A NONPARAMETRIC DYNAMIC ADDITIVE REGRESSION MODEL FOR LONGITUDINAL DATA

BY TORBEN MARTINUSSEN AND THOMAS H. SCHEIKE[1]

The Royal Veterinary and Agricultural University and University of Copenhagen

In this work we study additive dynamic regression models for longitudinal data. These models provide a flexible and nonparametric method for investigating the time-dynamics of longitudinal data. The methodology is aimed at data where measurements are recorded at random time points. We model the conditional mean of responses given the full internal history and possibly time-varying covariates. We derive the asymptotic distribution for a new nonparametric least squares estimator of the cumulative time-varying regression functions. Based on the asymptotic results, confidence bands may be computed and inference about time-varying coefficients may be drawn. We propose two estimators of the cumulative regression function. One estimator that involves smoothing and one that does not. The latter, however, has twice the variance as the smoothing based estimator. Goodness of fit of the model is considered using martingale residuals. Finally, we also discuss how partly-conditional mean models in which the mean of the response is regressed onto selected time-varying covariates may be analysed in the same framework. We apply the methods to longitudinal data on height development for cystic fibrosis patients.

1. Introduction. In bio-medical research one often deals with responses and covariates obtained over time for independent subjects. This type of longitudinal data is encountered in many studies, for example when interest is on a treatment effect on health status. There has been important progress for parametric problems with the GEE approach of Liang and Zeger (1986). Murphy and Li (1995) considered estimating equations and Scheike (1994) dealt with conditional least squares for longitudinal data in the general framework for longitudinal data presented here.

Nonparametric methods provide a more flexible tool for longitudinal data analysis. Scheike and Zhang (1998) considered smooth nonparametric regression models to study effects of covariates on the longitudinal response. Their approach may, however, require very large sample sizes and is not suited for a situation with more than a few covariates. The results from such an analysis may further be difficult to interpret. In this work we present a new nonparametric method for studying the effects of several covariates on the longitudinal
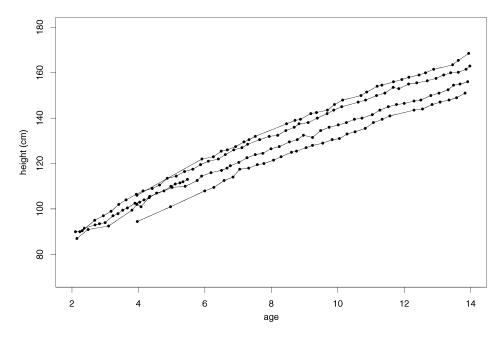
FIG. 1. *The growth of 5 randomly chosen patients with cystic fibrosis.*

response. We impose further structure through an underlying additive time-varying coefficients regression model.

We present a motivating example where the growth of cystic fibrosis patients is monitored over a long period of time; see Laursen et al. (1999) for details. Figure 1 shows the growth of 5 randomly chosen patients as a function of age.

The objective is to see whether the growth has changed over calendar time. The data consist of approximately 250 patients that have had their height measured longitudinally. For these data, responses for different subjects were recorded at random times rather than being observed at fixed times, this is often the case for longitudinal data collected in bio-medical research. Interest now centers on whether the growth of the patients has improved over calendar time due to changes in treatment and in particular whether the magnitude of any improvement in growth over calendar time depends on the age of the patients. To answer this question it is natural to consider conditional models for the current response given what has been previously observed and to allow the regression coefficients to depend on age.

The general and flexible class of regression models investigated in this paper are such time varying coefficients models in which the mean response of the quantity of interest is given as a linear combination of possible time-varying covariates and time-varying coefficients:

$$E(Z_t \,|\, \text{Covariates at time } t) = \beta_0(t) + X_1(t)\beta_1(t) + \cdots + X_q(t)\beta_q(t)$$

where Z_t is the response measured at time t, $\beta_1(t), \ldots, \beta_q(t)$ are the time-varying coefficients and $X_1(t), \ldots, X_q(t)$ the covariates. The covariates may represent two types of information collected over time. The internal information about the history of the longitudinal data such as previous responses and the time at which these responses were obtained; and the external information about the development of external covariates that contain information about the response in question.

   This type of model has been studied extensively in survival analysis over the years and is known as Aalen's additive hazard model; see Aalen (1980, 1989), McKeague (1988), Huffer and McKeague (1991) and Andersen, Borgan, Gill and Keiding (1993) for an overview. In the case of cross-sectional data, the time varying regression model has been described by a number of authors in the regression setting. Hastie and Tibshirani (1993) proposed various ways of fitting these models and derived conditions under which their back-fitting algorithm will converge. The theoretical properties of the estimation procedure is however not well known. Recently, Fan and Zhang (1999) and Hoover, Rice, Wu and Yang (1998) considered local regression techniques.

   Existing nonparametric methods for longitudinal data are typically based on smoothing techniques focusing on estimation of the regression functions directly. We take a completely different approach by estimating the cumulative regression functions $\int_0^t \beta_j(s)ds$. It turns out that these quantities are much easier to estimate than is the case for the regression functions. It is further possible to give a satisfactory description of the large sample properties of the estimators of the cumulative regression functions. We extend the methods used in survival analysis to longitudinal data exploiting the fact that longitudinal data with random measurement times may be viewed as a marked point process. This allows us to draw on martingale calculus, which in turn enables us to provide a detailed large sample theory for the suggested new least squares estimators of the cumulative regression functions. We may then compute confidence bands and draw inference about the regression functions based on the asymptotic results. When the distribution of the measurement times is known, the estimator of the cumulative regression function needs no smoothing. Typically, however, the distribution of the measurement times is unknown and then a smoothing parameter is needed to estimate this quantity. The estimator is similar to Aalen's least squares estimator for Aalen's additive intensity model in survival analysis. Even for an unknown distribution of the measurement times, we suggest an alternative nonparametric least squares estimators that does not involve any smoothing, but this estimator has twice the variance of the smoothing based estimator. The estimator where smoothing is avoided may, however, be useful as an initial estimator since the choice of the smoothing parameter is then not needed.

   When specifying the above model in a longitudinal data setting the covariates in the above mean value may represent knowledge collected over time. Pepe and Couper (1997) draw attention to the distinction between a fully conditional model where the above conditional mean is equal to the conditional mean given the entire history of the process as well as external covariates

and partly conditional mean models where the mean is regressed onto only selected covariates. These selected covariates are a subset of the covariates necessary to describe the full conditional mean. Aalen's additive model is an example of a fully conditional model since the intensity is given the entire history. Pepe and Couper study the partly conditional models fitting the parameters of the model by GEE [Liang and Zeger (1986)] and derive the asymptotic distribution under regularity conditions. We consider also these issues for the nonparametric time-varying additive regression model.

The paper is structured as follows: Section 2 contains a description of the model in a stochastic process framework (marked point process). Section 3 presents the nonparametric least squares estimators of the cumulative regression functions and in Section 4 we describe their asymptotic properties which may be used for inference. Section 5 deals with evaluation of "goodness of fit" for the dynamic regression model. Section 6 considers partly specified conditional models. Section 7 contains the application to height development for cystic fibrosis patients, and finally, Section 8 contains a discussion.

2. A nonparametric additive model for longitudinal data. In this section, we present a dynamic conditional model for the development of responses observed longitudinally. The model takes its basis around modeling of the conditional mean of the current response and is as such a moment based approach. We have the situation in mind where the responses are recorded at measurement times that are not too regular and can be described by a counting process. Therefore it makes sense to look for continuous time description of the development of the conditional mean of the response variable.

