,cv})^T$ is then defined to be the unique minimizer of $\mathrm{CV}(h)$.

When the dimensionality of $\mathbf{X}_i$ is high, the search for h_{cv} may be difficult. However, it is usually easy to find a suitable range of the bandwidths by examining the plots of the fitted curves. Within a given range of $h = (h_0, \ldots, h_k)$, one can approximate the value of h_{cv} by computing $\mathrm{CV}(h)$ through a series of $h = (h_0, \ldots, h_k)$ choices. This method for searching h_{cv}, although rather *ad hoc*, may actually speed up the computation and give a satisfactory bandwidth vector in practice. However, a systematic search for h_{cv}, particularly when k is large, may require sophisticated optimization algorithms beyond the scope of this paper.

Remark 2.3. To see why $\mathrm{CV}(h)$ is a reasonable criterion in practice, we consider the following decomposition

$$\mathrm{CV}(h)=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\frac{1}{nn_i}(Y_{ij}-\mathbf{X}_i^T\beta(t_{ij}))^2\right\}+\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\frac{1}{nn_i}\left[\mathbf{X}_i^T\left(\beta(t_{ij})-\widetilde{\beta}_{-i}(t_{ij})\right)\right]^2\right\}$$
$$+2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\frac{1}{nn_i}\left(Y_{ij}-\mathbf{X}_i^T\beta(t_{ij})\right)\left[\mathbf{X}_i^T\left(\beta(t_{ij})-\widetilde{\beta}_{-i}(t_{ij})\right)\right]\right\}. \qquad (2.8)$$

The first term of the right hand side of (2.8) does not depend on the bandwidths, while, because of the definition of $\widetilde{\beta}_{-i}(t)$, the expectation of the third term is zero. Let $\mathrm{ASE}(\widetilde{\beta})$ be the average squared error of $\mathbf{X}_i^T\widetilde{\beta}(t_{ij})$, i.e.,

$$\mathrm{ASE}(\widetilde{\beta})=\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left\{\frac{1}{nn_i}\left[\mathbf{X}_i^T\left(\beta(t_{ij})-\widetilde{\beta}(t_{ij})\right)\right]^2\right\}.$$

It is easy to see that the expectation of the second term of the right hand side of (2.8) is actually the expectation of $\mathrm{ASE}(\widetilde{\beta}_{-i})$, which approximates the expectation of $\mathrm{ASE}(\widetilde{\beta})$ when n is large. Thus, h_{cv} approximately minimizes the average squared error $\mathrm{ASE}(\widetilde{\beta})$. Consistency of a similar "leave-one-subject-out" cross-validation procedure in a different but simpler nonparametric regression setting has been shown by Hart and Wehrly (1993). But, under the current setting, the asymptotic and finite sample properties of h_{cv} have not been fully investigated.

2.3. Bootstrap confidence intervals

Statistical inferences, such as confidence regions, hypotheses testing and model diagnoses, are usually developed based on either asymptotic distributions of the estimators or bootstrap methods (cf. Efron and Tibshirani (1993)). In the context of cross-sectional data, asymptotic distributions are derived by letting the number of subjects n go to infinity. Thus the resulting inferences are reliable, at least when the sample size is large. However, the longitudinal data structure considered in this paper is more complicated because of two reasons. First, since the numbers of repeated measurements n_i, $i = 1, \ldots, n$, are allowed to be different, the corresponding asymptotic distributions of the estimators may also be different depending on how fast n_i, $i = 1, \ldots, n$, converge to infinity relative to n. It will be seen from the asymptotic properties of Section 5 that, in order to get a meaningful asymptotic result, n must converge to infinity, but n_i may be either bounded or converging to infinity as $n \to \infty$. Second, because of the possible intra-correlation structures of the data, which are assumed to be completely unknown, the asymptotic distributions may involve bias and correlation terms which are difficult to estimate. Thus, inferences which are purely based on the asymptotic distributions of $\widetilde{\beta}(t)$ may be difficult to implement in practice.

Since subjects are independently selected, a natural bootstrap sampling scheme is to resample the entire repeated measurements of each subject with replacement from the original data set. Based on $\widetilde{\beta}(t)$ and this resampling subject bootstrapping, the following naive bootstrap procedure can be used to construct approximate pointwise percentile confidence intervals for $\beta_r(t)$:

1. Randomly sample n subjects with replacement from the original data set, and let $\{(t^*_{ij}, \mathbf{X}^*_i, Y^*_{ij});\ 1 \le i \le n, 1 \le j \le n_i\}$ be the longitudinal bootstrap sample. Here the entire repeated measurements of some subjects in the original sample may appear two or more times in the new bootstrap sample.
2. Compute the kernel estimator $\widetilde{\beta}^{boot}_r(t)$ of $\beta_r(t)$ based on (2.4) (or (2.6)) and the bootstrap sample.
3. Repeat the above two steps B times, so that B bootstrap estimators $\widetilde{\beta}^{boot}_r(t)$ of $\beta_r(t)$ are obtained.
4. Let $L_{(\alpha/2)}(t)$ and $U_{(\alpha/2)}(t)$ be the $(\alpha/2)$th and $(1-\alpha/2)$th, i.e. lower and upper $(\alpha/2)$th percentiles, respectively, calculated based on the B bootstrap estimators. An approximate $(1-\alpha)$ bootstrap confidence interval for $\beta_r(t)$ is given by $(L_{(\alpha/2)}(t), U_{(\alpha/2)}(t))$.

The main advantage of this naive bootstrap procedure is that it does not rely on the asymptotic distributions of $\widetilde{\beta}(t)$. An alternative bootstrap procedure suggested by Hoover, et al. (1998), which relies on normal approximations of the critical values, is to construct pointwise intervals of the form

$$\widetilde{\beta}_r(t) \pm z_{(1-\alpha/2)} \widetilde{se}^*_B(t), \tag{2.9}$$

where $\widetilde{se}^*_B(t)$ is the estimated standard error of $\widetilde{\beta}_r(t)$ from the B bootstrap estimators and $z_{(1-\alpha/2)}$ is the $(1-\alpha/2)$th percentile of the standard Gaussian distribution. Technically, both (2.9) and our naive bootstrap percentile procedure may lead to good approximations of the actual $(1-\alpha)$ confidence intervals when the biases of $\widetilde{\beta}_r(t)$ are negligible; but the biases of the estimators have to be adjusted when they are not negligible. Theoretical properties of these bootstrap procedures have not been developed.

Remark 2.4. The "resampling-subject" bootstrap can also be applied to other smoothing estimators, such as local polynomials or smoothing splines. When the simultaneous confidence regions for $\beta(\cdot)$ are of interest, these bootstrap procedures can be used to construct Bonferroni-type confidence bands. However, it is well known that the Bonferroni method usually gives excessively conservative bands in practice. Further study on simultaneous inferences for $\beta(\cdot)$ may be worthwhile.

