Journal of the Korean Statistical Society (2004), **33**: *4, pp 367–379*

QUASI-LIKELIHOOD REGRESSION FOR VARYING COEFFICIENT MODELS WITH LONGITUDINAL DATA†

CHOONGRAK KIM[1], MEESEON JEONG[2], WOOCHUL KIM[3] AND BYEONG U. PARK[3]

ABSTRACT

This article deals with the nonparametric analysis of longitudinal data when there exist possible correlations among repeated measurements for a given subject. We consider a quasi-likelihood regression model where a transformation of the regression function through a link function is linear in time-varying coefficients. We investigate the local polynomial approach to estimate the time-varying coefficients, and derive the asymptotic distribution of the estimators in this quasi-likelihood context. A real data set is analyzed as an illustrative example.

AMS 2000 subject classifications. Primary 62G08; Secondary 62G20.
Keywords. Local quasi-score function, Newton-Raphson iteration, product kernel, varying coefficients.

1. INTRODUCTION

Regression models with longitudinal data have been extensively studied in the literature. Let $Y(t)$ and $\mathbf{X}(t)$ be the real-valued response and the $\mathbb{R}^k$-valued covariate vector, respectively, observed at time t. Suppose there are n subjects, and for the i^{th} subject there are m_i repeated measurements of $(Y(t), \mathbf{X}(t), t)$ over time. Denote by $(Y_{ij}, \mathbf{X}_{ij}, t_{ij})$ the j^{th} observation of $(Y(t), \mathbf{X}(t), t)$ for the i^{th} subject. Let μ_{ij} be the conditional mean of the response Y_{ij} given t_{ij} and $\mathbf{X}_{ij}$.

Received April 2004; accepted July 2004.
†Choongrak Kim was supported by Korea Research Foundation Grant (KRF-99-015-DP0057). Meeseon Jeong was supported by the Brain Korea 21 Projects. Woochul Kim and Byeong U. Park were supported by Korea Research Foundation Grant (KRF-2002-070-C00017).
[1]Department of Statistics, Pusan National University, Busan 609-735, Korea
[2]Radiation Health Research Institute, Seoul 132-703, Korea
[3]Department of Statistics, Seoul National University, Seoul 151-747, Korea

Most analyses with longitudinal data have been done by parametric approaches such as classical or generalized linear models (see, *e.g.*, Pantula and Pollock, 1985; Ware, 1985; Diggle, 1988; Diggle *et al.*, 1994). Nonlinear models with longitudinal data have been discussed in Davidian and Giltinan (1995). Some smoothing methods for $Y(t)$ based on t have been suggested by Hart and Wehrly (1986), Altman (1990) and Hart (1991) to overcome restrictive and unrealistic assumptions in the parametric approach. Semiparametric models have been considered by Moyeed and Diggle (1994), and Zeger and Diggle (1994).

Recently, Wu *et al.* (1998) and Hoover *et al.* (1998) suggested the linear time-varying coefficient model of the form

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta}(t_{ij}) + \varepsilon_i(t_{ij}) \tag{1.1}$$

where $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_k(\cdot))^T$ is a vector of unknown real-valued functions and $\varepsilon_i(t)$ are mean 0 stochastic process. They established pointwise asymptotic normality for a Nadaraya-Watson type estimator of $\boldsymbol{\beta}(\cdot)$. Also, they constructed confidence regions based on Bonferroni's adjustments.

In this paper we extend the model (1.1) to the quasi-likelihood regression model. We assume that there exists a link function g such that

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta}(t_{ij}) \tag{1.2}$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m_i$. Write $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im_i})^T$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{im_i})$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})^T$ and $\mathbf{t}_i = (t_{i1}, \ldots, t_{im_i})^T$ for $i = 1, \ldots, n$. We assume that we can model the conditional variance matrix as

$$\mathrm{Var}(\mathbf{Y}_i | \mathbf{t}_i, \mathbf{X}_i) = \sigma^2 \mathbf{V}_i(\boldsymbol{\mu}_i) \tag{1.3}$$

for $i = 1, \ldots, n$ where $\mathbf{V}_i(\cdot)$'s are given variance functions and σ^2 is an unknown constant.

The model specified by (1.2) and (1.3) is also an extension of the quasi-likelihood approach for varying coefficient models with cross-sectional data as considered by Cai *et al.* (2000). It is more general than the varying coefficient model for longitudinal data with ordinal response considered in Kauermann (2000), which is a special case of the present model because it relies on the assumption that the repeated measurements within each subject are independent too. In the estimation procedure we accommodate possible correlations within subjects in the form of (1.2), which introduces the variance-covariance terms of the repeated measurements, and consequently requires different estimation process. This is also in contrast with the approaches made by Wu *et al.* (1998) and

Hoover *et al.* (1998) where correlations within subjects are ignored in their estimation procedures. Other related works include least squares type estimation for the present model (Lin and Ying, 2001), and dynamic additive regression models for longitudinal data (Martinussen and Scheike, 2000).

In Section 2, we introduce the local quasi-score function to estimate $\boldsymbol{\beta}(\cdot)$ and demonstrate asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}(\cdot)$. Also, we discuss estimation of variance function $\mathbf{V}_i$. In Section 3, an example based on a real data set is given.