To fix ideas, let data for the ith subject, $i = 1, \ldots, n$, consist of

$$(2.1) \qquad\qquad (T_i^k, Z_i^k, X_i(t))$$

where T_i^k is the time-point for the kth measurement Z_i^k of the longitudinal variable and $X_i(t)$ is a time-dependent piecewise constant or deterministic (given past information) covariate ($q \times 1$) associated with the ith subject. The covariates can reflect internal information such as the time since the previous measurement and the previous level of response as well as external information in terms of other covariate information such as gender and treatment. Extensions to non-predictable covariates are possible along the lines of Murphy and Li (1995) and Murphy (1995). In the application, T_i^k is the age for the ith subject at its kth measurement; Z_i^k is the height velocity on log-scale (see Section 7); and $X_i(t)$ is a covariate vector containing information about the previous height measurement, the time since the previous height measurement and calendar time.

We assume that we have n subjects from a generic model which we describe in the following. The description is based on the notion of marked point processes, see Brémaud [(1981), Chapter 8] for a more detailed exposition of this theory. The (T_i^k, Z_i^k) constitutes a marked point process, T_i^k being the kth jump time and Z_i^k the associated mark. The marks for the ith process

(Z_i^k) take their values in the measurable space $(E, \mathscr{E})$ referred to as the mark space. To each $A_i \in \mathscr{E}$, we associate a point process

$$N_i(t)(A_i) = \sum_{k \geq 1} 1_{(Z_i^k \in A_i)} 1_{(T_i^k \leq t)},$$

and let $N_i(t) = N_i(t)(E)$, $i = 1, \ldots, n$, denote the basic point processes. The marked point processes can also be identified by their respective induced counting measure $p_i(ds \times dz_i)$ defined by

$$p_i((0, t] \times A_i) = N_i(t)(A_i), \quad A_i \in \mathscr{E}.$$

We consider a history $\mathscr{F}_t = \mathscr{F}_t^1 \vee \cdots \vee \mathscr{F}_t^n$ such that

$$\mathscr{F}_t^i \supset \mathscr{F}_t^{p^i}, \qquad t \geq 0,$$

where $\mathscr{F}_t^{p^i}$ is the history generated by the ith marked point process and further assume that the processes $X_i(t)$, $i = 1, \ldots, n$, are $\mathscr{F}_t$-predictable, locally bounded and piecewise constant or deterministic given past information, that is,

$$X_i(t) = \begin{cases} f_k(t, X_i^0, T_i^1, X_i^1, Z_i^1, \ldots, T_i^k, X_i^k, Z_i^k), & \text{for } T_i^k < t < T_i^{k+1}, \\ g_k(t, X_i^0, T_i^1, X_i^1, Z_i^1, \ldots, T_i^k, X_i^k), & \text{for } t = T_i^k, \end{cases}$$

where $X_i^0, \ldots, X_i^k$ reflect external information, and f_k and g_k are deterministic functions. The $p_i(ds \times dz_i)$ admits the $(P, \mathscr{F}_t)$-intensity kernel $\lambda_t^i(dz_i) = \lambda_t^i \Phi_t^i(dz_i)$, $i = 1, \ldots, n$, that is, $\lambda_t^i \Phi_t^i(A_i)$ is the intensity of the point process $N_i(t)(A_i)$, $A_i \in \mathscr{E}$.

Denote the history

$$\sigma\{T_i^l, Z_i^l, 1 \leq l \leq k - 1; T_i^k; (X_i(t) : t \leq T_i^k)\}$$

of all observations up to and including T_i^k by $\mathscr{F}_{T_i^k-}^i$. We then have

$$\Phi_{T_i^k}^i(A_i) = P\Big(Z_i^k \in A_i \,\Big|\, \mathscr{F}_{T_i^k-}^i\Big),$$

that is, Φ_t^i is the conditional mark distribution given the past up to and including the time-point where the mark is obtained. The additive mean-model is now given by assuming that

$$m_i(t) = \int_E z_i \Phi_t^i(dz_i) = \beta_0(t) + \beta_1(t) X_{i1}(t) + \cdots + \beta_q(t) X_{iq}(t)$$

where $\beta_0(t), \ldots, \beta_q(t)$ are unspecified locally integrable time-dependent regression functions. The variance function $\int_E (z_i - m_i(t))^2 \Phi_t^i(dz_i)$ is assumed to be independent of i and denoted by $\sigma^2(t)$.

An alternative specification of the mean-model is

$$(2.2) \qquad Z_i^k = \beta_0(T_i^k) + \beta_1(T_i^k) X_{i1}(T_i^k) + \cdots + \beta_q(T_i^k) X_{iq}(T_i^k) + \varepsilon_i^k$$

with $E(\varepsilon_i^k \,|\, \mathscr{F}_{T_i^k-}^i) = 0$ and $V(\varepsilon_i^k \,|\, \mathscr{F}_{T_i^k-}^i) = \sigma^2(T_i^k)$.

The intensity λ_t^i for the ith subject, that is the intensity of $N_i(t)$, is assumed to have a multiplicative structure [Aalen (1975, 1978)] as follows:

$$\lambda_t^i = Y_i(t)\alpha(t)$$

where $\alpha(t)$ is an unknown deterministic function and $Y_i(t)$ is a stochastic process that only depends on the past of the observations (predictable). In this paper, we assume that $Y_i(t)$ is a simple indicator variable, indicating whether the ith individual is under observation just before time t. We are then able to deal with delayed entry and right censoring in a similar way as is done in survival analysis, see Andersen et al. (1993).

We end this section with some notation which we will use in the coming sections. Given marked point processes $p_i(dt \times dz)$ and a function $H_i(t, z)$ of time and the responses that may depend on observations prior to t, we use the notation

$$\int_0^t \int_E H_i(s, z_i) p_i(ds \times dz_i) = \sum_{k=1}^{\infty} H_i(T_i^k, Z_i(T_i^k)) I(T_i^k \leq t)$$

referring to the former as a marked point process integral. Collecting the n marked point processes $p(dt \times dz) = (p_1(dt \times dz_1), \ldots, p_n(dt \times dz_n))^T$ and considering $n \times r$ matrices $H(t, z) = (H_{ij}(t, z))$ and $K(t, z) = (K_{ij}(t, z))$ with $z = (z_1, \ldots, z_n)$, we define the jth entry in the $r \times 1$ matrix

$$\int_0^t \int_E H(s, z)^T p(ds \times dz)$$

by

$$\sum_{i=1}^{n} \sum_{k=1}^{\infty} H_{ij}(T_i^k, Z_i(T_i^k))^T I(T_i^k \leq t)$$

and the (j, l) element of the $r \times r$ matrix

$$\int_0^t \int_E H(s, z)^T \text{diag}(p(ds \times dz)) K(s, z)$$

by

$$\sum_{i=1}^{n} \sum_{k=1}^{\infty} H_{ij}(T_i^k, Z_i(T_i^k)) K_{il}(T_i^k, Z_i(T_i^k)) I(T_i^k \leq t).$$

3. Estimation. Trying to estimate the regression functions directly is rather complicated, see Hastie and Tibshirani (1993). We focus instead on estimating the cumulative regression functions $B_j(t) = \int_0^t \beta_j(s)\,ds$ by least squares estimation. Estimates of the regression functions may then be obtained by smoothing techniques. In applications, inferential issues about the regression functions may often be formulated in terms of the cumulative regression functions and then our approach will provide the possibility for formal tests.

Consider the processes $H_i(t, z_i) = z_i$ and let $B(t) = (B_0(t), \ldots, B_q(t))^T$. Then the sum of the responses occurring before time t can be written as a marked point process integral where we integrate $H_i(t, z_i)$ with respect to $p_i(ds \times dz_i)$. The sum of the responses thus gives the following martingale decomposition (see Proposition 1 in the Appendix):

$$\int_0^t \int_E z_i \, p_i(ds \times dz_i) = \int_0^t \alpha(s) Y_i(s) m_i(s) \, ds + M_i(z_i)(t),$$

for $i = 1, \ldots, n$, where $M_i(z_i)(t)$ is a marked point process martingale.