3. Application to CD4/HIV Study

Since CD4 cells are vital for immune function, CD4 cell count and percentage, i.e., CD4 cell count divided by the total number of lymphocytes, are currently the most commonly used markers for the health status of HIV infected persons. The dataset considered here is from the Multicenter AIDS Cohort Study, which includes the repeated measurements of physical examinations, laboratory results and CD4 percentages of 283 homosexual men who became HIV positive between 1984 and 1991. All individuals were scheduled to have their measurements made at semi-annual visits. However, since many individuals missed some of their scheduled visits and all the HIV infections happened randomly during the study, we have unequal numbers of repeated measurements and different design times t_{ij} per individual. Further details about the design, methods and medical implications of the study can be found in Kaslow, et al. (1987).

The objective here is to evaluate the effects of cigarette smoking, pre-HIV infection CD4 percentage and age at HIV infection on the mean CD4 percentage after the infection. Since the covariates are time independent, (1.3) and the kernel method of (2.4) can be applied. Let t_{ij} be the time (in years) of the jth measurement of the ith individual after HIV infection, Y_{ij} the ith individual's CD4 percentage at time t_{ij} and $X_i^{(1)}$ the ith individual's smoking status (equal to 1 if he always smokes cigarettes and 0 if he never smokes cigarettes). To ensure a clear biological interpretation of our results, we define $X_i^{(2)}$ to be the ith individual's centered pre-infection CD4 percentage, i.e., $X_i^{(2)}$ is obtained by subtracting the average pre-infection CD4 percentage of the sample from the ith individual's actual pre-infection CD4 percentage. Similarly, $X_i^{(3)}$, obtained by subtracting the sample average age at infection from the ith individual's age at infection, is the ith individual's centered age at HIV infection. The main advantage of using the centered covariates $X_i^{(2)}$ and $X_i^{(3)}$ is that $\beta_l(t)$, $l = 0, \ldots, 3$, have clear biological interpretations. Here, $\beta_0(t)$ represents the baseline CD4 percentage and can be interpreted as the mean CD4 percentage at time t for a nonsmoker with average pre-infection CD4 percentage and average age at HIV infection. Then, $\beta_1(t)$, $\beta_2(t)$ and $\beta_3(t)$ can be interpreted as the effects of cigarette smoking, pre-infection CD4 percentage and age at HIV infection on the post-HIV infection CD4 percentage at time t.

The estimators $\widetilde{\beta}_r(t)$, $r = 0, \ldots, 3$, were computed based on (2.4) and the standard Gaussian kernel. For the choice of bandwidths, a set of subjective bandwidths, $h_r = 1.5$ for all $r = 0, \ldots, 3$, was chosen by examining the plots of the estimated curves. The cross-validated bandwidths were then selected by minimizing the values of $\mathrm{CV}(h)$ over a series of $(h_0, \ldots, h_3)$ values. With 200 bootstrap replications, the bootstrap percentile procedure of Section 2.3 was used to construct pointwise confidence intervals for $\beta_0(t), \ldots, \beta_3(t)$ at 60 equally spaced time points between 0 and 5.9 years.

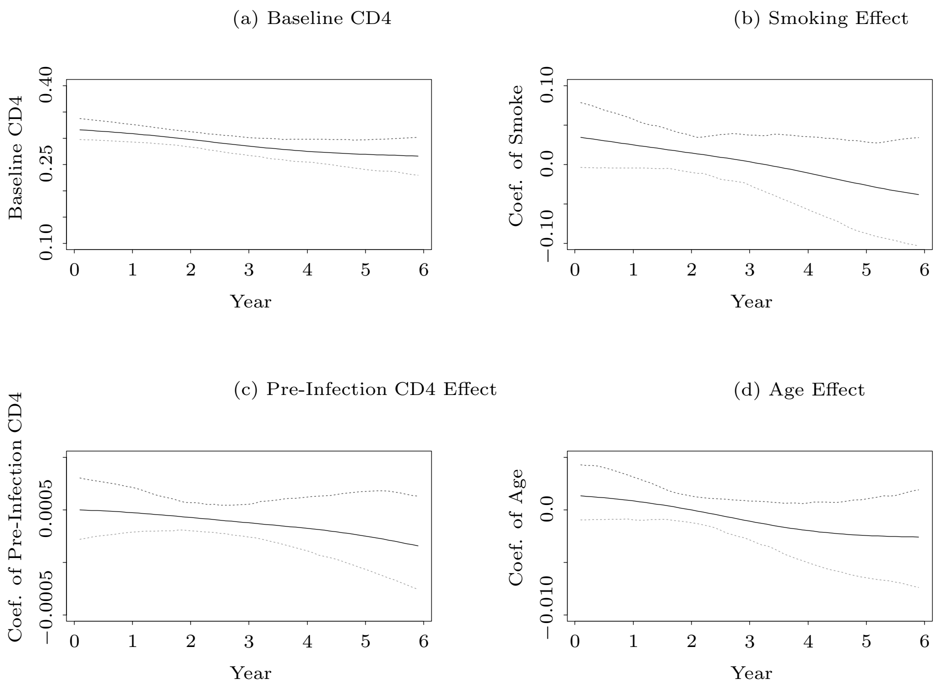


Figure 1. The solid curves in (a), (b), (c) and (d) show the kernel estimators $\widetilde{\beta}_r(t)$, $r = 0, 1, 2, 3$, respectively, based on (2.4), the standard Gaussian kernel and $h_S = (1.5, 1.5, 1.5, 1.5)^T$. The dotted curves represent the corresponding 95% bootstrap percentile pointwise intervals.

The solid curves of Figures (1a) through (1d) show the estimators $\widetilde{\beta}_0(t), \ldots, \widetilde{\beta}_3(t)$ based on the subjective bandwidths $h_r = 1.5$, $r = 0, \ldots, 3$, while the dotted curves give the corresponding 95% bootstrap percentile confidence intervals.

Similarly, the solid curves of Figures (2a) through (2d) show the estimators $\widetilde{\beta}_0(t)$, ..., $\widetilde{\beta}_3(t)$ based on the cross-validated bandwidths, and the dotted curves give the 95% bootstrap percentile confidence intervals for $\beta_r(t)$, $r = 0, \ldots, 3$.