2. INFERENCE AND ASYMPTOTICS

2.1. Local quasi-score function

For a real-valued function $a(\cdot)$ defined on $\mathbb{R}$ and for $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$, we write $a(\mathbf{x}) = (a(x_1), \ldots, a(x_d))^T$. Define

$$\mathbf{q}_i(\mathbf{z}_i, \mathbf{y}_i) = \mathbf{D}_i(\mathbf{z}_i)\mathbf{V}_i^{-1}(g^{-1}(\mathbf{z}_i))(\mathbf{y}_i - g^{-1}(\mathbf{z}_i))$$

where $\mathbf{z}_i = (z_{i1}, \ldots, z_{im_i})^T$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{im_i})^T \in \mathbb{R}^{m_i}$, and $\mathbf{D}_i(\mathbf{z}_i)$ is the $m_i \times m_i$ diagonal matrix with $\{g'(g^{-1}(z_{ik}))\}^{-1}$ as its k^{th} diagonal entry, and g' denotes the first derivative of g. If we model the conditional mean μ_{ij} as $g(\mu_{ij}) = \mathbf{X}_{ij}^T\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is an unknown vector, the quasi-score function for estimating $\boldsymbol{\beta}$ is given by $\sum_{i=1}^n \mathbf{X}_i\mathbf{q}_i(\mathbf{X}_i^T\boldsymbol{\beta}, \mathbf{Y}_i)$. Similarly, the quasi-score function for estimating $\boldsymbol{\beta}(\cdot)$ under the model (1.2) with (1.3) is given by

$$\sum_{i=1}^n \mathbf{X}_i\mathbf{q}_i(g(\boldsymbol{\mu}_i), \mathbf{Y}_i). \tag{2.1}$$

We consider a localized version of (2.1). It is based on an approximation of $\mathbf{X}_{ij}^T\boldsymbol{\beta}(t_{ij})$:

$$\mathbf{X}_{ij}^T\boldsymbol{\beta}(t_{ij}) \approx \mathbf{X}_{ij}^T \sum_{r=0}^{p} \frac{(t_{ij} - t_0)^r}{r!}\boldsymbol{\beta}^{(r)}(t_0) \tag{2.2}$$

for the t_{ij}'s which are close to t_0, where $\boldsymbol{\beta}^{(r)}(t_0) = (\beta_1^{(r)}(t_0), \ldots, \beta_k^{(r)}(t_0))^T$. Write $\boldsymbol{\gamma}_r = \boldsymbol{\beta}^{(r)}(t_0)/r!$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_0^T, \ldots, \boldsymbol{\gamma}_p^T)^T$, $\mathbf{z}_{ij}(t_{ij}) = (1, (t_{ij} - t_0), \ldots, (t_{ij} - t_0)^p)^T \otimes \mathbf{X}_{ij}$ and $\mathbf{Z}_i(\mathbf{t}_i) = (\mathbf{z}_{i1}(t_{i1}), \ldots, \mathbf{z}_{im_i}(t_{im_i}))$, where $\otimes$ denotes the Kronecker product (Searle, 1982). Then we can write the approximation given at (2.2) by $\mathbf{z}_{ij}(t_{ij})^T\boldsymbol{\gamma}$.

We construct a local quasi-score function from (2.1) by placing more emphasis on those observations at t_{ij}'s which are close to t_0 and using the approximation given at (2.2) for those t_{ij}'s. We define

$$\mathbf{q}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \mathbf{Z}_i(\mathbf{t}_i)\mathbf{q}_i(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma}, \mathbf{Y}_i)\mathcal{K}_i(\mathbf{t}_i, h) \tag{2.3}$$

where $\mathcal{K}_i(\mathbf{t}_i, h) = h^{-m_i}\prod_{j=1}^{m_i} K(h^{-1}(t_{ij} - t_0))$ and K is a given symmetric probability density function. The parameter $h > 0$ is usually called the bandwidth. Therefore, the local quasi-score function can be expressed as

$$\mathbf{q}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \mathbf{Z}_i(\mathbf{t}_i)\mathbf{D}_i\left(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma}\right)\mathbf{V}_i^{-1}\left(g^{-1}(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma})\right) \\ \times\left(\mathbf{Y}_i - g^{-1}(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma})\right)\mathcal{K}_i(\mathbf{t}_i, h),$$

and the local quasi-likelihood estimator (QLE) of $\beta^{(r)}(t_0)$ is then given by $r!\hat{\boldsymbol{\gamma}}_r$ where $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\gamma}}_0^T, \ldots, \hat{\boldsymbol{\gamma}}_p^T)^T$ is a solution of the local quasi-likelihood estimating equation $\mathbf{q}(\boldsymbol{\gamma}) = \mathbf{0}$. In general, the estimating equation $\mathbf{q}(\boldsymbol{\gamma}) = \mathbf{0}$ does not have a close form solution, and therefore, we use the Newton-Raphson iteration, *i.e.*,

$$\hat{\boldsymbol{\gamma}}^{(k+1)} = \hat{\boldsymbol{\gamma}}^{(k)} + \left[\mathbf{J}(\hat{\boldsymbol{\gamma}}^{(k)})\right]^{-1}\mathbf{q}(\hat{\boldsymbol{\gamma}}^{(k)}), \quad k = 0, 1, 2, \ldots$$

where $\hat{\boldsymbol{\gamma}}^{(k)}$ is the k^{th} iterated estimator of $\boldsymbol{\gamma}$ and

$$\mathbf{J}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \mathbf{Z}_i(\mathbf{t}_i)\mathbf{D}_i\left(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma}\right)\mathbf{V}_i^{-1}\left(g^{-1}(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma})\right)\mathbf{D}_i\left(\mathbf{Z}_i(\mathbf{t}_i)^T\boldsymbol{\gamma}\right) \\ \times\mathbf{Z}_i(\mathbf{t}_i)^T\mathcal{K}_i(\mathbf{t}_i, h)$$

is the Fisher information matrix.