Collecting these n equations in one vector equation, we obtain

$$(3.3) \qquad \int_0^t \int_E D(z) \, p(ds \times dz) = \int_0^t \alpha(s) Y(s) \, dB(s) + M(z)(t)$$

where $z = (z_1, \ldots, z_n)$, $D(z) = \text{diag}(z)$, $p(dt \times dz) = (p_1(dt \times dz_1), \ldots, p_n(dt \times dz_n))^T$, $M(z)(t) = (M_1(z_1)(t), \ldots, M_n(z_n)(t))^T$ and $Y(t) = (Y_{ij}(t))$ is the $n \times (q+1)$-matrix with ith row, $i = 1, \ldots, n$, given by

$$Y_i(t)(1, X_{i1}(t), \ldots, X_{iq}(t)).$$

Writing (3.3) in differential form

$$\int_E D(z) \, p(dt \times dz) = \alpha(t) Y(t) \, dB(t) + dM(z)(t),$$

where $E(dM(z)(t) \mid \text{Past observations}) = 0$, motivates least squares estimators for $B(t)$ of the form

$$(3.4) \qquad \hat{B}(t) = \int_0^t \int_E \frac{J(s)}{\hat{\alpha}(s)} Y^-(s) D(z) \, p(ds \times dz)$$

where $\hat{\alpha}(t)$ is an estimate of $\alpha(t)$, $Y^-(t)$ is a predictable generalized inverse of $Y(t)$, that is, a $(q+1) \times n$-matrix satisfying

$$Y^-(t) Y(t) = I_{(q+1)},$$

the $(q+1) \times (q+1)$ identity matrix and $J(t) = I(Y(t)$ has full rank and $\hat{\alpha}(t) > 0)$. One choice of generalized inverse is

$$(Y(t)^T Y(t))^{-1} Y(t)^T$$

which is based on an unweighted least squares principle. This choice is inefficient, but simple to compute and will be used throughout this paper. An asymptotically efficient estimator is available using a weighted least squares inverse as suggested by McKeague (1988) and Huffer and McKeague (1991) for the Aalen additive hazard model.

Using (3.4) we need to specify an estimate of $\alpha(t)$. For simplicity we suggest to use a kernel smoothing estimate, that is, we let

$$\hat{\alpha}(t) = \frac{1}{b_n} \int K\left(\frac{t-s}{b_n}\right) d\hat{A}(s)$$

where $\hat{A}(t) = \int_0^t \frac{1}{Y(s)} dN(s)$, $Y(t) = \sum_i Y_i(t)$, $N(t) = \sum_i N_i(t)$, K is a bounded kernel function having integral 1 and support $[-1, 1]$ and b_n is the bandwidth parameter. This is the Ramlau-Hansen (1983) kernel estimator. Note that the intensity estimator will be biased and that this bias will cause bias for the estimator (3.4). We suggest to undersmooth the intensity to avoid serious bias and return to this issue in the next section.

The estimator just presented is based on adjusting for the intensity of observations through a smoothing based estimator of the hazard $\alpha(t)$. We now present an alternative estimator that does not involve any smoothing. Denote the jump times of $N(t)$ by $\tau_k = \inf\{t \geq 0 | N(t) = k\}$. The alternative estimator of $B(t)$ is

$$\tilde{B}(t) = \int_0^t \int_E J(s)(s - \tau_{N(s-)})Y(s)Y^-(s)D(z)\, p(ds \times dz),$$

the idea being that

$$(\tau_k - \tau_{k-1}) \approx (Y(t)\alpha(t))^{-1}$$

for $t \in [\tau_{k-1}, \tau_k[$.

It may be shown that Theorem 1 is valid for this estimator also, except that the variance of the estimator is twice the variance of $\hat{B}(t)$, see an unpublished University of Copenhagen technical report by Martinussen and Scheike (1998) for a proof. The increase in variance is due to the fact that the distance to the nearest jump time is a much more variable estimator of the reciprocal of the intensity than a smoothing based estimate. In Scheike (2000) a similar approach for estimation of the cumulative regression function for the standard one-sample cross-sectional nonparametric regression set-up is considered.

Based on the cumulative regression functions which we estimated rather easily, estimates of the regression functions may be found by smoothing. A consistent estimator of $\beta_j(t)$, say, is given by

$$\hat{\beta}_j(t) = \int K\left(\frac{t-s}{w_j}\right) d\hat{B}_j(s)$$

where w_j is an appropriate bandwidth for smoothing $\hat{B}_j$. A more detailed study of this estimator is beyond the scope of this paper.

4. Large sample results. In this section we show that the least squares estimator described in the previous section is asymptotically Gaussian with a variance which we may estimate consistently. The result derived is based on Proposition 1 (see the Appendix) for marked point processes and martingale convergence results and is somewhat similar to the asymptotic analysis for Aalen's additive regression model provided by McKeague (1988) and Huffer and McKeague (1991).

We start by defining various functions depending on the design which are important for the asymptotic analysis. We suppose that the study interval is

$[0, S]$ where $S < \infty$. Define for $j, k, l, m = 0, \ldots, q, t \in [0, S]$,

$$R_j^{(1)}(t) = \sum_{i=1}^{n} Y_{ij}(t),$$

$$R_{jk}^{(2)}(t) = \sum_{i=1}^{n} Y_{ij}(t) Y_{ik}(t),$$

$$R_{jkl}^{(3)}(t) = \sum_{i=1}^{n} Y_{ij}(t) Y_{ik}(t) Y_{il}(t),$$

$$R_{jklm}^{(4)}(t) = \sum_{i=1}^{n} Y_{ij}(t) Y_{ik}(t) Y_{il}(t) Y_{im}(t).$$

Let $t \in [0, S]$ and assume that for $j, k, l, m = 0, \ldots, q$ there exists continuous functions $r_j^{(1)}$, $r_{jk}^{(2)}$, $r_{jkl}^{(3)}$ and $r_{jklm}^{(4)}$ such that as $n \to \infty$,

(A)

$$\sup_{t \in [0,S]} \left| \frac{1}{n} R_j^{(1)}(t) - r_j^{(1)}(t) \right| \xrightarrow{p} 0,$$

$$\sup_{t \in [0,S]} \left| \frac{1}{n} R_{jk}^{(2)}(t) - r_{jk}^{(2)}(t) \right| \xrightarrow{p} 0,$$

$$\sup_{t \in [0,S]} \left| \frac{1}{n} R_{jkl}^{(3)}(t) - r_{jkl}^{(3)}(t) \right| \xrightarrow{p} 0,$$

$$\sup_{t \in [0,S]} \left| \frac{1}{n} R_{jklm}^{(4)}(t) - r_{jklm}^{(4)}(t) \right| \xrightarrow{p} 0.$$

Assume further that for $h = 1, 2$, $j = 0, \ldots, q$,

(B)
$$G_h^{(n)} \overset{\text{def}}{=} \frac{1}{\sqrt{n}} \left(\sup_{\substack{t \in [0,S] \\ i=1,\ldots,n}} |Y_{ij}(t)| \right)^h \xrightarrow{p} 0 \qquad \text{as } n \to \infty$$

and (C) that the matrix $r^{(2)}(t)$ is nonsingular for all $t \in [0, S]$. We further assume (D) that the fourth central moment of the mark distribution (response) $\eta(t) = \int_E (z_i - m_i(t))^4 \Phi_t^i(dz_i)$ is independent of i and that $\sup_{t \in [0,S]} \eta(t) < \infty$. See Section 2 for a definition of the mark distribution $\Phi_t^i(dz_i)$. Finally, we assume that there exist a positive continuous function $y(t)$ such that

(E)
$$\sup_{t \in [0,S]} \left| \frac{1}{n} Y_{\cdot}(t) - y(t) \right| \xrightarrow{p} 0 \qquad \text{as } n \to \infty.$$

Using the central limit theorem for martingales, we can prove the following result.

