The Annals of Statistics
1998, Vol. 26, No. 4, 1328–1355

CUMULATIVE REGRESSION FUNCTION TESTS FOR REGRESSION MODELS FOR LONGITUDINAL DATA

BY THOMAS H. SCHEIKE AND MEI-JIE ZHANG[1]

University of Copenhagen and Medical College of Wisconsin

The longitudinal regression model $Y_{i,j} = m(V^i_{\tau_{i,j}}) + \varepsilon_{i,j}$ where $Y_{i,j}$, is the jth measurement of the ith subject at random time $\tau_{i,j}$, m is the regression function, $V^i_{\tau_{i,j}}$ is a predictable covariate process observed at time $\tau_{i,j}$ and $\varepsilon_{i,j}$ is noise, often provides an adequate framework for modeling and comparing groups of data. The proposed longitudinal regression model is based on marked point process theory, and allows a quite general dependency structure among the observations.

In this paper we find the asymptotic distribution of the cumulative regression function (CRF), and present a nonparametric test to compare the regression functions for two groups of longitudinal data. The proposed test, denoted the CRF test, is based on the cumulative regression function (CRF) and is the regression equivalent of the log-rank test in survival analysis. We show as a special case that the CRF test is valid for groups of independent identically distributed regression data. Apart from the CRF test, we also consider a maximal deviation statistic that may be used when the CRF test is inefficient.

1. Introduction. The aim of this paper is to present a nonparametric test to compare groups of longitudinal regression data obtained over time. The focus is on some measurement which we model conditionally on the observed history. We consider a nonparametric transition model [see Diggle, Liang and Zeger (1994)] that models the conditional mean of the current response given past outcomes, which in our model amounts to previously obtained measurements, the times for these measurements and outside random variation in the sense of Kalbfleisch and Prentice [(1980), Chapter 5]. The basis of transition models is the conditional mean structure, and inference is carried out by fitting the observations to their conditional means. In contrast to other approaches, including the generalized estimation equations [Liang and Zeger (1986)], for longitudinal data the transition model approach accommodates dynamic modeling of the processes under observation.

One challenging problem in longitudinal data analysis is to test whether two groups of independent longitudinal regression data have identical conditional mean functions. We introduce an approach to this problem based on the difference of estimators of the cumulative version of the conditional means

Received September 1996; revised September 1997.

[1]Supported in part by National Cancer Institute Grants 1 R01 CA54706-03 and PO1-CA-40053 and by American Cancer Society Institutional Research Grant 170H.

AMS 1991 *subject classifications*. Primary 62M10; secondary 62G07, 62G10, 62G20

Key words and phrases. Conditional regression models, cumulative regression function, longitudinal data, marked point process, maximal deviation statistic, nonparametric test, one-sample test, two-sample test.

between two groups. Later we generalize this nonparametric two-sample test to the k-sample test. To focus ideas, consider two groups of n_1 and n_2 subjects, each individual giving rise to N_i^k observations, $\{(Y_{k,i,j}, \tau_{k,i,j}) \mid j = 1, \ldots, N_i^k; i = 1, \ldots, n_k; k = 1, 2\}$. Here $Y_{k,i,j}$ is the jth measurement of the ith subject from group k at random time $\tau_{k,i,j} \in [0, t]$, where $[0, t]$ is the time period of observation. We assume that the conditional regression model is valid for each individual from the two groups such that

$$(1) \qquad Y_{k,i,j} = m_k(V^{k,i}_{\tau_{k,i,j}}) + \varepsilon_{k,i,j} \quad \text{for } j = 1, \ldots, N_i^k, \; i = 1, \ldots, n_k, \; k = 1, 2,$$

where m_k, $k = 1, 2$ are the regression functions of the two groups, $V_s^{k,i}$ is the d-dimensional covariate process of subject i from group k at time s and $\varepsilon_{k,i,j}$ is noise. The covariates, $V_s^{k,i}$, are allowed to depend on previous observations and outside random variation, and are thus predictable in the history of the observations. We return to a complete model specification in Section 2. Note that the covariates may be time ($V_s^{k,i} = s$). Estimation of the regression function for this longitudinal regression model has been discussed in a parametric setup in Scheike (1994) and nonparametrically in Scheike (1996). In this paper we construct a statistical procedure to compare the regression functions of two groups of data. This question arises often in statistical applications, for example, when the growth of two groups of patients is compared. In Section 4 we apply the procedure to compare the growth of patients with two different diagnoses of dwarfism (hypochondroplasia and achondroplasia) on the basis of longitudinal data. The proposed methodology may also be applied to compare reference charts based on cross-sectional data.

The proposed nonparametric test is the regression equivalent of the log-rank test of survival analysis; see, for example, Andersen, Borgan, Gill and Keiding (1993). Our approach is similar to that of McKeague and Zhang (1994) (MZ), who constructed a nonparametric test for comparing regression functions for stationary time series. The test proposed by MZ is based on comparing estimates of the integrals $\int_a^z m_1(x)\,dx$ and $\int_a^z m_2(x)\,dx$, where a and z are appropriately chosen points in d-dimensional regressor space. We consider the same test statistics and make inference based on the asymptotic distribution of $\int_a^{(\cdot)} \hat{m}_k(x)\,dx - \int_a^{(\cdot)} m_k(x)\,dx$, $k = 1, 2$, where $\hat{m}_k(x)$ is a nonparametric estimate of the regression function. The test can be applied to groups of independent identically distributed regression data.

The structure of the paper is as follows. Section 2 presents the model construction in a marked point process framework that allows a quite general dependency structure among the observations and also presents some Martingale results we later use to establish the asymptotic properties of our estimators. We further show that independent identically distributed regression data are a special case of our model. Section 3 presents the two-sample test for regression data and provides the asymptotic distribution of the proposed test statistic. Section 4 contains an application to data on the growth of two types of dwarfism and a simulation study to evaluate the power of the proposed test. In Section 5, some closing remarks are given. Finally, Section 6 contains the proofs.

2. A longitudinal regression model. Consider the regression model

$$Y_{k,i,j} = m_k(V^{k,i}_{\tau_{k,i,j}}) + \varepsilon_{k,i,j} \quad \text{for } j = 1, \dots, N^k_i,\ i = 1, \dots, n_k,\ k = 1, 2, \tag{2}$$

where $Y_{k,i,j}$ is the jth real-valued measurement of the ith subject from group k at a random time $\tau_{k,i,j}$ in the period of observation $[0, t]$, m_k, $k = 1, 2$, are the regression functions of the two groups, $V^{k,i}_s$ is a d-dimensional covariate process of subject i from groups k at time s and $\varepsilon_{k,i,j}$ is noise. Some additional structure is required to specify the model, so we formulate the model in terms of marked point processes; see Brèmaud (1981).

Let $\mathscr{B}$ denote the Borel σ-field on $\Re$. For $A \in \mathscr{B}$, define the counting process

$$N^{k,i}_s(A) = \sum_j I(Y_{k,i,j} \in A) I(\tau_{k,i,j} \le s) \quad \text{as defined,}$$

that counts the number of observations in the set A for the ith individual of group k, and the associated *marked point process* $P^{k,i}(ds \times dz)$:

$$P^{k,i}([0, s] \times A) = N^{k,i}_s(A) \quad \text{as defined, for } s \ge 0,\ A \in \mathscr{B}.$$

Define further the history of the subjects, that is, the history of the marked point processes, as

$$\mathscr{F}_u = \sigma\big(N^{k,i}_s(A) \colon s \le u,\ A \in \mathscr{B},\ i = 1, \dots, n_k,\ k = 1, 2\big) \vee \mathscr{A} \quad \text{as defined.}$$

The σ-algebra $\mathscr{A}$ is independent of $\sigma(N^{k,i}_s(A) \colon s \le u,\ A \in \mathscr{B},\ i = 1, \dots, n_k,$ $k = 1, 2)$ and represents knowledge prior to time 0. We further need to define the σ-field $\mathscr{F}_{\tau_{k,i,j}-} = \sigma((Y_{k,i,m}, \tau_{k,i,m}) \colon \tau_{k,i,m} < \tau_{k,i,j}; \tau_{k,i,j}) \vee \mathscr{A}$ that contains the information just prior to observation of a jump size.

Define further

$$N^{k,i}_s = N^{k,i}_s(\Re) \quad \text{as defined,}$$

the counting process associated with the $\tau_{k,i,j}$'s, where $\Re = (-\infty, \infty)$. It is assumed that no two of the counting processes $N^{k,i}_s$ jump at the same time, and this is needed in Proposition 1.

We assume that $V^{k,i}_s$ is predictable with respect to the history $\mathscr{F}_s$, that $N^{k,i}_s$ has a random intensity $\lambda^{k,i}_s \ge 0$ and that all processes are cadlag. The existence of an intensity implies that measurement times vary continuously and do not clump at, for example, weekly intervals. Clumping of measurements due to measurements at regular intervals makes the compensator discontinuous. Murphy (1995) dealt with asymptotic analysis of Martingales where measurement times can vary continuously as well as clump at regular intervals. We return to this issue in the remark after Proposition 2. The intensity has the following interpretation: $\lambda^{k,i}_s\, ds$ is the probability of a jump in the time interval $(s; s + ds]$ given the history of $\mathscr{F}_s$. One particular form for the intensity, $\lambda^{k,i}_s$, of interest for applications is Aalen's multiplicative intensity model

$$\lambda^{k,i}_s = \alpha_k(s) Z^{k,i}(s) \quad \text{as defined,}$$

where $\alpha_k(s)$ is a deterministic function and $Z^{k,i}(s)$ is $\mathscr{F}_s$ predictable; see Aalen (1975, 1978).