2.2. Asymptotics

In this section, we derive the asymptotic distributions of the local QLE $\hat{\beta}^{(r)}(t_0) = r!\hat{\boldsymbol{\gamma}}_r$, $r = 0, 1, \ldots, p$ at a fixed point $t_0 \in \mathbb{R}$. We assume that the design points t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, m_i$ are random and *iid* according to an underlying density f. We also assume that t_0 is an interior point on support of f.

Denote $\kappa_r = \int u^r K(u)du$, $\nu_r = \int u^r K^2(u)du$ for $r = 0, 1, \ldots, 2p$, and $\mathbf{H} = \text{diag}\{1, h, \ldots, h^p/p!\} \otimes \mathbf{I}_k$ where $\mathbf{I}_k$ is $k \times k$ identity matrix. Define $\mathbf{S}_1 = (\kappa_{p+1}, \ldots, \kappa_{2p+1})^T$ and $\mathbf{S}_2 = \kappa_{p+1}(\kappa_0, \ldots, \kappa_p)^T$. And let $\mathbf{M}_1$ and $\mathbf{M}_2$ be $(p+1) \times (p+1)$ matrices having the $(r, s)^{th}$ entry equal to κ_{r+s} and $\kappa_r\kappa_s$, $0 \le r, s \le p$, respectively. Also define $\mathbf{N}_{i1}$ and $\mathbf{N}_{i2}$ are $(p+1)\times(p+1)$ matrices having the $(r,s)^{th}$ entry equal to $\nu_0^{m_i-1}\nu_{r+s}$ and $\nu_0^{m_i-2}\nu_r\nu_s$, $0\le r,s\le p$, respectively. Let $\rho_i(\mathbf{t}_i,\mathbf{X}_i)=\mathbf{D}_i(g(\boldsymbol{\mu}_i))\mathbf{V}_i^{-1}(\boldsymbol{\mu}_i)\mathbf{D}_i(g(\boldsymbol{\mu}_i))$ and $\rho_{j,k}^{(i)}$ be its $(j,k)^{th}$ element. Define $\Gamma_{j,k}^{(i)}(\mathbf{t}_i)=E[\rho_{j,k}^{(i)}\mathbf{X}_{ij}\mathbf{X}_{ik}^T|\mathbf{t}_i]$, $j,k=1,\ldots,m_i$. Let $c_n=n\delta_n$ and $\delta_n=(\sum_{i=1}^n h^{-m_i})^{-1/2}$. Define

$$\boldsymbol{\Lambda}_n=\sigma^2\delta_n^2\sum_{i=1}^n h^{-m_i}f(t_0)^{m_i}\left\{\mathbf{N}_{i1}\otimes\mathbf{G}_{i1}+\mathbf{N}_{i2}\otimes\mathbf{G}_{i2}\right\},\tag{2.4}$$

$$\boldsymbol{\Delta}_n=n^{-1}\sum_{i=1}^n f(t_0)^{m_i}\left\{\mathbf{M}_1\otimes\mathbf{G}_{i1}+\mathbf{M}_2\otimes\mathbf{G}_{i2}\right\},\tag{2.5}$$

where $\mathbf{G}_{i1}=\sum_{j=1}^{m_i}\Gamma_{j,j}^{(i)}(t_0\mathbf{1}_i)$, and $\mathbf{G}_{i2}=\sum_{j\neq k}\Gamma_{j,k}^{(i)}(t_0\mathbf{1}_i)$ with $\mathbf{1}_i$ being $m_i\times 1$ vector all entries of which equal one. The following theorem states the asymptotic distribution of $\boldsymbol{\beta}^{(r)}$'s when n tends to infinity.

THEOREM 1. *Suppose that the conditions (C1) $\sim$ (C6) in Appendix hold. Then,*

$$\begin{aligned}&c_n\boldsymbol{\Lambda}_n^{-1/2}\boldsymbol{\Delta}_n\Bigg(\mathbf{H}\cdot\mathrm{vec}\left[\hat{\boldsymbol{\beta}}^{(r)}(t_0)-\boldsymbol{\beta}^{(r)}(t_0),\ r=0,1,\ldots,p\right]\\&-\boldsymbol{\Delta}_n^{-1}\frac{h^{p+1}}{(p+1)!}\frac{1}{n}\sum_{i=1}^n f(t_0)^{m_i}(\mathbf{S}_1\otimes\mathbf{G}_{i1}+\mathbf{S}_2\otimes\mathbf{G}_{i2})\boldsymbol{\beta}^{(p+1)}(t_0)\{1+o(1)\}\Bigg)\\&\xrightarrow{\mathcal{D}}\ N(\mathbf{0},\mathbf{I}_{k(p+1)})\end{aligned}\tag{2.6}$$

as n tends to infinity, where vec *denotes the vec operation (Searle, 1982).*

2.3. Estimation of variance function

When we solve the local quasi-likelihood estimating equation $\mathbf{q}(\boldsymbol{\gamma})=\mathbf{0}$, the variance function $\mathbf{V}_i$ should be given. If we have no information on $\mathbf{V}_i$, we have to estimate it. Several methods for estimating $\mathbf{V}_i$ have been suggested, and here we take Liang and Zeger (1986)'s suggestion among them. Note that $\mathbf{V}_i$ can be written as