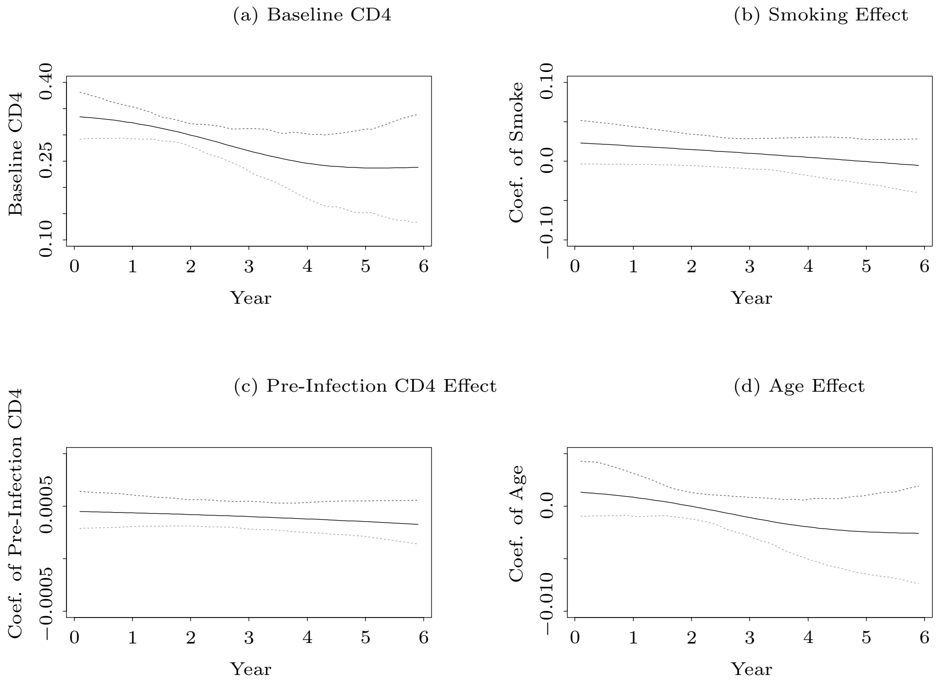


Figure 2. The solid curves in (a), (b), (c) and (d) show the kernel estimators $\widetilde{\beta}_r(t)$, $r = 0, 1, 2, 3$, respectively, based on (2.4), the standard Gaussian kernel and the cross-validated bandwidths $h_{cv} = (3.0, 3.0, 1.5, 3.0)^T$. The dotted curves represent the corresponding 95% bootstrap percentile pointwise intervals.

From these figures, we see that the mean baseline CD4 percentage of the population depletes rather quickly at the beginning of HIV infection, but the rate of depletion appears to be slowing down at four years after the infection. Cigarette smoking and age of HIV infection do not show any significant effect on the post-infection CD4 percentage. But pre-infection CD4 percentage appears to be positively associated with higher post-infection CD4 percentage.

4. Simulation Results

We consider a design that is similar to the nature of the CD4/HIV study of Section 3. Based on (1.3), the cross-sectional covariate vector $\mathbf{X} = (1, X^{(1)}, X^{(2)})^T$ is specified so that the covariates $X^{(1)}$ and $X^{(2)}$ are independent random variables with a joint density

$$f(x_1, x_2) = \frac{1}{8(2\pi)^{1/2}} \exp\left(\frac{-x_2^2}{32}\right) 1_{\{0,1\}}(x_1) 1_{(-\infty,\infty)}(x_2).$$

The coefficient curves are given by $\beta_0(t) = 3.5 + 6.5\sin(t\pi/60)$,

$$\beta_1(t) = -0.2 - 1.6\cos\left(\frac{(t-30)\pi}{60}\right) \quad \text{and} \quad \beta_2(t) = 0.25 - 0.0074\left(\frac{30-t}{10}\right)^3.$$

A simple random sample $\{\mathbf{X}_i;\ i = 1, \ldots, 400\}$ of $\mathbf{X}$ is generated. Each subject is designed to appear in the 31 scheduled equally spaced time points between 0 and 30, i.e., $0, \ldots, 30$, but has a probability of 60% to be randomly missing. This leads to unequal design and the remaining unequally spaced time points are denoted by t_{ij}, $1 \le i \le 400$, $1 \le j \le n_i$. The random errors ϵ_{ij} are generated from the Gaussian process with zero mean and covariance function

$$\text{cov}\left(\epsilon_{i_1 j_1}, \epsilon_{i_2 j_2}\right) = \begin{cases} 0.0625\exp\left(-\left|t_{i_1 j_1} - t_{i_2 j_2}\right|\right), & \text{if } i_1 = i_2, \\ 0, & \text{if } i_1 \ne i_2. \end{cases}$$

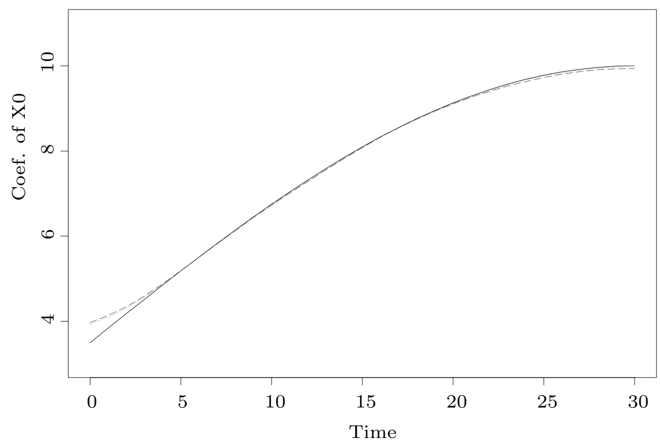
The time dependent responses Y_{ij} are obtained by substituting t_{ij}, $\mathbf{X}_i$, ϵ_{ij} and the above coefficient curves into (1.3).

Two hundred simulated longitudinal samples $\{(t_{ij}, \mathbf{X}_i, Y_{ij});\ 1 \le i \le 400, 1 \le j \le n_i\}$ were independently generated. For each simulated dataset, kernel estimators $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ were computed for $t \in [0, 30]$ using the standard Gaussian kernel and the cross-validated bandwidth vectors h_{cv} . Through a close examination of the CV scores for different bandwidths, the CV scores of $h_S = (2.0, 2.0, 2.0)^T$ for both $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ are very close to their corresponding global minima. Thus, kernel estimators $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ based on the standard Gaussian kernel and the subjective bandwidth vector $h_S = (2.0, 2.0, 2.0)^T$ were also computed. Estimators based on other commonly used kernels, such as the Epanechnikov kernel and the uniform kernel, gave similar results that are omitted. Using $\widetilde{\beta}_r(t)$, $\widetilde{\beta}_r^*(t)$ and the bootstrap percentile procedure of Section 2.3, we constructed the 95% pointwise confidence intervals for $\beta_r(t)$, $r = 0, 1, 2$, at a sequence of time points based on each simulated dataset and computed the estimated coverage probabilities of the intervals based on all the replicated datasets.

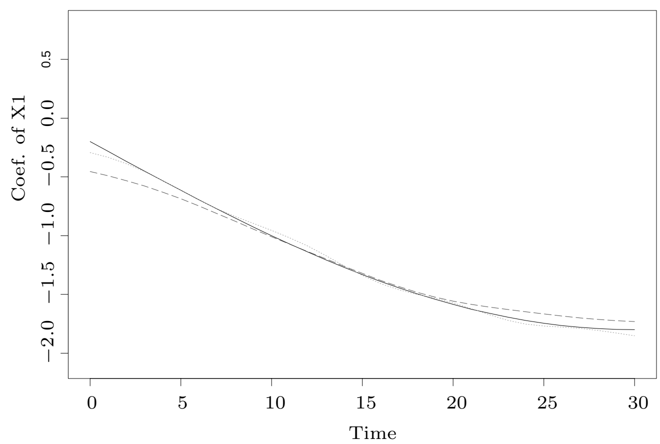
We only present the simulation results, including the graphs of the estimated curves and the estimated coverage probabilities of the bootstrap confidence intervals, of $\widetilde{\beta}(t)$. Although $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ appear to be intuitive for the current setting, the graphs of $\beta_r(t)$, $\widetilde{\beta}_r(t)$ and $\widetilde{\beta}_r^*(t)$ for $r = 0, 1, 2$ show that, at least for the simulated example considered, $\widetilde{\beta}(t)$ is superior to $\widetilde{\beta}^*(t)$. The bootstrap confidence intervals of $\widetilde{\beta}_r(t)$ also have much better coverage probabilities than those of $\widetilde{\beta}_r^*(t)$. One possible reason for the superiority of $\widetilde{\beta}(t)$ over $\widetilde{\beta}^*(t)$ is the unequal numbers of repeated measurements. When n_i, $i = 1, \ldots, n$, are equal or at least close to each other, the performance of $\widetilde{\beta}^*(t)$ is improved and similar to that of $\widetilde{\beta}(t)$.