Finally, it is assumed that the noise terms from (2) have conditional mean and variance given by

$$E(\varepsilon_{k,i,j} \mid \mathscr{F}_{\tau_{k,i,j}-}) = 0,$$
$$E(\varepsilon_{k,i,j}^2 \mid \mathscr{F}_{\tau_{k,i,j}-}) = \sigma_k^2(V_{\tau_{k,i,j}}^{k,i}),$$

so that

$$\begin{aligned} E(Y_{k,i,j} \mid \mathscr{F}_{\tau_{k,i,j}-}) &= m_k(V_{\tau_{k,i,j}}^{k,i}), \\ E(Y_{k,i,j}^2 \mid \mathscr{F}_{\tau_{k,i,j}-}) &= m_k^2(V_{\tau_{k,i,j}}^{k,i}) + \sigma_k^2(V_{\tau_{k,i,j}}^{k,i}), \end{aligned} \tag{3}$$

for $k = 1, 2$, $i = 1, \dots, n_k$, $j = 1, \dots, N_t^{k,i}$, where $\sigma_k^2(\cdot)$ is deterministic, continuous and bounded.

The law of $\varepsilon_{k,i,j}$ conditional on the past and the time of the jump, $\tau_{k,i,j}$, is denoted

$$F_{\tau_{k,i,j}}^{k,i}(z) = P(\varepsilon_{k,i,j} \le z \mid \mathscr{F}_{\tau_{k,i,j}-}) \quad \text{as defined.}$$

Therefore, the conditional distribution of $Y_{k,i,j}$ is

$$P(Y_{k,i,j} \le z \mid \mathscr{F}_{\tau_{k,i,j}-}) = F_{\tau_{k,i,j}}^{k,i}(z - m_k(V_{\tau_{k,i,j}}^{k,i})).$$

The actual construction of the processes can be carried through by specifying the particular form of $\lambda_s^{k,i}$ and $V_s^{k,i}$, and the conditional distribution of $Y_{k,i,j}$ given the history through the functions $F_s^{k,i}(z)$ and $m_k(\cdot)$; see Jacobsen [(1982), Chapter 2].

In this work we aim to estimate the regression function nonparametrically. We consider the situation where possibly one component of the covariates may be equal to time and the rest vary continuously. We assume that the conditional distribution of the d-dimensional covariates $V_s^{k,i}$ given $\lambda_s^{k,i}$ has a bounded density $f_s^{k,i}(v_1, \dots, v_d)$ with respect to Lebesgue measure and denote the distribution as $f_s^{k,i}(v_1, \dots, v_d)\, dl_d$, where l_d is the d-dimensional Lebesgue measure or is a product of the Dirac measure at s, ε_s and a measure given by a $d-1$-dimensional bounded density $f_s^{k,i}(v_1, \dots, v_{d-1})$ with respect to Lebesgue measure, that is, $\varepsilon_s \otimes f_s^{k,i}(v_1, \dots, v_{d-1})\, dl_{d-1}$. If we do not wish to specify which one of the two cases we consider, we write $G_s^{k,i}(dv)$.

The following Martingale result about marked point processes is the key to handling the dependencies and is used throughout the rest of the paper. Define a 0-Martingale to be a Martingale with mean 0.

PROPOSITION 1. *Let $P^{k,i}(ds \times dz)$ for $i = 1, \dots, n_k$, $k = 1, 2$, be marked point processes defined as above and let $H_{k,i}(s, z)$ be $\mathscr{F}_s$ predictable processes such that*

$$E\left(\int_0^t \int_{\Re} H_{k,i}^2(s, z) \lambda_s^{k,i}\, dF_s^{k,i}(z - m_k(V_s^{k,i}))\, ds\right) < \infty. \tag{4}$$

Then

$$M^{k,i}(H_{k,i})_t = \int_0^t \int_{\Re} H_{k,i}(s,z) P^{k,i}(ds \times dz) - \int_0^t \int_{\Re} H_{k,i}(s,z) \lambda_s^{k,i}\, dF_s^{k,i}(z - m_k(V_s^{k,i}))\, ds \quad \textit{as defined}$$

are orthogonal square-integrable 0-*Martingales with respect to* $\mathscr{F}_t$, *with variance processes*

$$\langle M^{k,i}(H_{k,i}), M^{k,i}(H_{k,i})\rangle_t = \int_0^t \int_{\Re} H^2_{k,i}(s,z) \lambda_s^{k,i}\, dF_s^{k,i}(z - m_k(V_s^{k,i}))\, ds$$

for $i = 1, \dots, n_k$, $k = 1, 2$.

The proposition follows from Boel, Varaiya and Wong (1975a, b) or Brèmaud (1981).

To present a Nadaraya–Watson (ND) type estimator, see Nadaraya (1964) and Watson (1964), of the regression functions, we need some definitions. Let $K(\cdot)$ be a kernel function with support on $[-1;1]$, $\int K(u)\,du = 1$, and let $b = (b_1, \dots, b_d)$ be a d-dimensional bandwidth, $|b| = b_1 \cdot \cdots \cdot b_d$, $b \in (0;\infty)^d$. Define further $c_K = \int K^2(u)\,du$ as defined, $d_K = \int u^2 K(u)\,du$ as defined and $e_K = \int u K(u)\,du$ as defined. We assume that $c_K = O(1)$, $d_K = O(1)$ and e_K is 0 to obtain an asymptotically unbiased result for our estimator. Finally, we assume that the kernel function satisfies a Lipschitz condition, that is, $|K(x) - K(y)| \le C|x - y|$. The assumptions on the kernel are denoted **K**. We abuse notation by letting K denote a d-dimensional as well as a one-dimensional kernel through the product kernel, that is, $K(y, b) = K(y_1/b_1, \dots, y_d/b_d) = \prod_{i=1}^d K(y_i/b_i)$ as defined.

The ND estimator, $\hat{m}_k(y)$, of $m_k(y)$ is defined by

$$\hat{m}_k(y) = \frac{\hat{r}_k(y)}{\hat{\alpha}_k(y)} \quad \text{as defined,} \tag{5}$$

where

$$\hat{r}_k(y) = \frac{1}{n_k} \sum_{i=1}^{n_k} \sum_{j=1}^{N_t^{k,i}} Y_{k,i,j} \frac{1}{|b|} K(y - V^{k,i}_{\tau_{k,i,j}}, b) \quad \text{as defined}$$

and

$$\hat{\alpha}_k(y) = \frac{1}{n_k} \sum_{i=1}^{n_k} \sum_{j=1}^{N_t^{k,i}} \frac{1}{|b|} K(y - V^{k,i}_{\tau_{k,i,j}}, b) \quad \text{as defined.} \tag{6}$$

The behavior of a similar nonparametric estimator was studied in Scheike (1993). In this paper, however, the focus is on the test statistics based on the cumulative version of $\hat{m}_k$, which is established in the next section, and a detailed study of the ND estimator is therefore omitted.

The term $\hat{\alpha}_k(y)$ is an estimator of the density of the covariates. Fusaro, Nielsen and Scheike (1993) and Nielsen and Linton (1995) used this estimator as the occurrence estimator in an occurrence/exposure estimator of a marker dependent hazard function.

Define

$$\alpha_k^\star(y) = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \frac{1}{|b|}K(y - V_s^{k,i}, b)\lambda_s^{k,i}\,ds, \tag{7}$$

$$D_k(y) = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re}(z - m_k(y))\frac{1}{|b|}K(y - V_{\tau_{k,i,j}}^{k,i}, b)P^{k,i}(ds\times dz) \tag{8}$$

and

$$\begin{aligned} D_k^\star(y) &= \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re}(z - m_k(y))\frac{1}{|b|}K(y - V_s^{k,i}, b)\lambda_s^{k,i}\,dF_s^{k,i} \\ &\qquad\qquad \times (z - m_k(V_s^{k,i})\,ds \\ &= \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t (m_k(V_s^{k,i}) - m_k(y))\frac{1}{|b|}K(y - V_s^{k,i}, b)\lambda_s^{k,i}\,ds \end{aligned} \tag{9}$$

and note that $\hat{m}_k(y) - m_k(y) = D_k(y)/\hat{\alpha}_k(y)$.

The following result regarding the asymptotic distribution of $\hat{\alpha}_k(y)$ and $\hat{m}_k(y)$, can be established.

PROPOSITION 2. *Assume the following statements*:

(i) $m_k(\cdot)$ *is twice continuously differentiable*, $\sigma_k^2(\cdot)$ *is bounded and the kernel satisfies condition* **K**.

(ii) *Conditional on the intensity*, $\lambda_s^{k,i}$, *the covariates*, $V_s^{k,i}$, *have distribution* $f_s^{k,i}(v)\,dl_d$ *or* $\varepsilon_s \otimes f_s^{k,i}(v_1,\dots,v_{d-1})\,dl_{d-1}$, *where the densities are bounded.*

(iii) $E((1/n_k)\sum_{i=1}^{n_k}\int_0^t \lambda_s^{k,i}\,ds) = O(1)$.

(iv) $b \to 0$, $n_k \to \infty$, *such that* $n_k|b| \to \infty$.

(v) *There exists* $\alpha_k(y) > \delta > 0$ *and a compact set in* d-*dimensional space*, A, *such that* $\sup_{y\in A} |\,\alpha_k^\star(y) - \alpha_k(y)| \to_p 0$.

It then follows that

$$\sup_{y\in A}\big|\hat{\alpha}_k(y) - \alpha_k(y)\big| \to_p 0, \qquad \sup_{y\in A}\big|\hat{m}_k(y) - m_k(y)\big| \to_p 0.$$

REMARK. The proof of Proposition 2 is based on the existence of an intensity, thus making the compensator continuous. In applications, however, the measurement times may be regular (e.g., every week) such that the compensator will have discontinuities. Murphy (1995) discussed this situation and establishes Martingale asymptotics which may be used to extend Proposition 2 to more regular measurement times.