THEOREM 1.   *Assume that conditions* (A), (B), (C), (D) *and* (E) *hold,* $\alpha(t)$ *is continuous differentiable and bounded away from zero,* $\hat{\alpha}(t)$ *is uniformly*

consistent and $b_n \to 0$, $\sqrt{n} b_n^2 \to 0$. Then

$$\sqrt{n}(\hat{B} - B) \xrightarrow{\mathscr{D}} U \qquad as\ n \to \infty$$

in $D[0, S]^{p+1}$, where U is a mean-zero Gaussian martingale with

$$
\begin{aligned}
\text{cov}&(U_j(t), U_k(t))\\
&= \sum_{f,g,l,m=0}^{p} \int_0^t \frac{\beta_f(s)\beta_g(s)}{\alpha(s)}(r^{(2)}(s))_{jl}^{-1}(r^{(2)}(s))_{km}^{-1} r_{fglm}^{(4)}(s)\, ds\\
&\quad - \int_0^t \frac{\beta_j(s)\beta_k(s)}{\alpha(s)y(s)}\, ds + \int_0^t \frac{\sigma^2(s)}{\alpha(s)}(r^{(2)}(s))_{jk}^{-1}\, ds.
\end{aligned}
$$

(4.5)

The proof of Theorem 1 is given in the Appendix.

REMARKS. (i) The Ramlau-Hansen kernel estimator, $\hat{\alpha}(t)$, is uniformly consistent when $nb_n^2 \to \infty$; see Ramlau-Hansen (1983).

(ii) The covariance function (4.5) is consistently estimated by the estimated optional variation process

$$\hat{\Sigma}(t) = n \int_0^t \int_E \frac{J(s)}{\hat{\alpha}^2(s)} H(z, s) \text{diag}(p(ds \times dz)) H(z, s)^T$$

where $H(z, t) = Y^-(t)(D(z) - \frac{1}{Y.(t)} Y(t)\hat{\beta}(t)a)$ with a the $1 \times n$-vector $(1, \ldots, 1)$.

(iii) Suppose that we have i.i.d replicates, that is, the $((p_i, Y_i, X_i), i = 1, \ldots, n)$ are assumed to be i.i.d. A sufficient condition for conditions (A) and (B) to hold is then

$$E\left(\sup_{t \in [0,S]} |Y_{ij}^4(t)|\right) < \infty$$

for $j = 0, \ldots, p$. Furthermore, the (j, l)-element of the matrix $r^{(2)}$ is

$$r_{jl}^{(2)}(t) = E(Y_{ij}(t)Y_{il}(t)),$$

and it is thus seen that condition (C) corresponds to a linear independence condition on the covariates. For a similar remark on Aalen's additive hazard model, see Andersen et al. [(1993), page 578].

(iv) In the one-sample situation, that is without any covariates, the two first terms in (4.5) cancel out leaving us with the following asymptotic variance

$$\int_0^t \frac{\sigma^2(s)}{\alpha(s)y(s)}\, ds,$$

see Scheike and Zhang (1998) for a similar result for the nonparametric regression set-up.

(v) The ideal bandwidth $b_{n,\mathrm{opt}}$ (for a kernel smoother), the one that balances the effect of the "squared bias term" and the "variance term" of the mean integrated squared error term in an optimal way, is of order $n^{-1/5}$. To ensure that $\sqrt{n}b_n^2 \to 0$, we have to require that b_n converges faster to zero than $b_{n,\mathrm{opt}}$, that is we have to undersmooth compared to the optimal situation.

(vi) Note that the results may be applied to regression models where responses are transformed by a transformation that depends only on past observations.

(vii) The theorem may be extended to predictable variance functions $(\sigma^2(\cdot))$ under further regularity conditions that ensure the convergence of the predictable variation process and the Lindeberg condition in the martingale convergence theorem. The presented estimators of the cumulative regression functions and their variance functions remain valid.

If one is unwilling to choose a smoothing parameter b_n, the alternative estimator $\tilde{B}(t)$ may be used. An estimator of the variance for $\tilde{B}(t)$ is given by its optional variation process, that is equivalent to the expression in remark (ii) with $1/\hat{\alpha}^2(t)$ replaced by $(t - \tau_{N.(t-)})^2 Y_.^2(t)$

Apart from the simplicity of estimation, one further benefit of direct estimation of the cumulative regression functions is that inference may be based on the cumulative regression functions whose asymptotic distribution are described by Theorem 1.

Considering one component of the cumulative regression functions, $B_j(t)$, say, one may now wish to test if this component is equal to zero. This is done easily through the above Theorem 1. Let $\hat{\Sigma}_{jj}(t)$ denote element (j, j) of $\hat{\Sigma}(t)$. A test statistic is

$$\sqrt{n}\,\frac{\hat{B}_j(S)}{\sqrt{\hat{\Sigma}_{jj}(S)}}$$

which is asymptotically standard normal. This test-statistic has been discussed in the nonparametric regression setting by Scheike and Zhang (1998) and Scheike (2000); see also McKeague and Zhang (1994) for a similar approach for time-series analysis.

Alternatively, when the studied regression functions are not consistently positive or negative one may use a maximal deviation test statistic of the process

$$\xi_j(t) = \hat{B}_j(t)(\hat{\Sigma}_{jj}(S))^{1/2}(\hat{\Sigma}_{jj}(t) + \hat{\Sigma}_{jj}(S))^{-1}$$

which converges towards a time-changed Brownian bridge. Therefore

$$\sup_{t \in [0, S]} |\xi_j(t)|$$

converges towards the supremum of the limit distribution, see Huffer and McKeague (1991) for details and Hall and Wellner (1980) for quantiles of the limit distribution.

5. Goodness of fit. We suggest to use martingale residual processes plots to investigate the goodness of fit of the model given by (2.2). Aalen (1993) considered similar plots for the additive hazard regression model and we adapt his suggestion to the current situation. In the marked point process set-up for longitudinal data, however, the suggested martingale residuals can only reveal whether the responses are well modelled by the conditional means given the covariates. As this is the primary object of interest we focus only on this part of the model validation. Note, as pointed out in Remark (vii) of Theorem 1, that the description of the variance function is not crucial for the estimates of the cumulative regression function and their variance functions to be correct. The model for the counting process that describes the measurement times, however, will influence how to estimate the conditional mean values. Scheike (1997) gave an example where extreme observations of the response made new visits at a clinic more likely, thus increased the intensity of the counting process. We return to this issue in the discussion. Many techniques for investigating the goodness of fit for counting process models are available and we refer to Andersen et al. (1993) for further details on such techniques.