To see the average performance of $\widetilde{\beta}_r(t)$, Figures (3a), (3b) and (3c) show the graphs of $\beta_r(t)$ and the mean curves over all 200 simulated samples of $\widetilde{\beta}_r(t)$ based on h_{cv} and h_S. Although both bandwidth choices give reasonable estimators, at least for the interior t points, the estimators based on h_S appear to be slightly better than those based on h_{cv}. For both bandwidth choices, the estimators are subject to larger biases near the boundary than at the interior points.

(3a) Beta0



(3a) Beta1



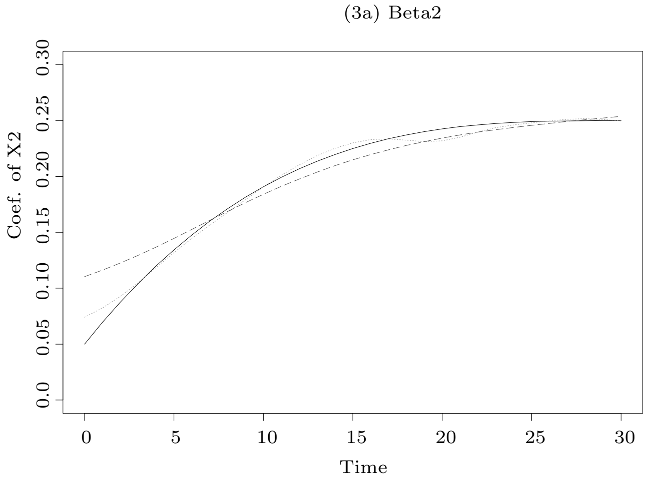


Figure 3. The solid curves in (a), (b) and (c) show the true values of $\beta_0(t)$, $\beta_1(t)$ and $\beta_2(t)$, respectively. The corresponding dashed curves show the means of $\widetilde{\beta}_0(t)$, $\widetilde{\beta}_1(t)$ and $\widetilde{\beta}_2(t)$ over all the simulated samples based on the cross-validated bandwidths and the standard Gaussian kernel. The corresponding dotted curves show the means of $\widetilde{\beta}_0(t)$, $\widetilde{\beta}_1(t)$ and $\widetilde{\beta}_2(t)$ over all the simulated samples based on $h_0 = h_1 = h_2 = 2.0$ and the standard Gaussian kernel.

Table 1 summarizes the estimated coverage probabilities of the bootstrap percentile confidence intervals for $\beta_r(t)$, $r = 0, 1, 2$, at nine equally spaced time points between 3.0 and 27.0. Except for $\beta_2(3.0)$, both bandwidth choices give similar coverage probabilities which are, for most part, close to satisfactory. The few lower than expected coverage values suggest that the naive bootstrap procedure may be subject to further improvement. One possibility is to adjust for the biases of the estimators. Practical bias adjustment methods are still subject to study.

Table 1. Estimated coverage probabilities of 95% bootstrap percentile confidence intervals for $\widetilde{\beta}_r(t)$, $r = 0, 1, 2$, at 9 time points based on h_{cv}, $h_S = (2.0, 2.0, 2.0)^T$ and the standard Gaussian kernel.

	Time Point	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0
	$\widetilde{\beta}_0$	0.90	0.97	0.93	0.94	0.89	0.94	0.91	0.92	0.96
h_{cv}	$\widetilde{\beta}_1$	0.91	0.95	0.94	0.92	0.92	0.93	0.94	0.95	0.97
	$\widetilde{\beta}_2$	0.64	0.96	0.90	0.82	0.87	0.94	0.95	0.96	0.96
	$\widetilde{\beta}_0$	0.91	0.95	0.94	0.94	0.88	0.94	0.90	0.92	0.94
h_S	$\widetilde{\beta}_1$	0.95	0.94	0.96	0.95	0.92	0.92	0.90	0.94	0.96
	$\widetilde{\beta}_2$	0.94	0.94	0.92	0.90	0.90	0.93	0.94	0.93	0.94

5. Asymptotic Risk Representations

We now derive the asymptotic representations of the mean squared errors and the mean integrated squared errors of $\widetilde{\beta}(t)$. The asymptotic risks of $\widetilde{\beta}^*(t)$ can be derived similarly, hence are only briefly discussed. These asymptotic results demonstrate that, when the covariates are cross-sectional, the componentwise kernel estimators $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ are generally superior to the ordinary least squares kernel estimator $\widehat{\beta}(t)$ obtained by minimizing (1.2).

5.1. Mean squared risks

For mathematical convenience, we specify that the time design points t_{ij} are randomly selected from a distribution function F with density f. But the n_i, $i = 1, \ldots, n$, are assumed to be nonrandom. This corresponds to random designs in regression analysis. However, by modifying the notation and several key steps in the derivations, the main results of this section can be extended to fixed designs, or to the case that the n_i are also random.

Because $\widetilde{\beta}(t)$ and $\widetilde{\beta}^*(t)$ are R^{k+1} valued estimators, their closeness to $\beta(t)$ can be measured in different ways. Suppose that we are only interested in the adequacy of one component of $\widetilde{\beta}(t)$, say $\widetilde{\beta}_r(t)$. A natural risk of $\widetilde{\beta}_r(t)$ at point t is the mean squared error $E\{[\widetilde{\beta}_r(t) - \beta_r(t)]^2\}$. For the risk of $\widetilde{\beta}(t)$ at t, one can then use $E\{(\widetilde{\beta}(t) - \beta(t))^T W(\widetilde{\beta}(t) - \beta(t))\} = \sum_{r=0}^{k} W_r E\{[\widetilde{\beta}_r(t) - \beta_r(t)]^2\}$ where W is a $(k+1) \times (k+1)$ diagonal matrix with nonnegative diagonal elements $W_0, \ldots, W_k$.

Unfortunately, a minor technical difficulty for kernel regression estimators is that their moments, hence $E\{[\widetilde{\beta}_r(t) - \beta_r(t)]^2\}$, may not exist (cf. Rosenblatt (1969)), so that modifications of the mean squared errors have to be used. By (2.4), it is easy to see that

$$\widetilde{\beta}_r(t) = \left(\widetilde{f}_r(t)\right)^{-1} \widetilde{m}_r(t), \tag{5.1}$$

where

$$\widetilde{m}_r(t) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Big[\Big(\frac{1}{n n_i h_r}\Big)\Big(\sum_{l=0}^{k} \widehat{e}_{rl} X_i^{(l)}\Big) Y_{ij} K_r\Big(\frac{t - t_{ij}}{h_r}\Big)\Big] \tag{5.2}$$

and

$$\widetilde{f}_r(t) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Big[\Big(\frac{1}{n n_i h_r}\Big) K_r\Big(\frac{t - t_{ij}}{h_r}\Big)\Big]. \tag{5.3}$$

Straightforward algebra, using (5.1), (5.2) and (5.3), shows that

$$(1 - \Delta_r(t)) \left(\widetilde{\beta}_r(t) - \beta_r(t)\right) = (f(t))^{-1} \left(\widetilde{m}_r(t) - \beta_r(t) \widetilde{f}_r(t)\right),$$

where $\Delta_r(t) = 1 - (\widetilde{f}_r(t)/f(t))$.