$$\mathbf{V}_i(\boldsymbol{\mu}_i)=\phi\mathbf{A}_i\mathbf{R}_i\mathbf{A}_i$$

where ϕ is a dispersion parameter, $\mathbb{A}_i=\mathrm{diag}\{(\mu_{ij}^*(1-\mu_{ij}^*))^{1/2},\ j=1,\ldots,m_i\}$, $\mu_{ij}^*=g^{-1}(\mathbf{z}_{ij}(t_{ij})^T\boldsymbol{\gamma})$, and $\mathbf{R}_i=\mathrm{Corr}\{Y_{ij},\ j=1,\ldots,m_i\}$ is a working correlation matrix to be estimated. As an estimator of $\mathbf{R}_i$, Liang and Zeger (1986) suggested a geometrically decreasing correlation, *i.e.*,

$$\hat{\mathbf{R}}_i = \{\hat{\tau}^{|t_{ij}-t_{is}|},\ j,s=1,\dots,m_i\}$$

where $\hat{\tau}$ can be estimated as follows: Let

$$r_{js} = \{n-k(p+1)\}^{-1}\hat{\phi}^{-1}\sum_{l=1}^{n}\bar{y}_{lj}\bar{y}_{ls},\ \ j,s=1,\dots,m_i$$

where

$$\bar{y}_{ij} = (y_{ij}-\mu^*_{ij})\{\mu^*_{ij}(1-\mu^*_{ij})\}^{-1/2}$$

and

$$\hat{\phi} = \left\{\sum_{i=1}^{n} m_i - k(p+1)\right\}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\bar{y}^2_{ij}.$$

Then, $\log\hat{\tau}$ is the slope after fitting a straight line on the data $\{|t_{ij}-t_{is}|,\ \log r_{js}\}$.

3. Example

As an illustration example, we consider prostate cancer data collected at the University of Chicago hospital. The response variable is genito-urinary (bladder) toxicities associated with radiation therapy of prostate cancer. The assessment was made on the ordinal scale (no symptoms, pain/local bleeding, bleeding lesion, and serious lesion), however, we treat them as binary data (0 for no symptoms, 1 for others). Covariates considered are X_1 = hospital (1 and 2), X_2 = race (white and black), X_3 = age of the patient at the timepoint of therapy, X_4 = stage of prostate cancer (minor, medium, and severe), and X_5 = total dose of radiation (weak, medium, and strong). These data are already analyzed by Kauermann (2000) using the varying coefficients model. However, his method is suitable for the case where the repeated measurements within each subject are independent. Our method allows existence of correlations between replications.

The original data consist of $n = 196$ observations. Here we take $n = 70$ observations with $m_i \geq 10$, because a subject with small replications can make trouble in computing the product kernel. The Epanechnikov kernel is used, and the bandwidth h is selected to minimize the Pearson's χ^2, *i.e.*,

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{\{y_{ij}-\hat{\mu}_{ij}(t_{ij})\}^2}{\hat{\mu}_{ij}(t_{ij})(1-\hat{\mu}_{ij}(t_{ij}))}$$