The processes in this section are only considered on $[0, S^*]$ where $S^* = \inf\{t \geq 0 : Y(t) \text{ loses full rank}\}$ is a stopping time (preserving martingale properties). Let

$$(5.6) \qquad M_{\mathbf{res}}(t) = \int_0^t \int_E (I - Y(s)Y^-(s))D(z)\,p(ds \times dz).$$

We can rewrite $M_{\mathbf{res}}(t)$ as

$$M_{\mathbf{res}}(t) = \int_0^t (I - Y(s)Y^-(s))\alpha(s)Y(s)\,dB(s) + \int_0^t \int_E (I - Y(s)Y^-(s))\,dM(z)(s)$$

$$= \int_0^t \int_E (I - Y(s)Y^-(s))\,dM(z)(s)$$

showing that $M_{\mathbf{res}}(t)$ is in fact a martingale. To investigate the linearity for a given covariate one can group it into r groups and cumulate the martingale residuals for each group. When linearity is fulfilled all r entries are martingales. Formally, we transform $M_{\mathbf{res}}(t)$ into a $r \times 1$ vector where r is the number of groups for the given covariate. The jth entry in this vector is the sum of $M_{\mathbf{res},i}(t)$ for those i where the covariate at time t belongs to the jth group. Let K_t be a $r \times n$-matrix defining this transformation, that may depend on time if the covariates are time-varying and let

$$M_{\mathbf{res}}^K(t) = \int_0^t \int_E K_s(I - Y(s)Y^-(s))\,dM(z)(s).$$

A martingale residual plot is then obtained by plotting $M_{\mathbf{res}}^K(t)$ against time t. If the model provides a good description of the data at hand, this plot should fluctuate around the zero line. The covariance matrix of $M_{\mathbf{res}}^K(t)$ may be esti-

mated by the optional variation process

$$\hat{\Sigma}_{\mathrm{res}}^K(t) = \int_0^t \int_E K_s (I - Y(s)Y^-(s)) \, D\,(z)\mathrm{diag}(p(ds \times dz))$$
$$D\,(z)(I - Y(s)Y^-(s))^T K_s^T$$

and confidence bands may then be added to the plot.

A uniform Hall-Wellner confidence band over the period from $[0, S^*]$ is given by

$$M_{\mathrm{res}}^K(t) \pm C_\gamma \hat{\Sigma}_{\mathrm{res}}^K(S^*)^{1/2} \left(1 + \frac{\hat{\Sigma}_{\mathrm{res}}^K(t)}{\hat{\Sigma}_{\mathrm{res}}^K(S^*)} \right)$$

where C_γ is the upper γ-quantile of $\sup_{s \in [0, 1/2]} |W^O(s)|$ with $W^O(s)$ the standard Brownian bridge [Hall and Wellner (1980)].

6. Partly conditional mean models. In the previous sections we have dealt with models for the conditional mean for a response given the history at the time of the measurement, that is,

$$E(Z_i^k \,|\, \text{Past history}) = \beta_0(T_i^k) + \beta_1(T_i^k)X_{i1}(T_i^k) + \cdots + \beta_q(T_i^k)X_{iq}(T_i^k)$$
$$= X_{Fi}(T_i^k)\beta(T_i^k)$$

where $\beta(t) = (\beta_0(t), \ldots, \beta_q(t))^T$ and $X_{Fi}(t) = (1, X_{i1}(t), \ldots, X_{iq}(t))$; see (2.1) for definitions. Note that the fully conditional model is assumed to have a fixed dimension. Letting the full design be partioned into a partly conditional model given by covariates $X_{Pi}(t) = (1, X_{i1}(t), \ldots, X_{im}(t))$ where $m < q$ and additional covariates $X_{Ai}(t) = (X_{i(m+1)}(t), \ldots, X_{iq}(t))$, we may write the full conditional design as $X_{Fi}(t) = (X_{Pi}(t), X_{Ai}(t))$. Below, we let subscripts F and P refer to the design for the fully and partly conditional mean models.

Suppose that the regression model is misspecified, either purposely or by accident, such that the data-analyst instead of the fully conditional mean considers the partly conditional model [see Pepe and Couper (1997)],

$$m_i(t) = X_{Pi}(t)\gamma(t)$$

where $\gamma(t) = (\gamma_0(t), \ldots, \gamma_m(t))^T$. Then with $\Gamma_i(t) = \int_0^t \gamma_i(s)\,ds$ and $\Gamma(t) = (\Gamma_0(t), \ldots, \Gamma_m(t))^T$, an estimator of $\Gamma(t)$ is given by

$$\hat{\Gamma}(t) = \int_0^t \int_E \frac{J_P(s)}{\hat{\alpha}(s)} Y_P^-(s)D(z)\,p(ds \times dz).$$

This estimator has the same mean value as (its compensator)

$$\int_0^t \frac{J_P(s)}{\hat{\alpha}(s)} Y_P^-(s)\alpha(s)Y_F(s)\,\beta(s)ds = \int_0^t \frac{J_P(s)}{\hat{\alpha}(s)} \alpha(s)Y_P^-(s)[Y_P(s), Y_A(s)]\,\beta(s)ds$$
$$= \int_0^t \frac{J_P(s)}{\hat{\alpha}(s)} \alpha(s)[I_m, Y_P^-(s)Y_A(s)]\,\beta(s)ds$$

and with the unweighted least squares generalized inverse we see that

$$Y_P^-(s)Y_A(s) = (Y_P(t)^T Y_P(t))^{-1} Y_P(t)^T Y_A(s)$$

$$\xrightarrow{p} S(t) = \text{var}(Y_i(t)X_{Pi}(t))\,\text{cov}(Y_i(t)X_{Pi}(t), Y_i(t)X_{Ai}(t))$$

It therefore can be shown (given some conditions) that the estimator converges to

$$\hat{\Gamma}(t) \xrightarrow{p} \int_0^t [I_m, S(s)]\,\beta(s)ds$$

as $n \to \infty$. The coefficients $[I_m, S(s)]\,\beta(s)$ provide the best linear (L_2) prediction of Z_i^k (or $X_{Fi}(T_i^k)\beta(T_i^k)$) by $X_{Pi}(t)$. In general, that is, without assuming that the fully conditional mean is linear, the partly conditional model will provide estimates of the best linear (L_2) prediction for these covariates.

Therefore, as Aalen (1989) points out for the additive hazard regression model, when omitted covariates are independent of the ones included in the model the effects of covariates included in the model may be estimated from the partly conditional model.

7. Application to growth history analysis. Longitudinal data on growth are commonly encountered in pediatric clinics and used as a tool to monitor the growth of children. In this section we consider growth data on 258 patients with cystic fibrosis made available to us by the Department of Growth and Reproduction at the University Hospital in Copenhagen; see Laursen et al. (1999) for further details on the study. The growth of these patients were followed over a long period of time (1960-1990) with about 9000 visits at the clinic. The number of visits ranged between 2 to 50 with a median at 30. The patients were monitored at rather regular intervals, approximately 4 times per year. The aim of the analysis is to investigate whether the growth patterns have been changing over calendar time, thus, indicating that improvement in medical procedure have had a beneficiary effect on height development. It is expected that the treatment effects will vary with the age of the patients.

We model the current height conditional on the previous measurement of height, the time since the previous measurement and calendar time. The height increase will in general be positive and we therefore model the increase on a logarithmic scale as described below. A negative increase can occur due to measurement error and biological variation such as variation during the day, see Figure 1.

To be specific and with the notation from the previous sections, the model we considered for the two sexes were

$$Z_i^k = \beta_0(T_i^k) + \beta_1(T_i^k)H_i^{k-1} + \beta_2(T_i^k)(T_i^k - T_i^{k-1}) + \beta_3(T_i^k)C_i + \beta_4(T_i^k)C_i^2 + \varepsilon_i^k$$

where H_i^{k-1} is the height at the previous measurement, T_i^{k-1} is the age the previous measurement, C_i is the year of birth for the ith individual and Z_i^k

is the observed height velocity (on log-scale)

$$Z_i^k = \frac{\log(H_i^k) - \log(H_i^{k-1})}{T_i^k - T_i^{k-1}}.$$

The covariate C_i^2 is included as an empirical device to allow some curvature in the model. The model thus explains how a relative change depends on covariates

$$H_i^k = H_i^{k-1}\Big(\beta_0(T_i^k) + \beta_1(T_i^k)H_i^{k-1} + \beta_2(T_i^k)(T_i^k - T_i^{k-1})$$
$$+ \beta_3(T_i^k)C_i + \beta_4(T_i^k)C_i^2 + \varepsilon_i^k\Big)^{(T_i^k - T_i^{k-1})}.$$

The three regression functions $\beta_0(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are considered nuisance parameters.