For any interior point t of the support of $f(\cdot)$, it can be shown by the method used in kernel density estimation with independent cross-sectional data (cf. Silverman (1986)) that $\Delta_r(t) \to 0$ in probability as $n \to \infty$ and $h_r \to 0$. Then, by (5.1), (5.2) and (5.3), we have the following approximation:

$$(1 + o_p(1)) \left(\widetilde{\beta}_r(t) - \beta_r(t)\right) = (f(t))^{-1} \widetilde{R}_r(t), \tag{5.4}$$

where $\widetilde{R}_r(t) = \widetilde{m}_r(t) - \beta_r(t)\widetilde{f}_r(t)$.

We define the local and global risks of $\widetilde{\beta}_r(\cdot)$ by its modified mean squared error,

$$\text{MSE}\left(\widetilde{\beta}_r(t)\right) = E\Big\{[(f(t))^{-1}\widetilde{R}_r(t)]^2\Big\}, \tag{5.5}$$

and modified mean integrated squared error,

$$\text{MISE}\left(\widetilde{\beta}_r\right) = \int \text{MSE}\left(\widetilde{\beta}_r(s)\right)\ \pi(s)ds, \tag{5.6}$$

respectively, where $\pi(s)$ is any non-negative weight function whose support is a compact subset in the interior of the support of $f(\cdot)$. Similar to nonparametric regression with independent cross-sectional data, the compact support of $\pi(s)$ is used to remove the boundary effects of the kernel estimators (cf. Marron and Härdle (1986)). The local and global risks of $\widetilde{\beta}(\cdot)$ can be defined by

$$\text{MSE}\left(\widetilde{\beta}(t)\right) = \sum_{r=0}^{k} W_r \text{MSE}\left(\widetilde{\beta}_r(t)\right) \tag{5.7}$$

and

$$\text{MISE}\left(\widetilde{\beta}\right) = \sum_{r=0}^{k} W_r \text{MISE}\left(\widetilde{\beta}_r\right), \tag{5.8}$$

respectively, where $W_0, \ldots, W_k$ are known non-negative constants.

5.2. Asymptotic representations

The following assumptions are made throughout this section.

(a) For all $t \in R$, $f(t)$ is continuously differentiable and there are non-negative constants p_r, $r = 0, \ldots, k$, so that $\beta_r(t)$ are $(p_r + 2)$ times continuously differentiable with respect to t.

(b) For all $r, l = 0, \ldots, k$, $E[|X^{(r)}|^4]$ and the $(2+\delta)$th moments of $|\widehat{e}_{rl}|$ are finite for some $\delta > 0$.

(c) The variance and covariance of the error process $\epsilon(t)$ satisfy

$$\sigma^2(t) = E\left[\epsilon^2(t)\right] < \infty \qquad \text{and} \qquad \rho_\epsilon(t) = \lim_{t' \to t} E\left[\epsilon(t)\epsilon(t')\right] < \infty.$$

Furthermore, $\sigma^2(t)$ and $\rho_\epsilon(t)$ are continuous for all $t \in R$.

(d) The kernel function $K_r(\cdot)$ is a compactly supported (p_r+2)th order kernel, which satisfies $\int u^j K_r(u)du = 0$ for all $1 \le j < p_r + 2$,

$$\mu_{(p_r+2)} = \int u^{p_r+2} K_r(u)du < \infty, \quad R(K_r) = \int K_r^2(u)du < \infty$$

and

$$\int K_r(u)du = 1.$$

(e) The bandwidth $h_r > 0$ satisfies $h_r \to 0$ and $nh_r \to \infty$ as $n \to \infty$.

Notice that, in general, $\sigma^2(t) \ne \rho_\epsilon(t)$. The strict inequality between $\sigma^2(t)$ and $\rho_\epsilon(t)$ happens, for example, when $\epsilon_{ij} = s(t_{ij}) + W_i$ where $s(t)$ is a mean zero Gaussian stationary process and W_i is an independent white noise (cf. Zeger and Diggle (1994)). Some of the above assumptions, such as the compactness of the support of K_r and the smoothness conditions of $f(t)$, $\beta_r(t)$, $\sigma^2(t)$ and $\rho_\epsilon(t)$, are merely made for the simplicity of the derivations. Analogous asymptotic results may be derived when these conditions are modified or even weakened. In practice some non-compactly supported kernels, such as the standard Gaussian kernels, can provide equally good estimators as well.

Let $B(\widetilde{\beta}_r(t))$ and $V(\widetilde{\beta}_r(t))$ be the bias and variance of $(f(t))^{-1}\widetilde{R}_r(t)$, respectively. Then by (5.5), we have the decomposition

$$\text{MSE}\left(\widetilde{\beta}_r(t)\right) = B^2\left(\widetilde{\beta}_r(t)\right) + V\left(\widetilde{\beta}_r(t)\right). \tag{5.9}$$

An important fact, which is useful to establish the asymptotic representations of $B(\widetilde{\beta}_r(t))$ and $V(\widetilde{\beta}_r(t))$, is $\widehat{e}_{rl} = e_{rl} + O_p(n^{-1/2})$. For mathematical simplicity and to avoid excessive notation, we only consider the rate $O_p(n^{-1/2})$, but not the exact asymptotic expression, of $(\widehat{e}_{rl} - e_{rl})$ in the derivation of $B(\widetilde{\beta}_r(t))$ and $V(\widetilde{\beta}_r(t))$. Define

$$M_r^{(0)}(t) = \sum_{r_1=0}^{k}\sum_{r_2=0}^{k}\Big\{\beta_{r_1}(t)\beta_{r_2}(t)E\Big[X_{r_1}X_{r_2}\Big(\sum_{l=0}^{k} e_{rl}X_l\Big)^2\Big]\Big\} - \beta_r^2(t),$$

$$M_r^{(1)}(t) = M_r^{(0)}(t) + \sigma^2(t)E\Big[\Big(\sum_{l=0}^{k} e_{rl}X_l\Big)^2\Big],$$

$$M_r^{(2)}(t) = M_r^{(0)}(t) + \rho_\epsilon(t)E\Big[\Big(\sum_{l=0}^{k} e_{rl}X_l\Big)^2\Big],$$

$$Q_{1r}(t) = \mu_{(p_r+2)}\Big[\frac{\beta_r^{(p_r+2)}(t)}{(p_r+2)!} + \frac{\beta_r^{(p_r+1)}(t)f'(t)}{(p_r+1)!f(t)}\Big]$$

and

$$Q_{2r}(t) = (f(t))^{-1}R(K_r)M_r^{(1)}(t).$$

Theorem 5.1. *Suppose that t is in the interior of the support of $f(\cdot)$ and Assumptions* (a) *through* (e) *are satisfied. When n is sufficiently large,*

$$B(\widetilde{\beta}_r(t)) = h_r^{p_r+2} Q_{1r}(t) + o(h_r^{p_r+2}) + O(n^{-1/2}) \tag{5.10}$$

and

$$\begin{aligned} V\left(\widetilde{\beta}_r(t)\right) = h_r^{-1}\Big[\sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big] Q_{2r}(t)(1+o(1)) \\ +\Big[n^{-1} - \sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big] M_r^{(2)}(t)\,(1+o(1)) \\ +B(\widetilde{\beta}_r(t))\,O(n^{-1/2})(1+o(1)). \end{aligned} \tag{5.11}$$

The asymptotic representations of $\mathrm{MSE}(\widetilde{\beta}_r(t))$, $\mathrm{MISE}(\widetilde{\beta}_r)$, $\mathrm{MSE}(\widetilde{\beta}(t))$ *and* $\mathrm{MISE}(\widetilde{\beta})$ *can be obtained by substituting* (5.10) *and* (5.11) *into* (5.9), (5.6), (5.7) *and* (5.8).