The variance functions, $\sigma_k^2(\cdot)$, can be estimated by the squared-residual kernel estimator

$$\hat{V}_k(y) = \frac{V_k(y)}{\hat{\alpha}_k(y)} - (\hat{m}_k(y))^2, \tag{10}$$

where

$$V_k(y) = \frac{1}{n_k}\sum_{i=1}^{n_k}\sum_{j=1}^{N_t^{k,i}}(Y_{k,i,j})^2\frac{1}{|b|}K(y - V^{k,i}_{\tau_{k,i,j}}, b) \quad \text{as defined.} \tag{11}$$

It follows from the next proposition that this estimator is consistent.

PROPOSITION 3. *Assume that the assumptions of Proposition* 2 *are satisfied and, further, that the conditional fourth moment of* $Y_{k,i,j}$ *is bounded. Then*

$$\sup_{y\in A}\left|\hat{V}_k(y) - \sigma_k^2(y)\right| \to_p 0.$$

In the case of the subjects being identically distributed, it follows that condition (iii) reduces to $E(\int_0^t \lambda_s^{k,i}\,ds) < C$. The existence of $\alpha_k(y)$ in condition (v) makes it equal to

$$\lim_{n_k\to\infty} E\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re^d}\frac{1}{|b|}K(y - v, b)G_s^{k,i}(dv)\lambda_s^{k,i}\,ds\right).$$

When subjects are identically distributed this reduces to $E(\int_0^t f_s^{k,i}(y)\lambda^{k,i}(s)\,ds)$ when $G_s^{k,i}(dv) = f_s^{k,i}(v)\,dl_d$ and to $E(f_s^{k,i}(y)\lambda^{k,i}(s))$ when $G_s^{k,i}(dv) = \varepsilon_s \otimes f_s^{k,i}(v_1,\dots,v_{d-1})\,dl_{d-1}$. When subjects are further independent, the existence of the $\alpha_k(y)$ is implied by the existence of the above mean value.

EXAMPLE. This example shows that the usual independent identically distributed regression data are a special case of the proposed model.

Let (Y_i, X_i), $i \in 1,\dots,n$, be independent and identically distributed and assume that the X_i's are observed in the time interval $[0, t]$ unless they are censored. The functional of interest is $m(s) = E(Y_i \mid X_i = s)$.

Denote the intensity function of X_i by $\alpha(\cdot)$ and the distribution function by $F(\cdot)$. Assume, furthermore, that there are given i.i.d. censoring times $C_1,\dots,C_n$, which are independent of (Y_i, X_i). Denote the distribution function of the C_i by $H(\cdot)$. Consider the counting processes

$$N_t^i = I(X_i \le t, X_i \le C_i) \quad \text{as defined}$$

with the intensity process $\alpha(s)Z^i(s)$, where

$$Z^i(s) = I(X_i \ge s, C_i \ge s),$$

so $Z^i(s)$ is 1 if subject i is still under risk at time s and 0 otherwise. Now,

$$\begin{aligned}E(\lambda^i(s)) &= \alpha(s)E(Z^i(s))\\ &= \alpha(s)((1 - F(s))(1 - H(s-))).\end{aligned}$$

Note that this expectation is strictly positive on $[0, t]$ if $F(t) < 1$ and $H(t-) < 1$. In this situation $\alpha^\star(s)$ is equal to

$$\alpha^\star(s) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t\frac{1}{|b|}K(y - V_s^i, b)\lambda_s^i\,ds,$$

where $V_s^i = s$. Let $\nu(s) = \alpha(s)(1 - F(s))(1 - H(s-))$ be twice continuously differentiable. Then it can be shown that

$$\sup_{[0,\,t]} |\alpha^\star(s) - \nu(s)| = O_p(b^2).$$

The assumptions of Proposition 2 are therefore valid if the functionals involved have sufficient smoothness.

3. Cumulative regression function tests for regression data. The aim of this work is to establish a nonparametric test to compare the regression functions for two independent groups of subjects. In this section, we provide a description of the asymptotic distribution of a test based on the cumulative regression function (CRF). Assume we have observed two groups of subjects over a time period from $[0, t]$, and want to compare the regression functions from the two groups. We consider the process, $T(z)$, which is defined as

$$T(z) = \int_a^z (\hat{m}_1(y) - \hat{m}_2(y))\,dy \quad \text{as defined,} \tag{12}$$

where a is introduced to avoid edge effects of the kernel estimators, and both a and z are points in d-dimensional Euclidean space, that is, in the space of the regressors. For given a and z, the statistic $T(z)$ gives the difference in the CRF's over the interval $[a, z]$.

Based on this quantity, we define the two-sample CRF test statistic of the hypothesis $H\colon m_1(\cdot) = m_2(\cdot)$ on the interval $[a, S-a]$ as $T(S-a)$, where S is a point in d-dimensional Euclidean space and the upper limit of comparison. The interval of comparison must be chosen as some relevant subset of the regressor space. Similarly, a maximal deviation test statistic of the same hypothesis may be defined as

$$M = \sup_{z \in [a,\, S-a]} |T(z)| \quad \text{as defined.} \tag{13}$$

The avoidance of the edge area through narrowing the interval of comparison by a at both ends is necessary because the ND estimator has a bias of order $O(b)$ in edge area, in contrast to the bias in nonedge areas, which is of order $O(b^2)$.

The two-sample CRF test will be sensitive to alternatives where the compared functions bound each other on the interval of interest. This is in analogy with the properties of the log-rank test of survival analysis and the Wilcoxon test. If the compared functions are ordered in size, however, one would expect the two-sample test to have superior power compared to the maximal deviation test. The asymptotics of the test statistics are obtained through the following theorem, which gives the asymptotic distribution of the process

$$I_k(z) = \int_a^z (\hat{m}_k(y) - m_k(y))\,dy \quad \text{as defined.} \tag{14}$$

PROPOSITION 4. *Assume that the assumptions of Proposition* 2 *are satisfied with* $[a, S-a] \subset A$ *and assume further that*

(i)
$$H_k(z) = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \sigma_k^2(V_s^{k,i})\left[\int_a^z \frac{1}{\alpha_k(y)}\frac{1}{|b|}K(y - V_s^{k,i}, b)\,dy\right]^2 \lambda_s^{k,i}\,ds \to_p h_k(z);$$

(ii) *with* $b_{\min} = \min(b_1, \dots, b_d)$, $\sup_{y\in[a,S-a]} b_{\min}^{-2}|\alpha_k^\star(y) - \alpha(y)| = O_p(1)$, *and* $b_{\max} = \max(b_1, \dots, b_d) = o(n^{-\{1/4\}})$.

Then $\sqrt{n_k}I_k(z)$ *converges in distribution in* $C[a, S-a]$ (*with respect to the Uniform topology*) *for* $d \leq 3$, *to a Gaussian Martingale with mean zero and variance function given by* $h_k(z)$.

The theorem is valid for three-dimensional covariates and may be extended to higher dimensions if a bias-corrected estimator replaces the ND estimator of the previous section. The restriction on the dimensionality of the covariate process is a consequence of $b_{\max} = o(n^{-\{1/4\}})$, which is necessary to make the bias asymptotically negligible. Note that a one-sample test may be established based on this theorem, see the discussion of confidence bands below.

The variance function can be estimated through the estimators of the previous section, namely, define

$$\widehat{H_k(z)} = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re} H_{k,i}\big(s, w, z, \widehat{\alpha_k}(\cdot), \widehat{m_k}(\cdot)\big)P^{k,i}(ds \times dw), \tag{15}$$

where

$$H_{k,i}(s, w, z, \alpha_k(\cdot), m_k(\cdot)) = \left[\int_a^z \frac{(w - m_k(y))}{\alpha_k(y)}\frac{1}{|b|}K(y - V_s^{k,i}, b)\,dy\right]^2.$$

PROPOSITION 5. $\widehat{H_k(z)}$ *is a consistent estimator of the variance function* $h_k(z)$. *In the case of independent identically distributed subjects, the variance function* $h_k(z)$ *is equal to*

$$\int_a^z \frac{\sigma_k^2(y)}{\alpha_k(y)}\,dy.$$

Note that the variance expression from the i.i.d. situation may be consistently estimated by simply plugging in the estimates of the functionals $\sigma_k^2(\cdot)$ and $\alpha_k(\cdot)$.

Confidence bands. For the one-sample case, it follows from Propositions 4 and 5 that an asymptotic $100(1-\alpha)\%$ confidence band for $\int_a^z m_k(y)\,dy$ is given by

$$\int_a^z \hat{m}_k(y)\,dy \pm c_\alpha n_k^{-1/2}\hat{H}_k(S-a)^{1/2}\left(1 + \frac{\hat{H}_k(z)}{\hat{H}_k(S-a)}\right), \qquad z \in [a, S-a],$$

where c_α is the upper α quantile of the distribution of $\sup_{t\in[0,1/2]}|B^0(t)|$ and B^0 is a Brownian bridge. Hall and Wellner (1980) gave the table for c_α.

Now, considering two groups of data, the test statistics T can be written as

$$T(z) = I_1(z) - I_2(z).$$

The following lemma that follows from Proposition 4 provides the asymptotic distribution of the test statistic.

LEMMA 1. *If $n_j/(n_1+n_2) \to p_j$ for $j = 1, 2$, it follows that $\sqrt{n_1+n_2}T(z)$ converges toward a Gaussian Martingale with mean zero (under the null hypothesis) and variance function*

$$p_1^{-1}h_1(z) + p_2^{-1}h_2(z)$$

that can be estimated consistently by the above estimators of $h_1(\cdot)$ and $h_2(\cdot)$.

REMARK. The lemma is valid for three-dimensional covariates only, since it is based on Proposition 4. When the underlying intensities $\alpha_k(y)$, $k = 1, 2$, are equal, however, the result of Lemma 1 will be valid in higher dimensions because the bias terms will cancel out.