We start by considering the martingale residual goodness of fit plots described in Section 5. The three continuous variables *previous height*, *time since previous measurement* and *year of birth* were each grouped into 5 equally large groups. We show only the martingale residuals for *year of birth* as this is the variable of primary interest and further is the one revealing some lack of fit for children born before 1970. This is done in Figure 2 where the martingale residual processes with their uniform 95% confidence bands are depicted.

Figure 2 does not reveal any dramatic fluctuations away from 0, and we therefore conclude that the model provides a reasonable fit of data, with some minor problems up to 4 years of age.

Estimates of the cumulative regression functions may be obtained by the estimator (3.4). We chose the bandwidth to 0.5 years. The estimator that does not involve the choice of a smoothing parameter gave essentially the same estimates. We also tried additional bandwidths and obtained similar results. In principle more objective bandwidth selection procedures may be developed based on optimality criterions. One notable difference from the usual smoothing techniques, however, is that we estimate the cumulatives at root-n such that the bias becomes asymptotically negligible. Therefore a more detailed asymptotic study of higher order terms are necessary to learn about for example mean squared error optimality.

We show only the results for boys, the behavior of estimates for girls were quite similar quantitatively, but with some significant differences between the regression coefficients as Figure 5 reveals.

Figure 3 shows the cumulative regression functions and their (pointwise) confidence limits, revealing that most effects are varying with age. The regression coefficient $\beta_2(\cdot)$ may, however, be equal to zero. A maximal deviation test would not be able to rule this out. In Figure 4 we show the estimates of the regression coefficients obtained by smoothing of the cumulative regression functions. Further, a comparison of the growth of boys and girls may be carried out by simply comparing the corresponding cumulative regression functions and this is done in Figure 5 which reveals that the development over time for
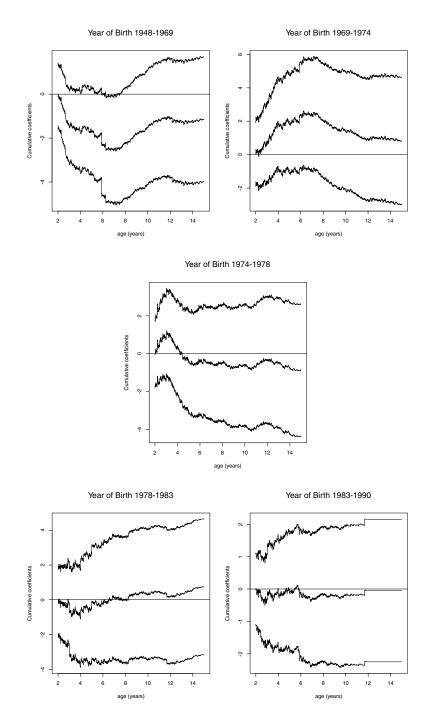
Year of Birth 1948-1969

Year of Birth 1969-1974

Year of Birth 1974-1978

Year of Birth 1978-1983

Year of Birth 1983-1990

FIG. 2. *Goodness of fit plot for* year of birth. *Martingale residual processes for* 5 *equally large groupings based on* year of birth *with* 95% *simultaneous confidence limits.*
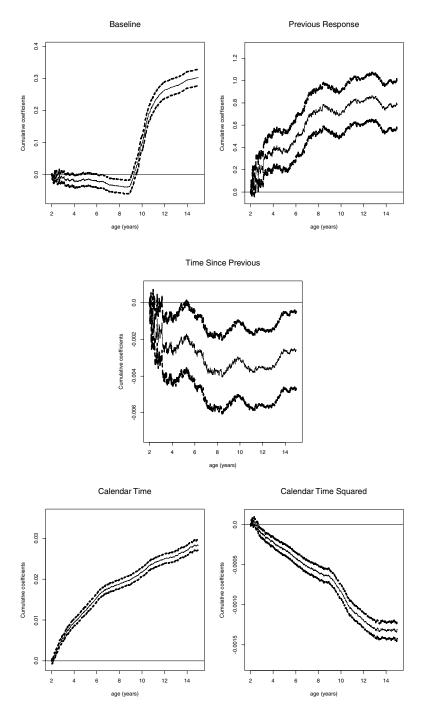
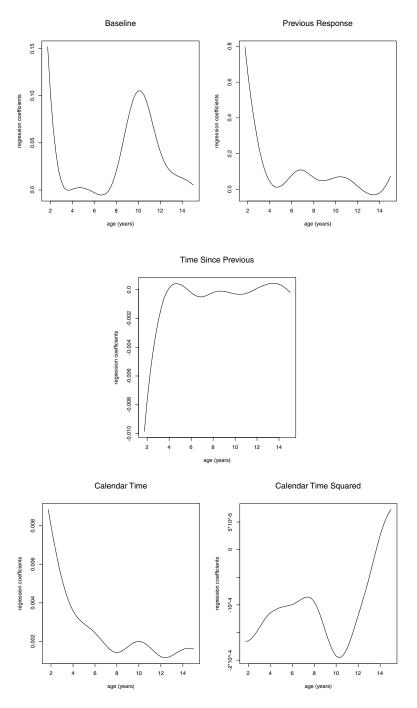FIG. 3.   *Cumulative regression coefficients for boys with 95% pointwise confidence limits.*

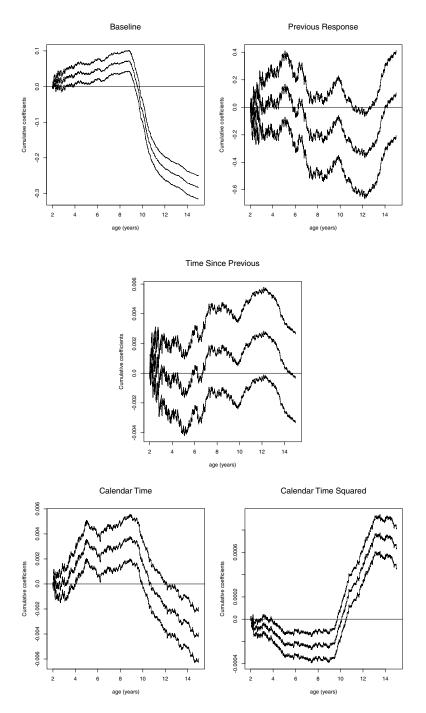FIG. 4. *Time-varying regression coefficients for boys.*

FIG. 5. *Difference between cumulative regression functions for boys and girls with* 95 % *pointwise confidence limits.*
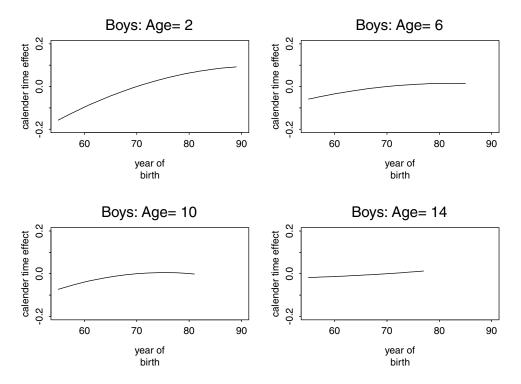
FIG. 6. *Effect of year of birth C on the conditional height velocities for different ages*: $\hat{\beta}_3(a)C + \hat{\beta}_4(a)C^2$, $a = 2, 6, 10, 14$ *years.*

boys and girls differs for the cystic fibrosis patients. The gender differences for $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are however not dramatic.