Proof. See Appendix A.

Say $\widetilde{\beta}_r(t)$ is a consistent estimator of $\beta_r(t)$ if $\mathrm{MSE}(\widetilde{\beta}_r(t)) \to 0$ as $n \to \infty$. It immediately follows from Theorem 5.1 that $\widetilde{\beta}_r(t)$ is consistent if and only if $h_r \to 0$ and $h_r^{-1}\sum_{i=1}^{n}(n_i^{-1}n^{-2}) \to 0$ as $n \to \infty$. The intra-correlations of the data have no effect on $B(\widetilde{\beta}_r(t))$, but may strongly influence $V(\widetilde{\beta}_r(t))$.

The next result shows that, when the numbers of repeated measurements are bounded, $\widetilde{\beta}_r(t)$ behaves like a Nadaraya-Watson type kernel regression estimator with independent cross-sectional data.

Theorem 5.2. *Suppose that the assumptions of Theorem* 5.1 *are satisfied and the n_i are bounded, i.e., $n_i \le c$ for some $c \ge 1$ and all $i = 1, \ldots, n$. The optimal bandwidths $h_{r,opt}(t)$ and $h_{r,opt}$, which minimize* $\mathrm{MSE}(\widetilde{\beta}_r(t))$ *and* $\mathrm{MISE}(\widetilde{\beta}_r)$, *respectively, for all $h_r > 0$, are given by*

$$h_{r,opt}(t) = \Big[\sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big]^{1/(2p_r+5)} \Big[\frac{Q_{2r}(t)}{2(p_r+2)Q_{1r}^2(t)}\Big]^{1/(2p_r+5)} \tag{5.12}$$

and

$$h_{r,opt} = \Big[\sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big]^{1/(2p_r+5)} \Big[\frac{\int Q_{2r}(s)\pi(s)\,ds}{2(p_r+2)\left(\int Q_{1r}^2(s)\pi(s)\,ds\right)}\Big]^{1/(2p_r+5)}. \tag{5.13}$$

The optimal $\mathrm{MSE}(\widetilde{\beta}_r(t))$ *and* $\mathrm{MISE}(\widetilde{\beta}_r)$ *corresponding to $h_{r,opt}(t)$ and $h_{r,opt}$ are given by*

$$\mathrm{MSE}(\widetilde{\beta}_{r,h_{r,opt}}(t)) = \Big[\sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big]^{\frac{2p_r+4}{2p_r+5}} [Q_{2r}(t)]^{\frac{2p_r+4}{2p_r+5}} [Q_{1r}(t)]^{2/(2p_r+5)}$$

$$\times\Big[(2p_r+4)^{\frac{-2p_r-4}{2p_r+5}}+(2p_r+4)^{1/(2p_r+5)}\Big](1+o(1)) \qquad (5.14)$$

and

$$\begin{aligned}\mathrm{MISE}(\widetilde{\beta}_{r,h_{r,opt}}) &= \Big[\sum_{i=1}^{n}(n_i^{-1}n^{-2})\Big]^{\frac{2p_r+4}{2p_r+5}}\Big[\int Q_{2r}(s)\pi(s)ds\Big]^{\frac{2p_r+4}{2p_r+5}}\\ &\quad\times\Big[\int Q_{1r}^2(s)\pi(s)ds\Big]^{1/(2p_r+5)}\\ &\quad\times\Big[(2p_r+4)^{\frac{-2p_r-4}{2p_r+5}}+(2p_r+4)^{1/(2p_r+5)}\Big](1+o(1)). \qquad (5.15)\end{aligned}$$

Proof. See Appendix B.

Remark 5.1. It can be shown as in (5.1) through (5.4) that

$$(1+o_p(1))\left(\widetilde{\beta}_r^*(t)-\beta_r(t)\right)=(f(t))^{-1}\widetilde{R}_r^*(t),$$

where $\widetilde{R}_r^*(t)$ is defined as $\widetilde{R}_r(t)$ in (5.4), with $1/(nn_i)$ in $\widetilde{m}_r(t)$ and $\widetilde{f}_r(t)$ substituted by $1/N$. The modified bias and variance of $\widetilde{\beta}_r^*(t)$ are then defined by, respectively,

$$B\left(\widetilde{\beta}_r^*(t)\right)=E\left[(f(t))^{-1}\widetilde{R}_r^*(t)\right] \qquad \text{and} \qquad V\left(\widetilde{\beta}_r^*(t)\right)=\mathrm{Var}\left[(f(t))^{-1}\widetilde{R}_r^*(t)\right].$$

The same calculations as in the proof of Theorem 5.1 show that, under the conditions of Theorem 5.1, when n is sufficiently large, $B(\widetilde{\beta}_r^*(t))$ is the same as the right hand side of (5.10), but $V(\widetilde{\beta}_r^*(t))$ is given by

$$\begin{aligned}V\left(\widetilde{\beta}_r^*(t)\right) &= N^{-1}h_r^{-1}Q_{2r}(t)(1+o(1))+N^{-2}\Big[\sum_{i=1}^{n}n_i^2-N\Big]M_r^{(2)}(t)(1+o(1))\\ &\quad+B\left(\widetilde{\beta}_r^*(t)\right)O\left(n^{-1/2}\right)(1+o(1)). \qquad (5.16)\end{aligned}$$

Optimal bandwidths and mean squared errors, analogous to those given in Theorem 5.2, for $\widetilde{\beta}_r^*(t)$, can be derived from (5.10) and (5.16). Since convergence rates of $V(\widetilde{\beta}_r(t))$ and $V(\widetilde{\beta}_r^*(t))$ depend on n_i differently, it is easy to see from (5.11) and (5.16) that neither $\widetilde{\beta}_r(t)$ nor $\widetilde{\beta}_r^*(t)$ asymptotically dominates the other for all situations.

Remark 5.2. Theorem 5.1 implies that, in general, the convergence rates of $\mathrm{MSE}(\widetilde{\beta}_r(t))$ depend on whether and how n_i, $i=0,\ldots,n$, converge to infinity relative to n. In practice, whether the n_i are sufficiently large or not is usually unknown, so that any bandwidth choices that purely rely on minimizing the asymptotic expressions of $\mathrm{MSE}(\widetilde{\beta}_r(t))$ or $\mathrm{MISE}(\widetilde{\beta}_r)$ may not work. Similar conclusions hold for $\widetilde{\beta}_r^*(t)$. The bandwidth and inference procedures suggested in Sections 2.2 and 2.3 only rely on available data, and are free of the asymptotic results that may depend on unrealistic assumptions.