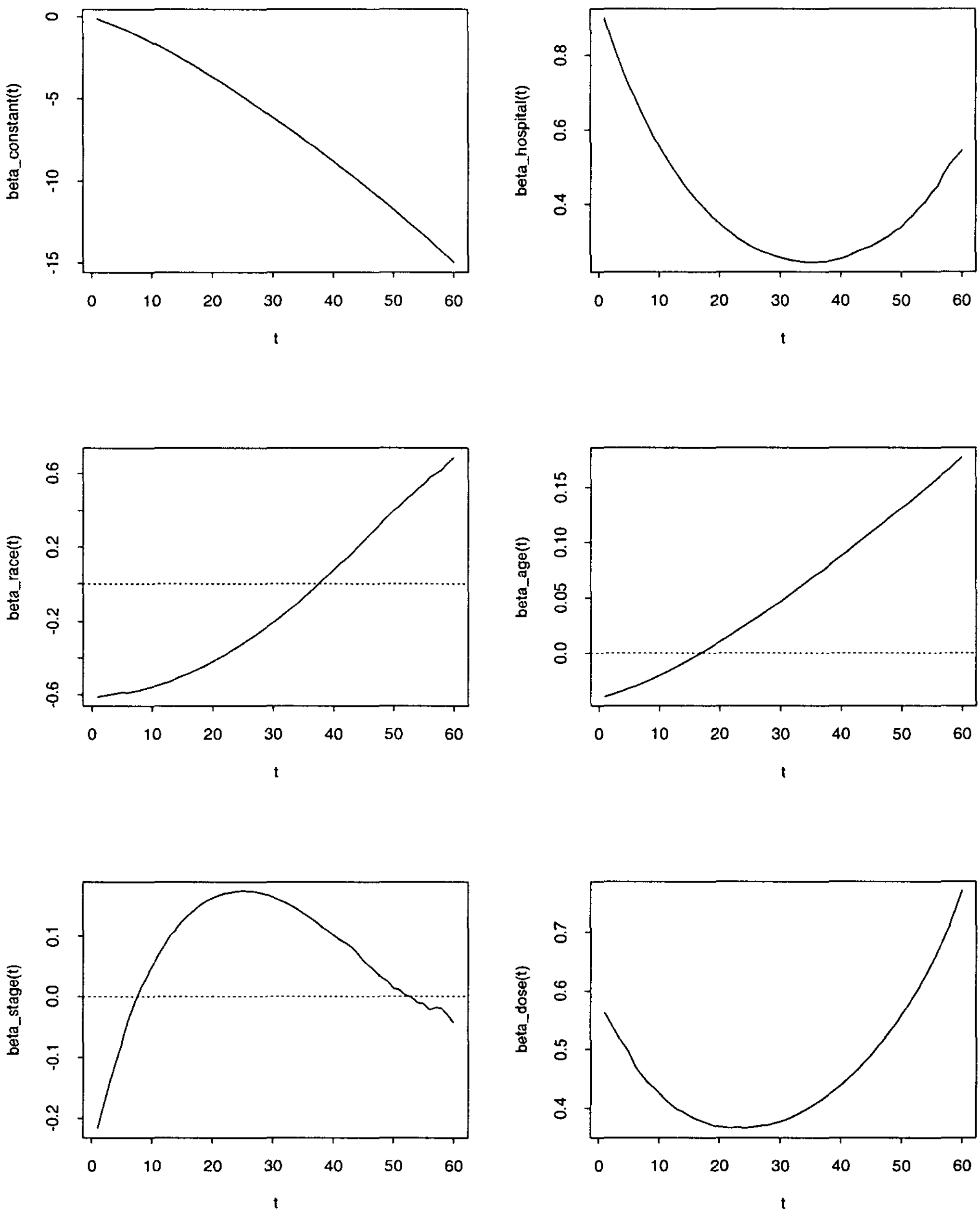

FIGURE 3.1 *Varying coefficients for radiation data*

where $\hat{\mu}_{ij}(t_{ij}) = g^{-1}(\mathbf{X}_{ij}^T\hat{\beta}(t_{ij}))$, $g(\cdot) = \text{logit}(\cdot)$. It turned out $h = 75$. We can use the cross-validation criterion to select h, but it is computationally expensive. The issue on choosing the smoothing parameter h in the varying coefficient model is not an easy problem, and it requires further research. The variance function $\mathbf{V}_i$ is computed using the method described in Section 2.3. Figure 3.1 shows the estimated regression coefficients $\hat{\beta}_0(t)$, $\hat{\beta}_1(t)$, $\hat{\beta}_2(t)$, $\hat{\beta}_3(t)$, $\hat{\beta}_4(t)$ and $\hat{\beta}_5(t)$. As a reference value, the zero line is given. It is clear that all the coefficients are different from constant $\hat{\beta}_i$, $i = 0, \ldots, 5$, say. This figure shows why the varying coefficient model should be used instead of the constant coefficient model. Effects of dose and stage are almost the same as Kauermann (2000)'s. However, hospital effect is different from Kauermann's results. He considered only dose, stage, and hospital as covariates, while we included two more covariares: race and age.

APPENDIX : PROOF OF THEOREM 1

A.1. The conditions

(C1) The density $f(\cdot)$ is continuous at the point $t = t_0$, and $f(t_0) > 0$.

(C2) The variance functions $\mathbf{V}_i$'s are twice continuously differentiable, and the link function g is three times continuously differentiable.

(C3) $K(\cdot)$ is supported on $[-1, 1]$.

(C4) $\beta_j^{(p+1)}(\cdot)$ is continuous in a neighborhood of t_0 for $j = 1, \ldots, k$.

(C5) $h \to 0$ and $n^{-2}\sum_{i=1}^n h^{-m_i} \to 0$ as $n \to \infty$.

(C6) $\mathbf{\Lambda}_n$ and $\mathbf{\Delta}_n$ converge to some nonsigular matrices as $n \to \infty$.

A.2. Proof of Theorem 1

Note that $\hat{\gamma}$ is a solution of $\mathbf{q}(\gamma) = \mathbf{0}$. Let $\hat{\boldsymbol{\theta}}_r = c_n h^r(\hat{\gamma}_r - \beta^{(r)}(t_0)/r!)$, $r = 0, 1, \ldots, p$, and $\beta^*(t) \equiv \beta^*(t, t_0) = \sum_{r=0}^p \beta^{(r)}(t_0)(t - t_0)^r/r!$. Define $\mathbf{z}_{ij}^*(t_{ij}) = (1, (t_{ij} - t_0)/h, \ldots, (t_{ij} - t_0)^p/h^p)^T \otimes \mathbf{X}_{ij}$ and form $\mathbf{Z}_i^*(\mathbf{t}_i) = (\mathbf{z}_{i1}^*(t_{i1}), \ldots, \mathbf{z}_{im_i}^*(t_{im_i}))$. Let $\mu_{ij}^* = g^{-1}(\mathbf{X}_{ij}^T\beta^*(t_{ij}))$ and $\boldsymbol{\mu}_i^* = (\mu_{i1}^*, \ldots, \mu_{im_i}^*)^T$. Then we can write

$$\mathbf{Z}_i(\mathbf{t}_i)^T\hat{\gamma} = g(\boldsymbol{\mu}_i^*) + c_n^{-1}\mathbf{Z}_i^*(\mathbf{t}_i)^T\hat{\boldsymbol{\theta}}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_0^T, \ldots, \hat{\boldsymbol{\theta}}_p^T)^T$. Therefore $\hat{\boldsymbol{\theta}}$ is a solution of the estimating equation $\mathbf{l}_n(\boldsymbol{\theta}) = \mathbf{0}$ where

$$\mathbf{l}_n(\boldsymbol{\theta}) = \delta_n \sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\, \mathbf{q}_i \left(g(\boldsymbol{\mu}_i^*) + c_n^{-1}\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta}, \mathbf{Y}_i\right) \mathcal{K}_i(\mathbf{t}_i, h). \tag{A.1}$$