In Figure 6 we have shown the effects of calendar time on the conditional height velocities for different ages (2, 6, 10 and 14), that is, we have plotted the following functions of year of birth C: $\hat{\beta}_3(a)C + \hat{\beta}_4(a)C^2$, $a = 2, 6, 10, 14$. The height velocities increases with calendar time, and in particular for young children.

In conclusion, the figures indicate that their is some evidence that the treatment has improved the growth of cystic fibrosis patients.

The conclusions that may be drawn are conditional on the cystic fibrosis patients being under observation, and this is a selected group of patients, where selection further may have changed over calendar time. The observed calendar time covers both changing selection effects and changing treatment effects and must therefore be interpreted cautiously. The proposed model may be thought of as a partly conditional model, where the mean is explained by the suggested covariates rather than being the fully conditional model.

8. Discussion. In this paper, we have studied a nonparametric additive regression model for longitudinal data. The influence of each covariate is al-

lowed to vary separately over time. We impose no parametric assumptions on the mean structure letting the influence of the covariates on the longitudinal variable vary freely. The model may be viewed as a varying coefficient model as defined by Hastie and Tibshirani (1993). Hastie and Tibshirani provide an estimation procedure for this more general class of models. However, the theoretical properties of their estimation procedure have not been well studied. Since our model may be considered as a marked point process model, we may draw on general results for marked point processes such as the central limit theorem for martingales. We showed that the limiting distribution of the estimated cumulative regression functions is a Gaussian martingale process with a variance function that may be estimated consistently which in turn makes inference about the regression functions possible. Simulation studies reported in an unpublished paper indicate that the small sample properties of the suggested estimators are satisfactory.

The suggested model is a very flexible tool in that the effect of the covariates on the longitudinal variable is specified nonparametrically. However, if there are several covariates to be considered, data may be to sparse to obtain reliable estimates. Therefore it is relevant to study whether some of the regression functions may be described by parametric expressions (e.g., the regression function is constant over time); see McKeague and Sasieni (1994), Lin and Ying (1994) and Martinussen and Sørensen (1998) in the context of Aalen's additive hazard model for event time data. Martinussen and Scheike (1999) considered estimation and test-procedures for such a semi-parametric regression model for longitudinal data.

Due to the additive structure of the proposed model it should be possible to handle the situation where there may be missing covariate values. This is currently being investigated and will be communicated in a separate report.

A different area where further research is needed is the modeling of the measurement times. It is seen from the suggested estimator that the design vectors are weighted according to the intensity of observations of the corresponding subjects. Such that, for example, an over-representation of extreme measurements does not bias the estimator. In this work we considered Aalen's multiplicative model where the predictable part of the intensity takes into account how the intensity depends on past observations. In applications it might be unclear how the intensity depends on the development of the process and in this case one may look to extend the model to estimate how the measurement process depends on past observations. A two step procedure, where the intensity is modeled in an initial step and then estimated and plugged into the estimation equations used for the regression functions for the conditional mean, may lead to satisfactory results. One may think of the modeling of the measurement time process as the price we are paying for the simple estimator of the cumulative regression functions. The usual approaches are based on conditioning with the measurement times and then estimating the regression functions directly by smoothing techniques. The smoothing based approach completely avoids the problem of specifying a correct model for the intensities. If, however, the intensity does interact with the measurements, the conditional

distribution given the measurement times should reflect this fact, just like the dynamic regression model should reflect it. For non-fixed designs the analysis should therefore always contain a detailed study of the the measurements times as well as the responses.

APPENDIX

We will apply marked point process theory to derive the large sample results presented in the paper.

The following result for marked point processes is used for handling the marked point process integrals in the model and is used throughout the rest of the Appendix. This basic result was also applied in Scheike (1994).

PROPOSITION 1. *Let p_i, $i = 1, \ldots, n$, be marked point processes as defined in Section 2, and let $H_{i1}(t, z_i)$, $H_{i2}(t, z_i)$ be $\mathscr{F}_t$-predictable processes such that*

$$\int_0^t \int_E H_{ij}^2(s, z_i) \lambda_s^i \Phi_s^i(dz_i) ds < \infty \qquad a.s,$$

for $j = 1, 2$. Further, define $q_i(ds \times dz_i) = p_i(ds \times dz_i) - \lambda_t^i \Phi_t^i(dz_i) ds$ and

$$M_i(H_{ij})(t) = \int_0^t \int_E H_{ij}(s, z_i) q_i(ds \times dz_i).$$

Then $M_i(H_{ij})(t)$, $i = 1, \ldots, n$, are orthogonal local square-integrable martingales with respect to $\mathscr{F}_t$. The predictable variation processes are given by

$$\langle M_i(H_{i1}), M_i(H_{i2}) \rangle(t) = \int_0^t \int_E H_{i1}(s, z_i) H_{i2}(s, z_i) \lambda_s^i \Phi_s^i(dz_i) ds$$

$i = 1, \ldots, n.$

PROOF OF THEOREM 1. Let $B^*(t) = \int_0^t J(s) \beta(s) ds$. Using (A), (C) and (E) it is seen that $\sqrt{n}(\hat{B} - B)$ and $\sqrt{n}(\hat{B} - B^*)$ have the same limiting distribution. Now,

$$\sqrt{n}(\hat{B} - B^*)(t)$$

(A.1)
$$= \sqrt{n} \int_0^t \frac{J(s)}{\hat{\alpha}(s)} Y^-(s) dM(z)(s) - \sqrt{n} \int_0^t \frac{J(s)}{\hat{\alpha}(s)} \beta(s) d(\hat{A}(s) - A(s))$$

$$+ \sqrt{n} \int_0^t \frac{J(s)}{\hat{\alpha}(s)} \beta(s) d(\hat{A}(s) - \tilde{A}(s)),$$

where $\hat{A}(t) = \int_0^t \frac{1}{Y.(s)} dN.(s)$ and $\tilde{A}(t) = \int_0^t \hat{\alpha}(s) ds$. Interchanging integration order in $\int_0^t \hat{\alpha}(s) ds$ and using a Taylor expansion, it can be seen that

$$\sup_{t \in [0,S]} (\hat{A}(t) - \tilde{A}(t)) = O_p(b_n^2).$$

Hence, using the continuous mapping theorem, the last term on the right-hand side of (A.1) converges uniformly to zero in probability. Let $X(t)$ denote the sum of the two first terms on the right-hand side of (A.1). We may write $X(t)$ as

$$X(t) = \sqrt{n} \int_0^t \int_E \frac{J(s)}{\hat{\alpha}(s)} Y^-(s)(D(z) - \frac{1}{Y.(s)} Y(s)\beta(s)a) q(ds \times dz)$$

where $q(ds \times dz) = (q_1(ds \times dz_1), \ldots, q_n(ds \times dz_n))^T$ and a is the $1 \times n$-vector $(1, \ldots, 1)$. We may replace $\hat{\alpha}(t)$ by $\alpha(t)$ in $X(t)$ and still obtain the same limiting distribution by use of the continuous mapping theorem. The jth component of $X(t)$ may then be written as

$$X_j(t) = \sum_{i=1}^n \int_0^t \int_E \frac{1}{\sqrt{n}} \frac{J(s)}{\alpha(s)} (V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)}) q_i(ds \times dz_i)$$