By Lemma 1, a two-sample test can be carried out using the test statistics $T(S-a)/\sqrt{\widehat{\mathrm{Var}}(T(S-a))}$, which has an asymptotically standard normal distribution under the null hypothesis of $m_1(z) = m_2(z)$ on $[a, S-a]$. Also simultaneous confidence bands for $\int_a^z m_1(y)\,dy - \int_a^z m_2(y)\,dy$ can be constructed as above in the one-sample case. Some plots of such confidence bands are given in Section 4.

The above analysis of the regression functions can also be carried out for the variance functions. Then one would study the integrated variance function and show that

$$S_k(z) = \int_a^z (\hat{V}_k(y) - \sigma_k^2(y))\,dy \quad \text{as defined} \tag{16}$$

converge toward a Gaussian Martingale under appropriate conditions.

The comparison of two groups of data is particularly simple. One may, however, need to compare K regression functions. Following the analogy of the K-sample test in survival analysis, a test may be based on the asymptotic distribution of estimators of $\int_a^{S-a} m_k\,dz - \int_a^{S-a} m_{\mathrm{tot}}\,dz$ for $k = 1, \ldots, K-1$. Here $m_{\mathrm{tot}}(\cdot)$ is the common regression function, that is, the true regression function under the null hypothesis. Various estimators of the true common regression function may be used, leading to slightly different tests.

4. Numerical results and example.

4.1. *Numerical results.* We have carried out three limited simulation studies to study the finite sample properties of our estimators and test statistics. Some dependency structure among the observations has been introduced in these simulation studies. In the first simulation study, the covariate is the random observation time, which depends on the previous measurement. In the second simulation study, we let the regression function be a function of the previous measurement and let the observation time depend on the previous measurement as well. Finally, a two-dimensional regression function was considered. This example indicates that for the multidimensional case a much larger sample size is needed in order to obtain a reasonable power of our test compared to the one-dimensional case. In these simulation studies, $\varepsilon_{k,i,j}$ is Gaussian white noise with mean zero and standard deviation 0.1, and a uniform kernel is used in all simulation studies. In the first simulation study, we also generate noise with a standard deviation of 0.5. A much larger variation among the measurements is introduced here and the simulation results showed that the test reached reasonable power slowly under the alternative case as the sample size increased. The two-dimensional case showed that that the CRF test is more sensitive to the choice of bandwidth than the one-dimensional case. To obtain a reasonable performance of our test statistic we therefore chose different bandwidths depending on the sample sizes in the third simulation study (see Table 3). Based on our limited simulation studies, we find that the choice of bandwidth is important to obtain a good performance of the test statistic; this is particularly important in the multidimensional case. Therefore, in practice, some care needs to be taken in choosing the bandwidth to ensure that the regression function estimates are not too unstable and at the same time do not introduce too much bias.

Simulation Study 1. Let the covariate $V^{k,i}_{\tau_{k,i,j}} = \tau_{k,i,j}$ be the random observation time. The two regression functions used here are:

Model 1. $m(\tau) = 0.1 + 0.9\tau$.

Model 2. $m(\tau) = 0.1 + 0.9\tau + 5\tau^{1.5}e^{-8\tau}$.

For each individual the observation times were generated from the Poisson process over the unit time period of $[0, 1]$ with parameter λ, where $\lambda = 20$ if the previous response value $Y_{k,i,j-1} \leq 0.6$ and $\lambda = 10$ otherwise, that is, a short follow-up time will occur when $Y_{k,i,j-1} \leq 0.6$, so that the random observation time depends on the value of the previous measurement.

First, we generate two independent samples, both from model 1 for the null case and from models 1 and 2, respectively, under the alternative case, to evaluate the confidence bands for our cumulative regression function estimates and our CRF test. Figure 1a and e show the plots of the raw data. There are 30 individuals in each sample. On the average there are about 15 observations for each individual. The exact average and range of observations are given in the plots. Examining these raw data plots, it is quite difficult to determine whether two samples have the same regression function. Figure 1b, c, f, and g. shows the plots of the estimated and true cumulative regression functions with 95% confidence bands. The bandwidth was taken as 0.075. Inspecting these plots, we find that all true cumulative regression functions are contained within the 95% confidence bands, which indicates that the estimates are consistent. In the two-sample problem, we plot the differences between the estimates of the cumulative function along with its 95% confidence bands

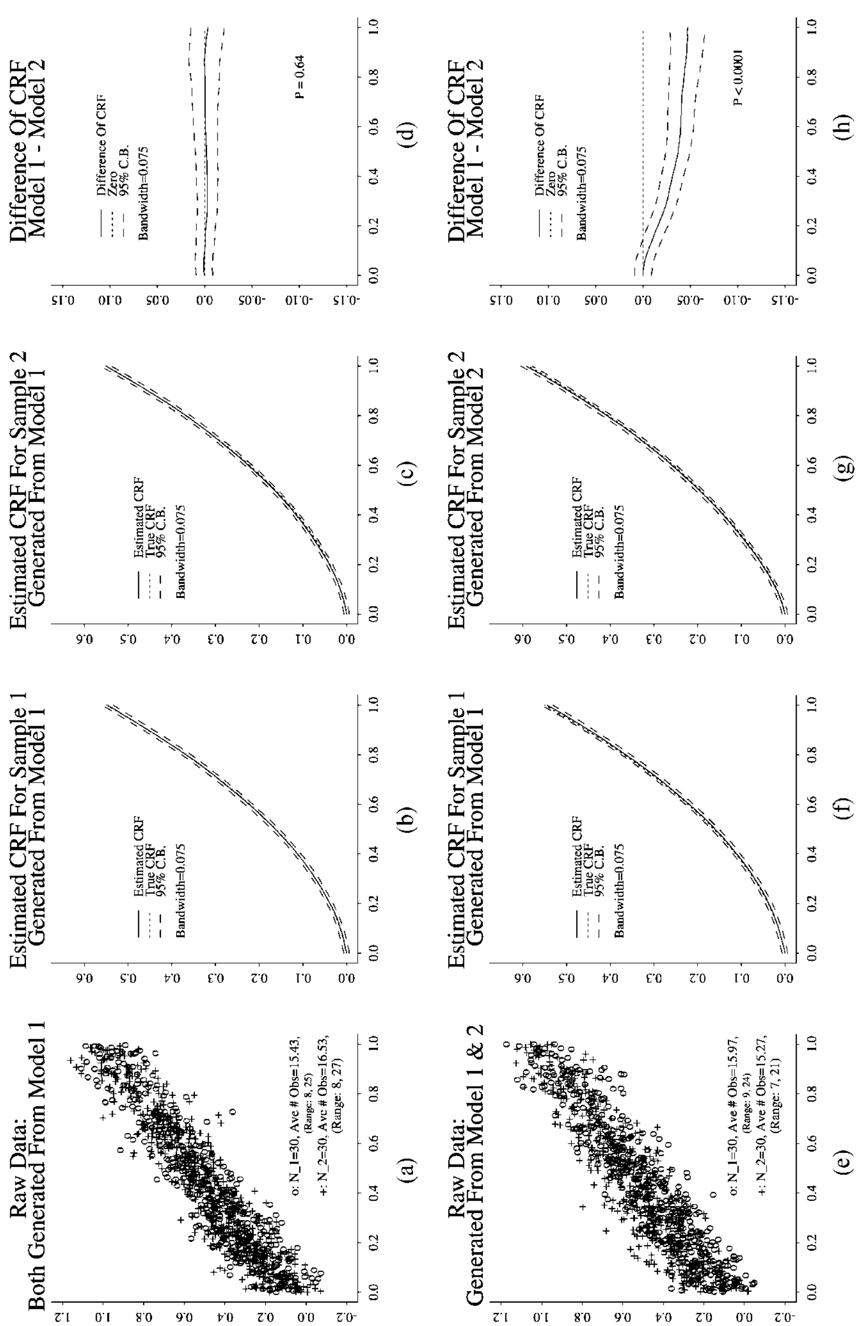

Raw Data:
Both Generated From Model 1
o: N_1=30, Ave # Obs=15.43, (Range: 8, 25)
+: N_2=30, Ave # Obs=16.53, (Range: 8, 27)
(a)
Estimated CRF For Sample 1
Generated From Model 1
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.075
(b)
Estimated CRF For Sample 2
Generated From Model 1
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.075
(c)
Difference Of CRF
Model 1 - Model 2
Difference Of CRF
Zero
95% C.B.
Bandwidth=0.075
P = 0.64
(d)
Raw Data:
Generated From Model 1 & 2
o: N_1=30, Ave # Obs=15.97, (Range: 9, 24)
+: N_2=30, Ave # Obs=15.27, (Range: 7, 21)
(e)
Estimated CRF For Sample 1
Generated From Model 1
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.075
(f)
Estimated CRF For Sample 2
Generated From Model 2
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.075
(g)
Difference Of CRF
Model 1 - Model 2
Difference Of CRF
Zero
95% C.B.
Bandwidth=0.075
P < 0.0001
(h)

FIG. 1.

(Figure 1d and h). Again the bandwidth was taken as 0.075 here. As we mentioned early, it is difficult to see the difference of two regression functions by examining the plots of the raw data (see Figure 1a and 1e). Also, in Figure 1d, the zero function is contained within the 95% confidence band, and our CRF test gives a p-value of 0.64, so our CRF test would correctly conclude that two regression functions are identical. In Figure 1h, the zero function is well outside the 95% confidence band with $p < 0.0001$ for the CRF test; thus the two-sample test correctly concludes that two regression functions are different.

We generated 5000 samples in each run to study the performance of our two-sample test (Table 1.1). Under the null case, for the small sample sizes ($N \leq 10$) the observed test sizes do not reach their nominal 5% values, so the test is not valid for these small sample size cases. Starting from sample size 15, the observed test sizes are close to their nominal 5% values and the powers are close to 100% under the alternative case.