For a function $\boldsymbol{\xi} : \mathbb{R}^d \to \mathbb{R}^d$, write $\boldsymbol{\xi}(\mathbf{x}) \equiv (\xi_1(\mathbf{x}), \ldots, \xi_d(\mathbf{x}))^T$ and let $(\partial/\partial\mathbf{x}^T)\times \boldsymbol{\xi}(\mathbf{x})$ denote the $d \times d$ matrix whose $(i,j)^{th}$ entry equals $(\partial/\partial x_j)\xi_i(\mathbf{x})$. Write also $(\partial^2/\partial\mathbf{x}\partial\mathbf{x}^T)\boldsymbol{\xi}(\mathbf{x})$ for the $d^2 \times d$ matrix consisting of $d \times d$ submatrices $(\partial^2/\partial\mathbf{x}\partial\mathbf{x}^T)\xi_1(\mathbf{x}), \ldots, (\partial^2/\partial\mathbf{x}\partial\mathbf{x}^T)\xi_d(\mathbf{x})$ where $(\partial^2/\partial\mathbf{x}\partial\mathbf{x}^T)\xi_l(\mathbf{x})$ has as its $(i,j)^{th}$ entry $(\partial^2/\partial x_i \partial x_j)\xi_l(\mathbf{x})$. Let

$$\mathbf{q}_i'(\mathbf{z}_i, \mathbf{y}_i) = \frac{\partial}{\partial \mathbf{z}_i^T}\mathbf{q}_i(\mathbf{z}_i, \mathbf{y}_i), \text{ and } \mathbf{q}_i''(\mathbf{z}_i, \mathbf{y}_i) = \frac{\partial^2}{\partial \mathbf{z}_i \partial \mathbf{z}_i^T}\mathbf{q}_i(\mathbf{z}_i, \mathbf{y}_i).$$

Using the Taylor expansion of $\mathbf{q}_i(\cdot, \mathbf{y}_i)$, we have from (A.1)

$$\begin{aligned} \mathbf{l}_n(\boldsymbol{\theta}) &= \delta_n \sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathcal{K}_i(\mathbf{t}_i, h) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i'(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta}\mathcal{K}_i(\mathbf{t}_i, h) \\ &\quad + \frac{1}{2}\delta_n c_n^{-2}\sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\{\mathbf{I}_{m_i} \otimes (\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta})\}^T \\ &\qquad \times \mathbf{q}_i''(\mathbf{g}_i, \mathbf{Y}_i)\{\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta}\}\mathcal{K}_i(\mathbf{t}_i, h) \end{aligned}$$

where $\mathbf{g}_i$ is between $g(\boldsymbol{\mu}_i^*)$ and $g(\boldsymbol{\mu}_i^*) + \mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta}$. Let

$$\begin{aligned} \mathbf{W}_n &= \delta_n \sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathcal{K}_i(\mathbf{t}_i, h), \\ \mathbf{A}_n &= -\frac{1}{n}\sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i'(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathbf{Z}_i^*(\mathbf{t}_i)^T\mathcal{K}_i(\mathbf{t}_i, h), \\ \mathbf{R}_n(\boldsymbol{\theta}) &= \frac{1}{2}\delta_n c_n^{-2}\sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\{\mathbf{I}_{m_i} \otimes (\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta})\}^T \\ &\qquad \times \mathbf{q}_i''(\mathbf{g}_i, \mathbf{Y}_i)\{\mathbf{Z}_i^*(\mathbf{t}_i)^T\boldsymbol{\theta}\}\mathcal{K}_i(\mathbf{t}_i, h). \end{aligned}$$

Then, (A.1) becomes

$$\mathbf{l}_n(\boldsymbol{\theta}) = \mathbf{W}_n - \mathbf{A}_n\boldsymbol{\theta} + \mathbf{R}_n(\boldsymbol{\theta}). \tag{A.2}$$

Note that $\mathbf{q}_i(\mathbf{z}_i, \mathbf{y}_i)$, $\mathbf{q}_i'(\mathbf{z}_i, \mathbf{y}_i)$ and $\mathbf{q}_i''(\mathbf{z}_i, \mathbf{y}_i)$ are linear in $\mathbf{y}_i$ for fixed $\mathbf{z}_i$. Using the fact that $\delta_n c_n^{-2} = o(n^{-1})$ one may show that for any $C > 0$

$$\sup_{|\boldsymbol{\theta}| \le C} |\mathbf{R}_n(\boldsymbol{\theta})| \to 0 \tag{A.3}$$

in probability. Note that

$$\begin{aligned} g(\boldsymbol{\mu}_i) &= g(\boldsymbol{\mu}_i^*) + \left[(t_{i1} - t_0)^{p+1}\mathbf{X}_{i1}, \ldots, (t_{im_i} - t_0)^{p+1}\mathbf{X}_{im_i}\right]^T \\ &\quad \times \frac{\boldsymbol{\beta}^{(p+1)}(t_0)}{(p+1)!} + o_p(h^{p+1}). \end{aligned} \tag{A.4}$$

for t_{ij}'s such that $|t_{ij} - t_0| < h$. By (A.4), we obtain

$$\mathbf{q}_i'(g(\boldsymbol{\mu}_i^*), \boldsymbol{\mu}_i) = -\rho_i(\mathbf{t}_i, \mathbf{X}_i) + o_p(h^p).$$