Acknowledgements

Partial support was provided by the National Institute on Drug Abuse grant R01 DA10184-01 for the first author and the National Science Council grant 88-2118-M-029-007 for the second author. We would like to thank an Associate Editor and two referees for helpful comments and suggestions which have greatly improved the presentation of the paper. We are also grateful to Dr. Alfred Saah for providing the MACS Public Use Data Set Release PO4 (1984-1991) and many valuable comments and discussions on HIV epidemiology.

Appendix A
Proof of Theorem 5.1.

By (1.3), we have

$$\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}=\sum_{l_1=0}^{k}\Big[X_i^{(l_1)}\Big(\sum_{l_2=0}^{k}\widehat{e}_{rl_2}X_i^{(l_2)}\Big)\beta_{l_1}(t_{ij})\Big]+\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)\epsilon_{ij}$$

and the obvious identity

$$\beta_r(t)=E\Big\{\sum_{l_1=0}^{k}\Big[X^{(l_1)}\Big(\sum_{l_2=0}^{k}e_{rl_2}X^{(l_2)}\Big)\beta_{l_1}(t)\Big]\Big\},\qquad r=0,\ldots,k. \tag{A.1}$$

Assumption (b) and the definition of $\widehat{e}_{rl}$ imply that

$$\begin{aligned}E\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}\Big|t_{ij}=s\Big]&=\sum_{l_1=0}^{k}E\Big[X_i^{(l_1)}\Big(\sum_{l_2=0}^{k}\widehat{e}_{rl_2}X_i^{(l_2)}\Big)\Big]\beta_{l_1}(s)\\&=\beta_r(s)+O(n^{-1/2}).\end{aligned}$$

Thus it follows from (5.2), (5.3), (5.4), assumption (c) and the change of variables that, when n is sufficiently large,

$$\begin{aligned}&B(\widetilde{\beta}_r(t))\\&=\frac{1}{nh_rf(t)}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\int\frac{1}{n_i}\Big\{E\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}\Big|t_{ij}=s\Big]-\beta_r(t)\Big\}K_r\Big(\frac{t-s}{h_r}\Big)f(s)\,ds\\&=(f(t))^{-1}\int[\beta_r(t-h_ru)-\beta_r(t)]f(t-h_ru)K_r(u)\,du+O(n^{-1/2}).\end{aligned}$$

Then (5.10) follows from the Taylor expansions of $\beta_r(t-h_ru)$ and $f(t-h_ru)$ at $\beta_r(t)$ and $f(t)$, respectively.

For the expression of $V(\widetilde{\beta}_r(t))$, we consider

$$\Big[(f(t))^{-1}\widetilde{R}_r(t)\Big]^2=A_r^{(1)}(t)+A_r^{(2)}(t)+A_r^{(3)}(t),$$

where

$$A_r^{(1)}(t) = (f(t))^{-2}(nh_r)^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{Z_{ijr}^2(t)K_r^2(\frac{t-t_{ij}}{h_r})\Big\},$$

$$A_r^{(2)}(t) = (f(t))^{-2}(nh_r)^{-2}\sum_{i=1}^{n}\sum_{j_1\neq j_2}\Big\{Z_{ij_1r}(t)Z_{ij_2r}(t)K_r\Big(\frac{t-t_{ij_1}}{h_r}\Big)K_r\Big(\frac{t-t_{ij_2}}{h_r}\Big)\Big\},$$

$$A_r^{(3)}(t) = (f(t))^{-2}(nh_r)^{-2}\sum_{i_1\neq i_2}\sum_{j_1,j_2}\Big\{Z_{i_1j_1r}(t)Z_{i_2j_2r}(t)K_r\Big(\frac{t-t_{i_1j_1}}{h_r}\Big)K_r\Big(\frac{t-t_{i_2j_2}}{h_r}\Big)\Big\},$$

$$Z_{ijr}(t) = n_i^{-1}\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)Y_{ij}-\beta_r(t)\Big].$$

By (1.3), straightforward computation shows that

$$n_i^2Z_{ijr}^2(t) = \Big(\xi_{ijr}(t)\Big)^2 + 2\xi_{ijr}(t)\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)\epsilon_{ij} + \Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)^2\epsilon_{ij}^2,$$

where $\xi_{ijr}(t) = \sum_{l_1=0}^{k}\{X_i^{(l_1)}(\sum_{l_2=0}^{k}\widehat{e}_{rl_2}X_i^{(l_2)})\beta_{l_1}(t_{ij})\} - \beta_r(t)$. Since $\mathbf{X}_i$ and ϵ_{ij} are independent, we have that, by (A.1) and the definition of $M_r^{(0)}(t)$,

$$E\Big[\xi_{ijr}(t)\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)\epsilon_i\Big(t_{ij}\Big)\Big|t_{ij}=s\Big] = 0,$$

$$\begin{aligned}E\Big[\Big(\xi_{ijr}(t)\Big)^2\Big|t_{ij}=s\Big] &= E\Big\{\Big[\sum_{l_1=0}^{k}\Big[X_i^{(l_1)}\Big(\sum_{l_2=0}^{k}\widehat{e}_{rl_2}X_i^{(l_2)}\Big)\beta_{l_1}\Big(t_{ij}\Big)\Big]-\beta_r(t)\Big]^2\Big|t_{ij}=s\Big\}\\ &= E\Big\{\Big[\sum_{l_1=0}^{k}\Big[X_i^{(l_1)}\Big(\sum_{l_2=0}^{k}e_{rl_2}X_i^{(l_2)}\Big)\beta_{l_1}(t_{ij})\Big]\Big]^2\Big|t_{ij}=s\Big\}\\ &\quad -2\beta_r(t)\,E\Big\{\sum_{l_1=0}^{k}\Big[X_i^{(l_1)}\Big(\sum_{l_2=0}^{k}e_{rl_2}X_i^{(l_2)}\Big)\beta_{l_1}(t_{ij})\Big]\Big|t_{ij}=s\Big\}\\ &\quad +\beta_r^2(t)+o(1)\\ &= M_r^{(0)}(s)+\beta_r^2(s)-2\beta_r(t)\beta_r(s)+\beta_r^2(t)+o(1),\end{aligned}$$

and

$$E\Big[\Big(\sum_{l=0}^{k}\widehat{e}_{rl}X_i^{(l)}\Big)^2\epsilon_{ij}^2\Big|t_{ij}=s\Big] = \sigma^2(s)E\Big[\Big(\sum_{l=0}^{k}e_{rl}X^{(l)}\Big)^2\Big](1+o(1)).$$