Next, we considered the same models and generated $\varepsilon_{k,i,j}$ with a standard deviation of 0.5. A much larger variation among the measurements is introduced here. The bandwidth was taken as 0.05 to be comparable with the sample sizes. The number of samples in each run was 5000. The numerical results are given in Table 1.2. Under the null case, the observed test sizes are close to their nominal 5% values for a relatively small sample size, but the power increases slowly under the alternative case as the sample size increases. For sample size 120 the test only has power of 0.55. This numerical example indicates that a much larger sample size is needed in order to obtain a reasonable power for the test when the variation among the measurements increases.

In practice, we suggest plotting the raw data, estimated cumulative regression function with its confidence band and estimated (cumulative) vari-

TABLE 1.1

Observed test sizes and powers of CRF test at nominal level of 5%; *bandwidth* $b = 0.075$. *The data were generated with* $\varepsilon_{k,i,j} \sim N(0, 0.1)$. *The number of samples in each run was* 5000

	Sample 1		**Sample 2**		
	N_1	**Ave # obs (range)**	N_2	**Ave # obs (range)**	***P*** **(95% C.I.)**
Null case	5	15.78(3, 35)	5	15.79(4, 35)	0.106(0.097, 0.115)
	10	15.78(3, 33)	10	15.80(2, 36)	0.072(0.065, 0.079)
	15	15.79(2, 35)	15	15.75(3, 35)	0.054(0.048, 0.061)
	20	15.78(2, 36)	20	15.77(3, 36)	0.060(0.053, 0.066)
	25	15.80(2, 36)	25	15.77(2, 36)	0.048(0.042, 0.054)
	30	15.79(2, 35)	30	15.79(2, 36)	0.046(0.040, 0.052)
Alternative case	5	15.78(3, 36)	5	15.35(2, 32)	0.506(0.492, 0.520)
	10	15.77(2, 36)	10	15.33(3, 34)	0.762(0.751, 0.774)
	15	15.79(2, 36)	15	15.32(2, 35)	0.902(0.894, 0.911)
	20	15.80(3, 33)	20	15.34(2, 36)	0.961(0.955, 0.966)
	25	15.79(2, 35)	25	15.34(3, 34)	0.978(0.974, 0.982)
	30	15.80(2, 37)	30	15.33(1, 35)	0.990(0.987, 0.992)

TABLE 1.2
Observed test sizes and powers of CRF test at nominal level of 5%; bandwidth $b = 0.05$. The data were generated with $\varepsilon_{k,i,j} \sim N(0, 0.5)$. The number of samples in each run was 5000

	Sample 1		Sample 2		
	N_1	Ave # obs (range)	N_2	Ave # obs (range)	P (95% C.I.)
Null case	20	14.52(1, 36)	20	14.52(1, 33)	0.069(0.062, 0.076)
	40	14.51(1, 34)	40	14.51(1, 35)	0.055(0.048, 0.061)
	60	14.52(1, 36)	60	14.53(1, 37)	0.056(0.050, 0.062)
	80	14.51(1, 39)	80	14.53(1, 37)	0.055(0.049, 0.062)
	100	14.52(1, 36)	100	14.52(1, 36)	0.050(0.044, 0.056)
	120	14.53(1, 35)	120	14.52(1, 38)	0.044(0.038, 0.049)
Alternative case	20	14.54(1, 36)	20	14.24(1, 35)	0.142(0.133, 0.152)
	40	14.54(1, 40)	40	14.23(1, 35)	0.232(0.221, 0.244)
	60	14.52(1, 34)	60	14.24(1, 37)	0.316(0.303, 0.329)
	80	14.53(1, 37)	80	14.24(1, 35)	0.411(0.397, 0.425)
	100	14.53(1, 36)	100	14.23(1, 37)	0.478(0.464, 0.492)
	120	14.52(1, 36)	120	14.24(1, 36)	0.548(0.534, 0.562)

ance function. Then we would examine these plots carefully to determine the smoothness of the regression function and whether a large variation is involved among the observations. Clearly, more research into this area is needed.

Simulation Study 2. Here the covariate is the previous response value, that is, $V^{k,i}_{\tau_{k,i,j}} = Y_{k,i,j-1}$. The data were generated from the following two regression functions:

Model 1. Linear autoregressive model, $m(V^{k,i}_{\tau_{k,i,j}}) = 0.8Y_{k,i,j-1}$.

Model 2. Threshold autoregressive model,

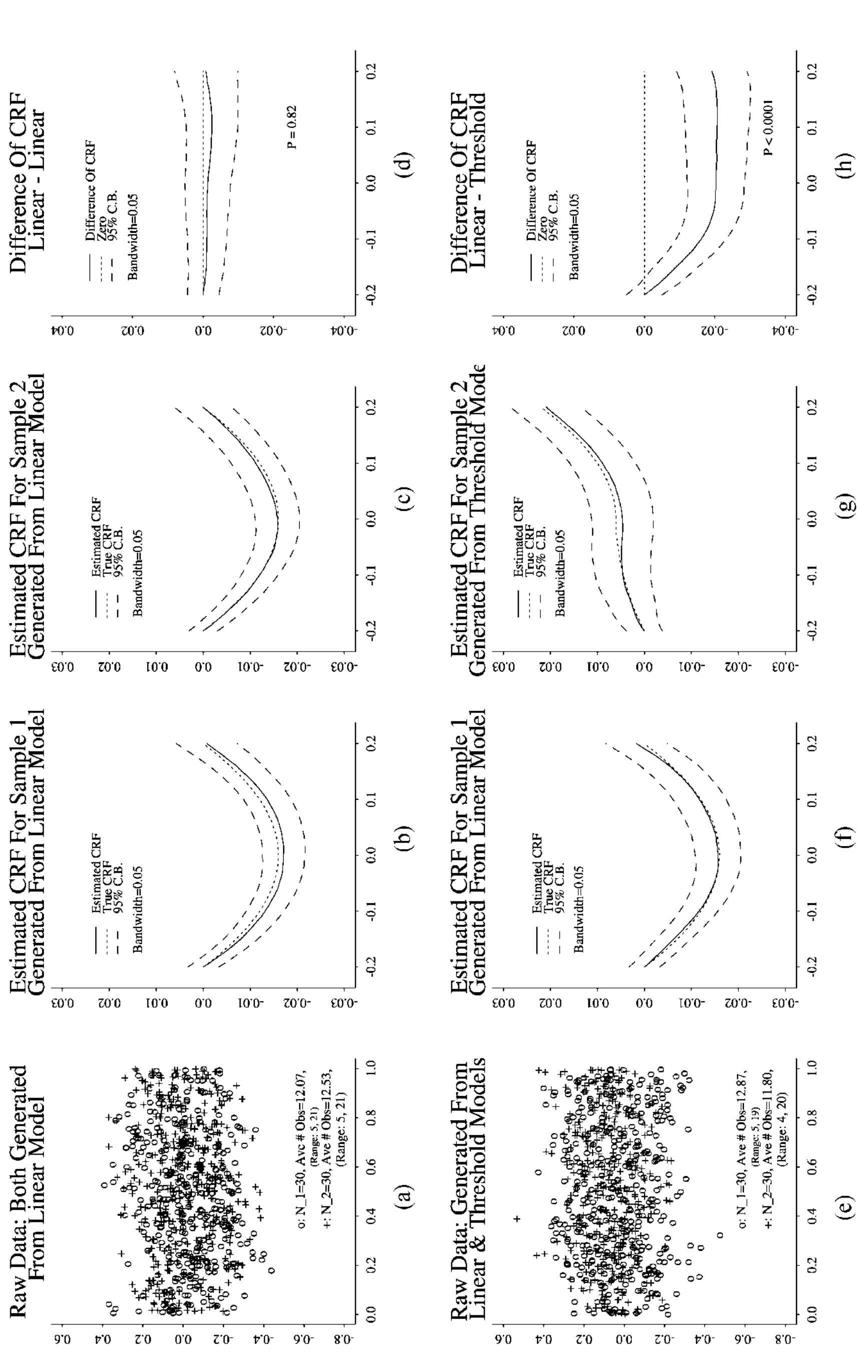
$$m(V^{k,i}_{\tau_{k,i,j}}) = \begin{cases} -0.3Y_{k,i,j-1}, & \text{if } Y_{k,i,j-1} \le 0, \\ 0.8Y_{k,i,j-1}, & \text{if } Y_{k,i,j-1} > 0. \end{cases}$$

The observation times were generated from the Poisson process over the unit time period of $[0, 1]$ with parameter λ, where $\lambda = 20$ if the previous response value $Y_{k,i,j-1} > 0.15$ or $Y_{k,i,j-1} < -0.15$, and $\lambda = 10$ otherwise. The simulation results are given in Figure 2 and Table 2. Examining the plots and the table, a similar conclusion is obtained. Furthermore, inspecting Figure 2f and g, we find that these two models have distinct shapes of the cumulative regression functions, and they are easily distinguishable.

Simulation Study 3 (Two-dimensional case). In this study, the covariate $V^{k,i}_{\tau_{k,i,j}} = (X_{k,i}, \tau_{k,i,j})$, where $X_{k,i}$ is an additional covariate for each individual and $\tau_{k,i,j}$ is the random observation times. The models used in this simulation study are:

Model 1. $m(V^{k,i}_{\tau_{k,i,j}}) = 0.3X_{k,i} + (0.1 + 0.9\tau_{k,i,j})$.

Model 2. $m(V^{k,i}_{\tau_{k,i,j}}) = 0.3X_{k,i} + (0.1 + 0.9\tau_{k,i,j}) + 5\tau^{1.5}_{k,i,j}\exp(-8\tau_{k,i,j})$.