The mean of $\mathbf{A}_n$ equals

$$E(\mathbf{A}_n) = \frac{1}{n}\sum_{i=1}^n E\left[\mathbf{Z}_i^*(\mathbf{t}_i)\rho_i(\mathbf{t}_i, \mathbf{X}_i)\mathbf{Z}_i^*(\mathbf{t}_i)^T \mathcal{K}_i(\mathbf{t}_i, h)\right]\{1 + o(1)\},$$

and thus the $(r,s)^{th}$ block of $E(\mathbf{A}_n)$ is

$$\begin{aligned} E(\mathbf{A}_n)_{r,s} &= \frac{1}{n}\sum_{i=1}^n E\left\{\sum_{j=1}^{m_i}\sum_{k=1}^{m_i}\left(\frac{t_{ij} - t_0}{h}\right)^r \left(\frac{t_{ik} - t_0}{h}\right)^s \rho_{j,k}^{(i)} \mathbf{X}_{ij}\mathbf{X}_{ik}^T \mathcal{K}_i(\mathbf{t}_i, h)\right\} \\ &\qquad \times \{1 + o(1)\} \\ &= \frac{1}{n}\sum_{i=1}^n f(t_0)^{m_i}\left\{\kappa_{r+s}\mathbf{G}_{i1} + \kappa_r\kappa_s\mathbf{G}_{i2}\right\}\{1 + o(1)\} \end{aligned}$$

for $0 \le r, s \le p$. Therefore,

$$\begin{aligned} E(\mathbf{A}_n) &= \frac{1}{n}\sum_{i=1}^n f(t_0)^{m_i}\left\{\mathbf{M}_1 \otimes \mathbf{G}_{i1} + \mathbf{M}_2 \otimes \mathbf{G}_{i2}\right\}\{1 + o(1)\} \\ &= \boldsymbol{\Delta}_n\{1 + o(1)\} \end{aligned}$$

where $\boldsymbol{\Delta}_n$ is as given at (2.5). Similarly, we can show

$$\mathrm{Var}(\mathbf{A}_n) = O\Big(n^{-2}\sum_{i=1}^n h^{-m_i}\Big).$$

By the conditions (C5) and (C6) we obtain $\mathbf{A}_n = \boldsymbol{\Delta}_n + o_p(1)$, and

$$\mathbf{l}_n(\boldsymbol{\theta}) = \mathbf{W}_n - \boldsymbol{\Delta}_n\boldsymbol{\theta} + \mathbf{R}_{n1}(\boldsymbol{\theta}) \tag{A.5}$$

where $\sup_{|\boldsymbol{\theta}|\le C} |\mathbf{R}_{n1}(\boldsymbol{\theta})|$ converges to zero in probability. Thus, we get the following stochstic expansion for $\hat{\boldsymbol{\theta}}$:

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\Delta}_n^{-1}\mathbf{W}_n + o_p(1). \tag{A.6}$$

Next, we calculate the mean vector and the covariance matrix of $\mathbf{W}_n$. By using (A.4), the mean of $E(\mathbf{W}_n)$ equals

$$\begin{aligned} E(\mathbf{W}_n) &= \delta_n h^{p+1} \sum_{i=1}^{n} E\Bigg[\mathbf{Z}_i^*(\mathbf{t}_i)\rho_i(\mathbf{t}_i, \mathbf{X}_i) \\ &\quad \times \left\{ \left(\frac{t_{i1}-t_0}{h}\right)^{p+1} \mathbf{X}_{i1}, \ldots, \left(\frac{t_{im_i}-t_0}{h}\right)^{p+1} \mathbf{X}_{im_i} \right\}^T \\ &\quad \times \frac{\boldsymbol{\beta}^{(p+1)}(t_0)}{(p+1)!} \mathcal{K}_i(\mathbf{t}_i, h)\Bigg] \{1+o(1)\}. \end{aligned}$$

The r^{th} block of $E(\mathbf{W}_n)$ is

$$\begin{aligned} E(\mathbf{W}_n)_r &= \delta_n h^{p+1} \sum_{i=1}^{n} E\Bigg\{ \sum_{j=1}^{m_i}\sum_{k=1}^{m_i} \left(\frac{t_{ij}-t_0}{h}\right)^r \left(\frac{t_{ik}-t_0}{h}\right)^{p+1} \\ &\quad \times \rho_{j,k}^{(i)} \mathbf{X}_{ij}\mathbf{X}_{ik}^T \mathcal{K}_i(\mathbf{t}_i, h) \Bigg\} \frac{\boldsymbol{\beta}^{(p+1)}(t_0)}{(p+1)!} \{1+o(1)\} \\ &= \delta_n h^{p+1} \sum_{i=1}^{n} f(t_0)^{m_i} \{\kappa_{r+p+1}\mathbf{G}_{i1} + \kappa_r\kappa_{p+1}\mathbf{G}_{i2}\} \\ &\quad \times \frac{\boldsymbol{\beta}^{(p+1)}(t_0)}{(p+1)!} \{1+o(1)\} \end{aligned}$$

for $0 \le r \le p$. Therefore, we have

$$\begin{aligned} E(\mathbf{W}_n) &= \delta_n h^{p+1} \sum_{i=1}^{n} f(t_0)^{m_i} \{\mathbf{S}_1 \otimes \mathbf{G}_{i1} + \mathbf{S}_2 \otimes \mathbf{G}_{i2}\} \\ &\quad \times \frac{\boldsymbol{\beta}^{(p+1)}(t_0)}{(p+1)!} \{1+o(1)\}. \end{aligned} \tag{A.7}$$