$$= \sum_{i=1}^n M_i(H_{ij})(t)$$

where

$$H_{ij}(t, z_i) = \frac{1}{\sqrt{n}} \frac{J(s)}{\alpha(s)} (V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)})$$

with $V_{ji}(t) = \sum_{l=0}^p (\frac{1}{n} R^{(2)}(t))_{jl}^{-1} Y_{il}(t)$. To identify the asymptotic variance, we have to compute the predictable variation process of $X(t)$. Using Proposition 1, it is found to be (suppressing the dependency of time on the integrands)

$$\langle X_j, X_k \rangle(t) = \sum_{l,m=0}^p \int_0^t \frac{J}{\alpha} \left(\frac{1}{n} R^{(2)}\right)_{jl}^{-1}$$

$$\times \left(\frac{1}{n} R^{(2)}\right)_{km}^{-1} \left(\frac{1}{n} \sum_{i=1}^n Y_{il} Y_{im} \left(\sum_{f,g=0}^p Y_{if} Y_{ig} \beta_f \beta_g + \sigma^2\right)\right) ds$$

$$- \int_0^t \frac{J\beta_j \beta_k}{\alpha \frac{1}{n} Y.} ds.$$

From the above expressions and using (A), (C) and (E) it is seen that $\langle X_j, X_k \rangle(t)$ converges in probability to the expression for $\text{cov}(U_j(t), U_k(t))$

stated in Theorem 1. We will now turn to the Lindeberg condition in the martingale central limit theorem. The process containing all the jumps of $X_j(t)$ larger in absolute value than ε is given by

$$X_{j\varepsilon}(t) = \sum_{i=1}^{n} \int_0^t \int_E \frac{1}{\sqrt{n}} \frac{1}{\alpha(s)} \left(V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)} \right)$$

$$\times I \left(\frac{1}{\sqrt{n}} \frac{1}{\alpha(s)} \left| V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)} \right| > \varepsilon \right) q_i(ds \times dz_i)$$

The predictable variation process for $X_{j\varepsilon}(t)$ is

$$\langle X_{j\varepsilon} \rangle(t) = \sum_{i=1}^{n} \int_0^t \int_E \frac{1}{n} \frac{Y_i(s)}{\alpha(s)} \left(V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)} \right)^2$$

$$\times I \left(\frac{1}{\sqrt{n}} \frac{1}{\alpha(s)} \left| V_{ji}(s)z_i - \frac{\beta_j(s)}{\frac{1}{n}Y.(s)} \right| > \varepsilon \right) \Phi_s^i(dz_i) ds$$

and we have to show that $\langle X_{j\varepsilon} \rangle(t) \xrightarrow{p} 0$ as $n \to \infty$. By applying the elementary inequality

$$(a-b)^2 I(|a-b| > \varepsilon) \leq 4a^2 I(|a| > \varepsilon/2) + 4b^2 I(|b| > \varepsilon/2)$$

twice, it may be seen that

$$\langle X_{j\varepsilon} \rangle(t) \leq (c_1 G_1^{(n)})^2 16 \int_0^t \frac{1}{\alpha(s)} \left(\frac{1}{n} \sum_{i=1}^{n} V_{ji}^2(s) \right) \frac{16}{\varepsilon^2} \eta(s) ds$$

$$+ 16 \int_0^t \frac{1}{\alpha(s)} \left(\frac{1}{n} \sum_{i=1}^{n} V_{ji}^2(s) m_i^2(s) \right) I(c_2 G_2^{(n)} > \varepsilon/4) ds$$

$$+ 4 \int_0^t \frac{\beta_j^2(s)}{\alpha(s)\frac{1}{n}Y.(s)} I \left(\frac{1}{\sqrt{n}} \frac{1}{\alpha(s)} \left| \frac{\beta_j(s)}{\frac{1}{n}Y.(s)} \right| > \varepsilon/2 \right) ds$$

where c_1 and c_2 are (finite) constants. Using the assumptions (A), (B), (C), (D) and (E) it follows that $\langle X_{j\varepsilon} \rangle(t) \xrightarrow{p} 0$ as $n \to \infty$, and this completes the proof. $\quad\square$

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 534–545.

AALEN, O. O. (1980). A model for non-parametric regression analysis of counting processes. *Mathematical Statistics and Probability. Lecture Notes in Statist.* **2** 1–25. Springer, New York.

AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in Medicine* **8** 907–925.

AALEN, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statistics in Medicine* **12** 1569–1588.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

BRÉMAUD, P. (1981). *Point Processes and Queues, Martingale Dynamics*. Springer, New York.

DIGGLE, P., LIANG, K. and ZEGER, S. (1994). *Analysis of Longitudinal Data*. Cambridge Univ. Press.

FAN, J. and ZHANG, W. (1999). Statistical estimation in varying-coefficient models. *Ann. Statistics* **27** 1491–1518.

HALL, W. and WELLNER, J. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67** 133–143.

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

HOOVER, P., RICE, J., WU, C. and YANG, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822.

HUFFER, F. W. and MCKEAGUE, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Amer. Statist. Assoc.* **86** 114–129.

LAURSEN, E., KOCH, C., PETERSEN, J. and MULLER, J. (1999). Secular changes in anthropometric data in cystic fibrosis patients. *Acta Pædiatrica* **88** 171–179.

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. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81** 61–71.

MARTINUSSEN, T. and SCHEIKE, T. H. (1998). A non-parametric dynamic additive regression model for longitudinal data. Research report, Dept. Biostatistics, Univ. Copenhagen.

MARTINUSSEN, T. and SCHEIKE, T. H. (1999). A semi-parametric additive regression model for longitudinal data. *Biometrika* **86** 691–702.

MARTINUSSEN, T. and SØRENSEN, T. I. A. (1998). Age-dependent *U*-shaped risk functions and Aalen's additive risk model. *Biometrics* **54** 989–1001.

MCKEAGUE, I. W. (1988). Asymptotic theory for weighted least squares estimators in Aalen's additive risk model. *Contemp. Math.* **80** 139–152.

MCKEAGUE, I. W. and SASIENI, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81** 501–514.

MCKEAGUE, I. W. and ZHANG, M. J. (1994). Identification of nonlinear time series from first order cumulative characteristics. *Ann. Statist.* **22** 495–514.

MURPHY, S. A. and LI, B. (1995). Projected partial likelihood and its application to longitudinal data. *Biometrika* **82** 399–406.

MURPHY, S. A. (1995). A central limit theorem for local martingales with applications to the analysis of longitudinal data. *Scandinavian J. Statist.* **22** 279–294.

PEPE, M. and COUPER, D. (1997). Modeling partly conditional means with longitudinal data. *J. Amer. Statist. Assoc.* **92** 991–998.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

SCHEIKE, T. H. (1994). Parametric regression for longitudinal data with counting process measurement times. *Scandinavian J. Statist.* **21** 245–263.

SCHEIKE, T. H. and ZHANG, M. (1998). Cumulative regression function tests for longitudinal data. *Ann. Statist.* **26** 1328–1355.

SCHEIKE, T. H. (1997). A general framework for longitudinal data through conditional distributions. *Biometrical J.* **39** 57–67.

SCHEIKE, T. H. (2000). Comparison of non-parametric regression functions through their cumulatives. *Statist. Probab. Lett.* **46** 21–32.

DEPARTMENT OF MATHEMATICS AND PHYSICS
THE ROYAL VETERINARY
   AND AGRICULTURAL UNIVERSITY
THORVALDSENSVEJ 40
DK-1871 FREDERIKSBERG C
DENMARK
E-MAIL: torbenm@dina.kvl.dk

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF COPENHAGEN
BLEGDAMSVEJ 3
DK-2200 COPENHAGEN N
DENMARK
E-MAIL: ts@biostat.ku.dk