Raw Data: Both Generated From Linear Model
o: N_1=30, Ave # Obs=12.07, (Range: 5, 21)
+: N_2=30, Ave # Obs=12.53, (Range: 5, 21)
(a)
Estimated CRF For Sample 1 Generated From Linear Model
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.05
(b)
Estimated CRF For Sample 2 Generated From Linear Model
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.05
(c)
Difference Of CRF Linear - Linear
Difference Of CRF
Zero
95% C.B.
Bandwidth=0.05
P = 0.82
(d)
Raw Data: Generated From Linear & Threshold Models
o: N_1=30, Ave # Obs=12.87, (Range: 5, 19)
+: N_2=30, Ave # Obs=11.80, (Range: 4, 20)
(e)
Estimated CRF For Sample 1 Generated From Linear Model
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.05
(f)
Estimated CRF For Sample 2 Generated From Threshold Mode
Estimated CRF
True CRF
95% C.B.
Bandwidth=0.05
(g)
Difference Of CRF Linear - Threshold
Difference Of CRF
Zero
95% C.B.
Bandwidth=0.05
P < 0.0001
(h)

FIG. 2.

TABLE 2
Observed test sizes and powers of CRF test at nominal level of 5%; bandwidth $b = 0.05$. The data were generated with $\varepsilon_{k,i,j} \sim N(0, 0.1)$. The number of samples in each run was 5000

	Sample 1		Sample 2		
	N_1	Ave # obs (range)	N_2	Ave # obs (range)	P (95% C.I.)
Null case	5	12.07(2, 31)	5	12.08(2, 30)	0.109(0.100, 0.117)
	10	12.11(2, 32)	10	12.11(2, 32)	0.064(0.057, 0.071)
	15	12.10(2, 33)	15	12.08(2, 33)	0.059(0.052, 0.066)
	20	12.12(2, 30)	20	12.10(2, 31)	0.052(0.046, 0.058)
	25	12.10(2, 34)	25	12.10(2, 35)	0.056(0.049, 0.062)
	30	12.10(2, 35)	30	12.10(2, 32)	0.050(0.044, 0.056)
Alternative case	5	12.09(2, 31)	5	11.80(2, 30)	0.677(0.664, 0.690)
	10	12.10(2, 31)	10	11.84(2, 30)	0.905(0.896, 0.913)
	15	12.11(2, 32)	15	11.83(2, 33)	0.980(0.976, 0.984)
	20	12.10(2, 31)	20	11.81(2, 32)	0.997(0.996, 0.999)
	25	12.11(2, 35)	25	11.80(2, 32)	1.000(0.999, 1.00)
	30	12.08(2, 32)	30	11.81(2, 33)	1.000(1.00, 1.00)

The observation times $\tau_{k,i,j}$ were generated from a Poisson process with the same parameter given in Simulation Study 1, and the covariate processes $X_{k,i}$ were generated from the random variables with density functions $f_1(x) = 3/4 + x/2$ and $f_2(x) = 5/4 - x/2$ for $x \in [0, 1]$, respectively. Here, in the alternative case, $m_1 < m_2$ for the same given covariate values of X and τ, and $E(X_{1,i}) = 0.54 > E(X_{2,i}) = 0.46$; therefore it would be difficult to detect the difference without adjusting the covariate of $X_{k,i}$.

Table 3 gives the observed test sizes and powers of our two-sample test. The observed test sizes are close to their nominal 5% values starting from

TABLE 3
Observed test sizes and powers of CRF test at nominal level of 5%. The data were generated with $\varepsilon_{k,i,j} \sim N(0, 0.1)$. The number of samples in each run was 5000

Sample 1			Sample 2			
N_1	$b_1 = b_2$	Ave # obs (range)	N_2	$b_1 = b_2$	Ave # obs (range)	P (95% C.I.)
			Null case			
25	0.10	14.00(2, 34)	25	0.10	14.27(2, 33)	0.177(0.167, 0.188)
50	0.08	14.00(1, 35)	50	0.08	14.26(1, 34)	0.058(0.052, 0.065)
75	0.07	13.99(1, 34)	75	0.07	14.26(1, 34)	0.041(0.036, 0.047)
100	0.06	13.99(1, 33)	100	0.06	14.26(1, 35)	0.045(0.040, 0.051)
125	0.05	13.99(1, 36)	125	0.05	14.28(1, 36)	0.056(0.050, 0.063)
			Alternative case			
25	0.10	13.99(2, 33)	25	0.10	13.54(2, 33)	0.864(0.855, 0.874)
50	0.08	13.99(1, 37)	50	0.08	13.57(1, 34)	0.993(0.990, 0.995)
75	0.07	13.99(1, 33)	75	0.07	13.56(1, 34)	1.000(0.999, 1.00)
100	0.06	13.99(1, 34)	100	0.06	13.57(1, 35)	1.000(1.00, 1.00)
125	0.05	13.98(1, 36)	125	0.05	13.57(1, 35)	1.000(0.999, 1.00)

sample size 50 ($P = 0.058$ for $N_1 = N_2 = 50$) under the null case. Therefore, relatively large sample sizes are required in the higher dimensional cases. The powers for the alternative case are very close to 100% for sample sizes ≥ 50, where the observed test sizes are close to their nominal 5% values.

The simulations indicate that our two-sample CRF test is consistent and has sufficient power to detect the difference between the regression functions in our simulations. The simulations further reveal that some caution has to be taken in choosing the smoothing parameters appropriately, and this is an area for further research.

4.2. *Example.* Finally, we applied the regression two-sample test to compare the growth for two diagnoses of dwarfism, namely, hypochondroplasia (Hypo) and achondroplasia (Acho). Our data were provided by the Department of Growth and Reproduction at the University Hospital in Copenhagen and consist of longitudinal measurements of height and weight for 36 patients with hypochondroplasia and 42 patients with achondroplasia. The data on height versus age for the two groups are displayed in Figure 3a and b, and it appears that the Hypo diagnosis results in a somewhat better growth. Figure 3c and 3d shows the similar weight versus age plots for the two groups. The weight versus age charts suggest that the differences between the two groups are much less pronounced than for height. This is consistent with the general impression of Acho patients being somewhat stockier. Due to the longitudinal nature of the data, a comparison of the growth of the two groups cannot be based on a simple height/weight for age reference chart. We therefore consider the longitudinal regression model where the regression function is a function of the previous measurement and the time since the previous measurement, that is,

$$Y_{k,i,j} = m_k(Y_{k,i,j-1}, \tau_{k,i,j} - \tau_{k,i,j-1}) + \varepsilon_{k,i,j}, \quad (17)$$

for $j = 2, \dots, N_i^k$, $i = 1, \dots, n_k$, $k = 1, 2$.

A biologically more satisfactory model would be to include the current age in the above model, but age did not contribute much additional information in terms of predicting the growth of Hypo or Acho patients; this is partly due to the design of the regressors. Therefore, when the objective is solely to compare the growth patterns, the above simpler submodel may be used to provide a good approximation. When a difference is found, it can be concluded that the longitudinal growth is different for the two diagnoses. In contrast, however, equivalent behavior for two groups based on the limited model will only make this conclusion valid for the observed age span. The residuals in the model represent further biological variation and measurement error and were expected to be right-skewed and biased for small values of time increments. Residual plots, however, revealed that this was not a serious problem for our data.

Note that we consider only individuals with more than one measurement and analyze the data conditional on the first measurement for each individual. First considering height, it appears that patients with hypochondroplasia grow

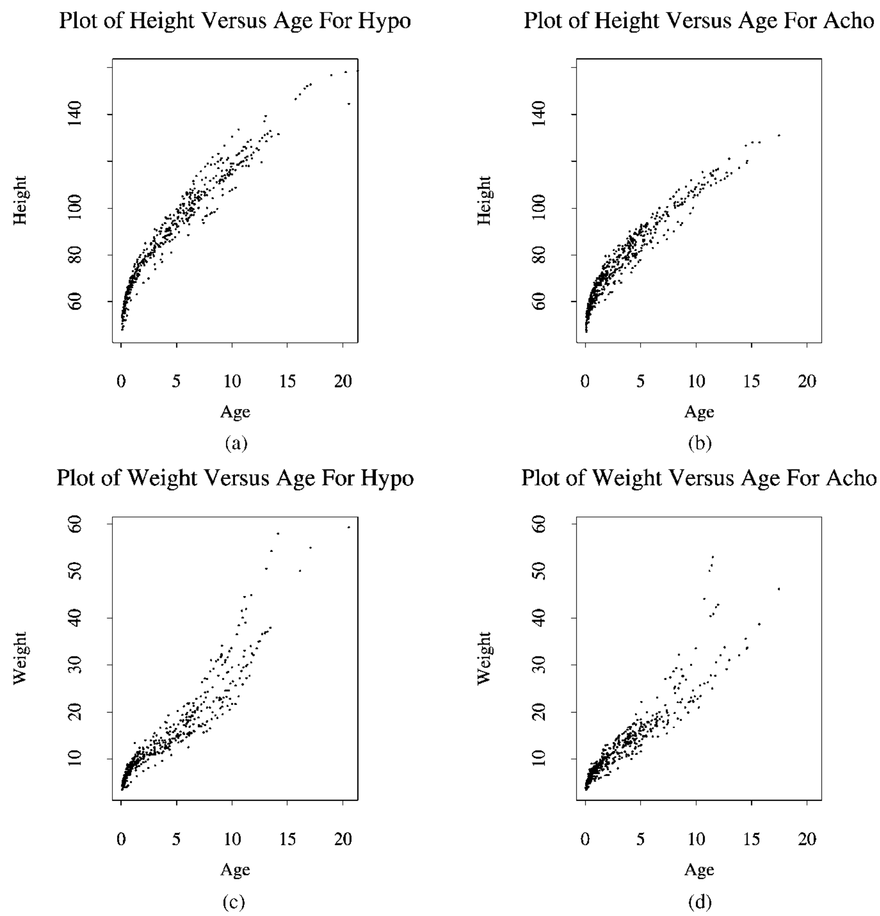


FIG. 3.