For the variance of $\mathbf{W}_n$, write $\mathbf{W}_n\mathbf{W}_n^T = \mathbf{B}_n + \mathbf{C}_n$ where

$$\begin{aligned} \mathbf{B}_n &= \delta_n^2 \sum_{i=1}^{n} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathbf{q}_i^T(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathbf{Z}_i^*(\mathbf{t}_i)^T \mathcal{K}_i^2(\mathbf{t}_i, h), \\ \mathbf{C}_n &= \delta_n^2 \sum_{i \ne i'} \mathbf{Z}_i^*(\mathbf{t}_i)\mathbf{q}_i(g(\boldsymbol{\mu}_i^*), \mathbf{Y}_i)\mathbf{q}_{i'}^T(g(\boldsymbol{\mu}_{i'}^*), \mathbf{Y}_{i'})\mathbf{Z}_{i'}^*(\mathbf{t}_{i'})^T \mathcal{K}_i(\mathbf{t}_i, h)\mathcal{K}_{i'}(\mathbf{t}_{i'}, h). \end{aligned}$$

By independence between different subjects, we can easily show that $E(\mathbf{C}_n) = (E\mathbf{W}_n)(E\mathbf{W}_n)^T\{1+o(1)\}$. Similar arguments show that

$$
\begin{aligned}
E(\mathbf{B}_n) &= \sigma^2\delta_n^2 \sum_{i=1}^{n} E\left\{\mathbf{Z}_i^*(\mathbf{t}_i)\rho_i(\mathbf{t}_i,\mathbf{X}_i)\mathbf{Z}_i^{*T}(\mathbf{t}_i)\mathcal{K}_i^2(\mathbf{t}_i,h)\right\}\{1+o(1)\} \\
&= \sigma^2\delta_n^2 \sum_{i=1}^{n} h^{-m_i} f(t_0)^{m_i}\left\{\mathbf{N}_{i1}\otimes\mathbf{G}_{i1}+\mathbf{N}_{i2}\otimes\mathbf{G}_{i2}\right\}\{1+o(1)\}.
\end{aligned}
$$

Therefore, we obtain $\mathrm{Var}(\mathbf{W}_n) = \mathbf{\Lambda}_n\{1+o(1)\}$ where $\mathbf{\Lambda}_n$ is as defined at (2.4). This together with (A.6), (A.7) and the fact that $\mathrm{Var}(\mathbf{W}_n)^{-1/2}\{\mathbf{W}_n - E(\mathbf{W}_n)\}$ converges in distribution to $N(\mathbf{0},\mathbf{I}_{k(p+1)})$ complete the proof of the theorem.

REFERENCES

ALTMAN, N. S. (1990). "Kernel smoothing of data with correlated errors", *Journal of the American Statistical Association*, **85**, 749–759.

CAI, Z., FAN, J. AND LI, R. (2000). "Efficient estimation and inferences for varying-coefficient models", *Journal of the American Statistical Association*, **95**, 888–902.

DAVIDIAN, M. AND GILTINAN, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, London.

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.

HART, J. D. (1991). "Kernel regression estimation with time series errors", *Journal of the Royal Statistical Society*, **B53**, 173–187.

HART, J. D. AND WEHRLY, T. E. (1986). "Kernel regression estimation using repeated measurements data", *Journal of the American Statistical Association*, **81**, 1080–1088.

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.

KAUERMANN, G. (2000). "Modeling longitudinal data with ordinal response by varying coefficients", *Biometrics*, **56**, 692–698.

LIANG, K. Y. AND ZEGER, S. L. (1986). "Longitudinal data analysis using generalized linear models", *Biometrika*, **73**, 13–22.

LIN, D. Y. AND YING, Z. (2001). "Semiparametric and nonparametric regression analysis of longitudinal data (with discussion)", *Journal of the American Statistical Association*, **96**, 103–126.

MARTINUSSEN, T. AND SCHEIKE, T. H. (2000). "A nonparametric dynamic additive regression model for longitudinal data", *The Annals of Statistics*, **28**, 1000–1025.

MOYEED, R. A. AND DIGGLE, P. J. (1994). "Rates of convergence in semi-parametric modelling of longitudinal data", *The Australian Journal of Statistics*, **36**, 75–93.

PANTULA, S. G. AND POLLOCK, K. H. (1985). "Nested analysis of variance with autocorrelated errors", *Biometrics*, **41**, 909–920.

SEARLE, S. R. (1982). *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.

WARE, J. H. (1985). "Linear models for the analysis of longitudinal studies", *The American Statistician*, **39**, 95–101.

WU, C. O., CHIANG, C. T. AND HOOVER, D. R. (1998). "Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data", *Journal of the American Statistical Association*, **93**, 1388–1402.

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.