It then follows that

$$E\Big[A_r^{(1)}(t)\Big] = \Big(\frac{1}{nh_rf(t)}\Big)^2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\int E\Big(Z_{ijr}^2(t)\Big|t_{ij}=s\Big)K_r^2\Big(\frac{t-s}{h_r}\Big)f(s)\,ds\Big\}$$

$$= \Big(\frac{1}{nh_r f(t)}\Big)^2 \sum_{i=1}^{n}\sum_{j=1}^{n_i}\Big\{\int \frac{1}{n_i^2}\Big[M_r^{(0)}(s)+\beta_r^2(s)-2\beta_r(t)\beta_r(s)+\beta_r^2(t)$$
$$+\sigma^2(s)E\Big[\Big(\sum_{l=0}^{k} e_{rl}X^{(l)}\Big)^2\Big]\Big]K_r^2\Big(\frac{t-s}{h_r}\Big)f(s)\,ds\Big(1+o(1)\Big)\Big\}$$
$$= (f(t))^{-2}\sum_{i=1}^{n}\Big\{\Big(\frac{1}{n_i^2 n^2 h_r}\Big)\sum_{j=1}^{n_i}\Big[M_r^{(1)}(t)R(K_r)f(t)\Big](1+o(1))\Big\}$$
$$= \frac{1}{h_r}\Big[\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big](f(t))^{-1}R(K_r)M_r^{(1)}(t)(1+o(1)). \tag{A.2}$$

Using similar computations as those for $A_r^{(1)}(t)$, we can show that

$$E\Big[A_r^{(2)}(t)\Big] = \Big[\frac{1}{n} - \sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]M_r^{(2)}(t)(1+o(1)) \tag{A.3}$$

and

$$E\Big[A_r^{(3)}(t)\Big] = \Big\{B\Big(\widetilde{\beta}_r(t)\Big)+O(n^{-1/2})\Big\}^2. \tag{A.4}$$

Then, when n is sufficiently large, (5.11) follows from (A.2), (A.3), (A.4) and

$$V(\widetilde{\beta}_r(t)) = \sum_{l=1}^{3}E[A_r^{(l)}(t)] - B^2(\widetilde{\beta}_r(t))$$
$$= E[A_r^{(1)}(t)] + E[A_r^{(2)}(t)] + B(\widetilde{\beta}_r(t))\,O(n^{-1/2}) + O(n^{-1}).$$

This completes the proof of Theorem 5.1.

Appendix B

Proof of Theorem 5.2. Because $n_1,\ldots,n_n$ are bounded by $c \geq 1$, we have

$$\frac{1}{cn} \leq \sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big) \leq \frac{1}{n} \tag{B.1}$$

and

$$0 \leq \frac{1}{n} - \sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big) \leq \frac{1}{n}(1-c^{-1}). \tag{B.2}$$

When $n^{1/(2p_r+5)}h_r \to 0$ as $n \to \infty$, it follows from (B.1) and (B.2) that, as $n \to \infty$,

$$n^{\frac{2p_r+4}{2p_r+5}} h_r^{2p_r+4} Q_{1r}^2(t) \to 0,$$
$$n^{\frac{2p_r+4}{2p_r+5}} h_r^{-1}\Big[\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]Q_{2r}(t) \geq h_r^{-1} n^{-1/(2p_r+5)} c^{-1} Q_{2r}(t) \to \infty,$$

and

$$n^{\frac{2p_r+4}{2p_r+5}}\Big|\Big[\frac{1}{n}-\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]M_r^{(2)}(t)\Big| \le n^{\frac{2p_r+4}{2p_r+5}}n^{-1}(1-c^{-1})\Big|M_r^{(2)}(t)\Big| \to 0.$$

Thus

$$n^{\frac{2p_r+4}{2p_r+5}}\mathrm{MSE}\Big(\widetilde{\beta}_r(t)\Big) \to \infty, \qquad \text{as } n\to\infty. \tag{B.3}$$

When $n^{1/(2p_r+5)}h_r \to \infty$ as $n\to\infty$, similar calculations show that

$$n^{\frac{2p_r+4}{2p_r+5}}h_r^{2p_r+4}Q_{1r}^2(t)\to\infty, \qquad n^{\frac{2p_r+4}{2p_r+5}}h_r^{-1}\Big[\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]Q_{2r}(t)\to 0$$

and

$$n^{\frac{2p_r+4}{2p_r+5}}\Big|\Big[\frac{1}{n}-\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]M_r^{(2)}(t)\Big| \to 0,$$

so that (B.3) still holds.

It suffices to consider the case that $h_r = n^{-1/(2p_r+5)}c_n$ for some c_n which does not converge to either 0 or ∞ when $n\to\infty$. Since, by (B.2),

$$n^{\frac{2p_r+4}{2p_r+5}}\Big[\frac{1}{n}-\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big] \le n^{-1/(2p_r+5)}\Big(1-c^{-1}\Big) = o(1),$$

(5.9), (5.10), (5.11) and (B.1) imply that

$$n^{\frac{2p_r+4}{2p_r+5}}\mathrm{MSE}\Big(\widetilde{\beta}_r(t)\Big) = c_n^{2p_r+4}Q_{1r}^2(t) + c_n^{-1}n\Big[\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]Q_{2r}(t) + o(1). \tag{B.4}$$

The right hand side of (B.4) is uniquely minimized by

$$c_n = n^{1/(2p_r+5)}\Big[\sum_{i=1}^{n}\Big(\frac{1}{n_i n^2}\Big)\Big]^{1/(2p_r+5)}\Big[\frac{Q_{2r}(t)}{(2p_r+4)Q_{1r}^2(t)}\Big]^{1/(2p_r+5)},$$

which then implies (5.12).

Substituting h_r of (5.10) and (5.11) with (5.12), (5.14) is a direct consequence of (5.9), (5.10) and (5.11). Similar calculations then show that (5.13) and (5.15) also hold.

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.* London, Chapman Hall, New York.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959-971.

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

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

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

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

Hart, T. D. and Wehrly, T. E. (1993). Consistency of cross-validation when the data are curves. *Stochastic Processes and Their Applications* **45**, 351-361.

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. 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.

Jones, R. H. (1987). Serial correlation in unbalanced mixed models. *Bull. Internat. Statist. Inst.* **46**, 105-122.

Jones, R. H. and Ackerson, L. M. (1990). Serial correlation in unequal spaced longitudinal data. *Biometrika* **77**, 721-731.

Jones, R. H. and Boadi-Boteng, F. (1991). Unequally spaced longitudinal data with serial correlation. *Biometrics* **47**, 161-175.

Kaslow, R. A., Ostrow, D. G., Detels, R., Phair, J. P., Polk, B. F. and Rinaldo, C. R. (1987). The multicenter AIDS cohort study: rationale, organization and selected characteristics of the participants. *Amer. J. Epidemiology* **126**, 310-318.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Marron, J. S. and Härdle, W. (1986). Random approximation to an error criterion of nonparametric statistics. *J. Multivariate Anal.* **20**, 91-113.

Pantula, S. G. and Pollock, K. H. (1985). Nested analysis of variance with autocorrelated errors. *Biometrics* **41**, 909-920.

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 estimators. *Multivariate Analysis II*, 25-31. Academic Press, New York.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

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

Ware, J. H. (1985). Linear models for the analysis of longitudinal studies. *Amer. Statist.* **39**, 95-101.

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.

Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, U.S.A.

E-mail: colin@mts.jhu.edu

Department of Statistics, Tunghai University, Taiwan.

E-mail: chiangct@mail.thu.edu.tw

(Received February 1997; accepted August 1999)