faster than patients with achondroplasia, and if we apply our two-sample test to the the following region of previous height and time since previous measurement, $[50, 120] \times [0.2, 1.9]$, our test statistic can be calculated for a choice of the two-dimensional bandwidth. Figure 4a shows the difference in the CRF's normalized with its variance for $b_1 = 5.0$ and $b_2 = 0.2$. For this choice of bandwidths we get a two-sample test statistic at 94 with variance 586, and this results in a test statistic that is asymptotically normal at 3.8, which results in a p-value at 0.0001. Further smoothing of the regression functions results in the same conclusion although the test statistic decreases some. Note that one would expect the test statistic to have good power in this application since the Hypo diagnosis appears to result in a better growth than the Acho diagnosis all at ages. Comparing the two conditional regression functions at fixed points, the conclusion is less clear. For example, considering previous height 90 cm and time since previous measurement 0.2 years, the standardized difference is 1.08, which is asymptotically standard normal (since the bias becomes asymp-

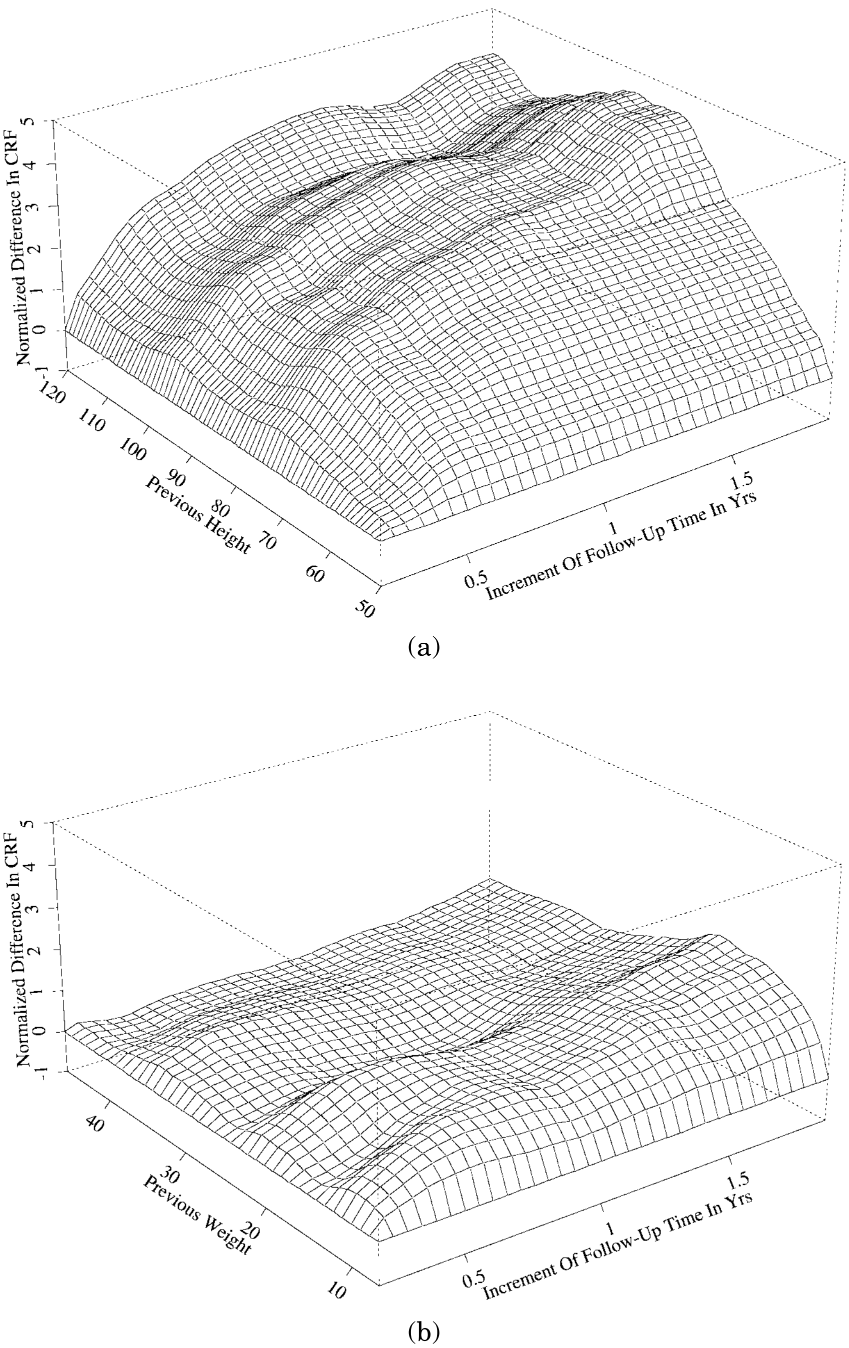


FIG. 4. *Normalized difference in cumulative regression function of* (*Hypo–Acho*) *with bandwidth* $= 5.0, 0.2$) *and CRF test of* (a) $p = 0.0001$ *and* (b) $p = 0.48$.

totically negligible) and thus results in a p-value at 0.2762. There are actually no single point where a comparison would result in a statistically significant difference.

A similar comparison of the increase of weight did not result in a significant difference between the two groups, with a p-value at 0.48 for the bandwidths chosen as for height. Figure 4b shows the difference in the CRF's normalized with its variance for $b_1 = 5.0$ and $b_2 = 0.2$. It appears that Hypo results in a slightly higher weight that, however, is not statistically different.

5. Discussion. We have in this work presented a nonparametric two-sample test for comparison of regression data. We believe that the test statistic, which is based on the cumulative regression function (CRF), often provides a reasonable nonparametric two-sample test. An asymptotic result is established for the CRF in low dimensions ($d \leq 3$). This result may be utilized in many directions. One may, for example, construct a test for additivity as in McKeague and Utikal (1990b), see also McKeague and Utikal (1990a), or establish a goodness-of-fit test along the lines of McKeague and Zhang (1994). The methodology may be used to compare longitudinal data through comparison of the conditional regression functions, and we showed as a special case that usual independent identically distributed regression data may be compared by the proposed methodology.

The power and applicability of the CRF test as well as the maximal deviation test depend on the size of the asymptotically negligible bias. The bias depends on the smoothness of the regression functions and may be a serious problem when the regression functions are varying rapidly. However, when the underlying intensities $\alpha_k(y)$, $k = 1, 2$, are equal, the bias terms will cancel out and the result may be extended beyond three dimensions.

The CRF test may be extended slightly by introducing a weight function. This may be a relevant area for further research.

A small difficulty arises from estimating the regression function by a Nadaraya–Watson estimator: for this estimator, the bias is more severe at the edges of the data and therefore one must limit the results to nonedge areas. One may in practice use an estimator without edge effects, for example, the local linear smoother (LLS) [see Fan (1992)], and thereby avoid the narrowing of the interval of comparison. The proof presented in this work, however, is based on the Nadaraya–Watson (ND) estimator. The ND estimator has a bias of $O(b)$ in the edge area, whereas the LLS smoother has a bias of $O(b^2)$ everywhere. The asymptotic variance is the same for both estimators $[O(1/(nb))]$ and therefore the results of this work also should carry through for edge areas when the LLS estimator is used, although some corrections in the variance formulas still are needed at the edges. The LLS estimator is more difficult to analyze since the design is used more actively, and a particular complication is that things no longer can be written as integrals of predictable functions.

APPENDIX

This Appendix contains the proofs of Propositions 2–5. Convergence of processes are with respect to the uniform topology.

PROOF OF PROPOSITION 2. Note that $\hat{\alpha}_k(y)-\alpha^\star_k(y)$ and $D_k(y)-D^\star_k(y)$ [see (6)–(9)] are square-integrable 0-Martingales if (4) is satisfied with $H_{k,i}(s,z) = (1/|b|)K(y-V^{k,i}_s, b)$ and $H_{k,i}(s,z) = (z-m_k(y))(1/|b|)K(y-V^{k,i}_s, b)$, respectively.

In order to show the proposition, it suffices to show that (a) $\sup_{y\in A}|\hat{\alpha}_k(y)-\alpha^\star_k(y)| \to_p 0$, (b) $\sup_{y\in A}|D_k(y)-D^\star_k(y)|\to_p 0$ and (c) $\sup_{y\in A}|D^\star_k(y)|\to_p 0$. We start with (a) and prove that $|\hat{\alpha}_k(y)-\alpha^\star_k(y)|\to_p 0$ for all y and that $|\hat{\alpha}_k(y)-\alpha^\star_k(y)|$ is tight. Tightness is proved by a slight extension of the multiparameter-processes result of Bickel and Wichura (1971).

Here we consider only the case where the conditional density of $V^{k,i}_s$ given $\lambda^{k,i}_s$ is $f^{k,i}_s(v)$, since the other case, where $V^{k,i}_s$ given $\lambda^{k,i}_s$ is a product of the Dirac measure at s, ε_s and a bounded density $f^{k,i}_s(v_1,\dots,v_{d-1})$, follows similarly.

Let $(1/|b|)\tilde{K}(s) = (1/|b|)K(y-V^{k,i}_s, b)$. Note that

$$\begin{aligned} E|\hat{\alpha}_k(y)-\alpha^\star_k(y)|^2 &= E\biggl(\frac{1}{n_k^2}\sum_{i=1}^{n_k}\int_0^t \frac{1}{|b|^2}\tilde{K}^2(s)\lambda^{k,i}_s\,ds\biggr)\\ &= \frac{1}{n_k^2|b|^2}E\biggl(\sum_{i=1}^{n_k}\int_0^t\int_{\Re^d}K^2(y-v,b)f^{k,i}_s(v)\,dv\lambda^{k,i}_s\,ds\biggr)\\ &= \frac{1}{n_k^2|b|}E\biggl(\sum_{i=1}^{n_k}\int_0^t\int_{-1}^1 K^2(u)f^{k,i}_s(y-bu)\,du\lambda^{k,i}_s\,ds\biggr)\\ &\le C\frac{1}{n_k^2|b|}E\biggl(\sum_{i=1}^{n_k}\int_0^t\lambda^{k,i}_s\,ds\biggr). \end{aligned}$$

Here we have used boundedness of the conditional density. If $V^{k,i}_s$ is time, the proof follows similarly. This proves that $|\hat{\alpha}_k(y)-\alpha^\star_k(y)|\to_p 0$ for all y, because of (iii) and (iv).

We now prove tightness. Following Bickel and Wichura (1971), a block B in $[0,1]^d$ is a subset of the form $(y_1,y_2] = \prod_{p=1}^d(y_{1,p},y_{2,p}]$. For a block we define

$$\begin{aligned} X(B) = \sum_{\varepsilon_1=0,1}\cdots\sum_{\varepsilon_d=0,1}(-1)^{\{d-\sum\varepsilon_p\}}X\bigl(y_{1,1}+\varepsilon_1(y_{2,1}-y_{1,1}),\dots,y_{1,d}\\ +\varepsilon_d(y_{2,d}-y_{1,d})\bigr). \end{aligned}$$

According to Bickel and Wichura, tightness follows if we can show that $E(|X(B)||X(C)|)\le\mu(B)\mu(C)$ for all neighboring blocks B and C, where μ is some measure. Using Cauchy–Schwarz, it suffices to show that $E(|X(B)|^2)\le\mu(B)^2$. We prove that $E(|X(B)|^2)\le\mu(B)^2+o(1)$, where $o(1)$ is independent of B and thus we use a slight extension of the result of Bickel and Wichura; see McKeague and Zhang [(1994), page 507].

For a d-dimensional block $(y_1,y_2]=\prod_{p=1}^d(y_{1,p},y_{2,p}]$, we consider

$$E\bigl(\hat{\alpha}_k(B)-\alpha^\star_k(B)\bigr)^2.$$

Let $V_s^{k,i}(p)$ denotes the pth component of $V_s^{k,i}$. Due to the choice of product kernels we can write $\hat{\alpha}_k(B) - \alpha_k^\star(B)$ as an integral of

$$H_{k,i}(s,z) = \frac{1}{|b|}\prod_{p=1}^{d}\left[K(y_{2,p} - V_s^{k,i}(p), b_p) - K(y_{1,p} - V_s^{k,i}(p), b_p)\right]$$

with respect to the marked point process Martingale, and can thus utilize Proposition 1 when calculating the above second moment.

We consider the one-dimensional case, since the multidimensional case follows similarly, and thus we need to show that

$$E\big(\big[\hat{\alpha}_k(y) - \alpha_k^\star(y)\big] - \big[\hat{\alpha}(u) - \alpha_k^\star(u)\big]\big)^2 \le C|y-u|^2 + o(1)$$

for all y, u.

The difference on the left-hand side of the inequality is

$$E\left(\frac{1}{n_k^2}\sum_{i=1}^{n_k}\int_0^t\left[\frac{1}{|b|}K(y - V_s^{k,i}, b) - \frac{1}{|b|}K(u - V_s^{k,i}, b)\right]^2 \lambda_s^{k,i}\, ds\right).$$

The integral over t can be bounded as

$$\begin{aligned}
&\frac{1}{n_k^2 b_1^2}\sum_{i=1}^{n_k}\int_0^t\int_{-1}^{1}\left[K(y-v,b_1) - K(u-v,b_1)\right]^2 f_s^{k,i}(v)\lambda_s^{k,i}\, dv\, ds\\
&\quad= \frac{1}{n_k^2 b_1}\sum_{i=1}^{n_k}\int_0^t\left[\int_{-1}^{1}K^2(v)f_s^{k,i}(y-b_1v)\,dv + \int_{-1}^{1}K^2(v)f_s^{k,i}(u-b_1v)\,dv\right.\\
&\qquad\qquad\left. -2\int_{-1}^{1}K(v)K\left(v - \frac{y-u}{b_1}\right)f_s^{k,i}(y-b_1v)\,dv\right]\lambda_s^{k,i}\,ds\\
&\quad\le \frac{C}{n_k b_1}\left(\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \lambda_s^{k,i}\, ds\right).
\end{aligned}$$

Note that the cross term above vanishes if $|y-u| > 2b_1$. Now, using (iii) and (iv) it follows that $E([\hat{\alpha}_k(y) - \alpha_k^\star(y)] - [\hat{\alpha}(u) - \alpha_k^\star(u)])^2 = o(1)$, since $n_k b_1 \to \infty$. The tightness condition is therefore satisfied if $n_k b_1 \to \infty$. The argument further implies that for any $a_n \to 0$, $\sqrt{n_k|b|}a_n(\hat{\alpha}_k(y) - \alpha_k^\star(y)) \to_p 0$ uniformly. If further the bias of $\hat{\alpha}_k(y)$ is uniformly small, and satisfies $\min(b_1, \dots, b_d)^{-2}\sup_{y\in A}|\alpha_k^\star(y) - \alpha_k(y)| = O_p(1)$, we can obtain a rate of convergence for $\hat{\alpha}_k(y)$. Considering the d-dimensional case, if $\sqrt{n_k|b|} \sim \min(b_1, \dots, b_d)^{-2}$ and $b_j \sim b$, $j = 1, \dots, d$, it follows that $b \sim n_k^{-1/(d+4)}$. Now, $n_k^\nu \sup|\hat{\alpha}_k(y) - \alpha_k(y)| \to_p 0$, where $\nu \le 2/(d+4)$ for $d \le 3$.

Similarly one gets

$$\begin{aligned}
&E\big|D_k(y) - D_k^\star(y)\big|^2\\
&\quad= E\left(\frac{1}{n_k^2}\sum_{i=1}^{n_k}\int_0^t\left[\sigma_k^2(V_s^{k,i}) + (m_k(V_s^{k,i}) - m_k(y))^2\right]\frac{1}{|b|^2}\tilde{K}^2(s)\lambda_s^{k,i}\,ds\right)\\
&\quad\le E\left(\frac{1}{n_k^2}\sum_{i=1}^{n_k}\int_0^t\frac{1}{|b|^2}\tilde{K}^2(s)\lambda_s^{k,i}\,ds\right),
\end{aligned}$$

where the boundedness of the expression inside the square brackets was used. Tightness follows by arguments similar to those above, that also show that for any $a_n \to 0$, $\sqrt{n_k|b|}a_n(D_k(y) - D_k^\star(y)) \to_p 0$ uniformly.

Now,

$$E\big(D_k^\star(y)\big) = E\bigg(\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \big(m_k(V_s^{k,i}) - m_k(y)\big)\frac{1}{|b|}K(y - V_s^{k,i}, b)\,\lambda_s^{k,i}\,ds\bigg)$$

$$= E\bigg(\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re^d}\big(m_k(v) - m_k(y)\big)\frac{1}{|b|}K(y - v, b)f_s^{k,i}(v)\,dv\,\lambda_s^{k,i}\,ds\bigg)$$

$$\le C(b_1^2 + \cdots + b_d^2)E\bigg(\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \lambda_s^{k,i}\,ds\bigg) + o(b_1^2 + \cdots + b_d^2),$$

where the compact support, $e_K = 0$, of the kernel was used and the regression function has a bounded second derivative. Next we verify that $D_k^\star(y)$ can be approximated by its expected value. It follows that

$$\frac{1}{n_k^2}\sum_{i=1}^{n_k}E\bigg(\int_0^t \big(m_k(V_s^{k,i}) - m_k(y)\big)\frac{1}{|b|}K(y - V_s^{k,i}, b)\lambda_s^{k,i}\,ds\bigg)^2 = o(1),$$

since $n_k|b| \to \infty$. Tightness follows similarly. □

PROOF OF PROPOSITION 3. Defining

$$V_k^\star(y) = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \Big[\sigma_k^2(V_s^{k,i}) + \big(m_k(V_s^{k,i}) - m_k(y)\big)^2\Big]\frac{1}{|b|}\tilde{K}(s)\lambda_s^{k,i}\,ds,$$

it follows that $V_k(y) - V_k^\star(y)$ is a square-integrable 0-Martingale (because of the assumption). The proof now follows along the lines of Proposition 2, using the bound on the conditional fourth moments. □

PROOF OF PROPOSITION 4. Recall that $\hat{m}_k(y) - m_k(y) = D_k(y)/\hat{\alpha}_k(y)$; see (5) and (6).

As in Proposition 4, we consider only the case where the conditional density of $V_s^{k,i}$ given $\lambda_s^{k,i}$ is $f_s^{k,i}(v)$; the other case follows similarly.

Define the square-integrable Martingale $X_k(y) = D_k(y) - D_k^\star(y)$; see (7) and (8). Then

$$I_k(z) = \int_a^z \frac{D_k(y)}{\hat{\alpha}_k(y)}\,dy$$

$$= \int_a^z \frac{X_k(y)}{\alpha_k(y)}\,dy - \int_a^z \frac{(\hat{\alpha}_k(y) - \alpha_k(y))}{\hat{\alpha}_k(y)}\frac{X_k(y)}{\alpha_k(y)}\,dy$$

$$+ \int_a^z \frac{D_k^\star(y)}{\alpha_k(y)}\,dy - \int_a^z \frac{(\hat{\alpha}_k(y) - \alpha_k(y))}{\hat{\alpha}_k(y)}\frac{D_k^\star(y)}{\alpha_k(y)}\,dy.$$

We will show that

(a) $$\tilde{I}_k(z) = \sqrt{n_k}\int_a^z \frac{X_k(y)}{\alpha_k(y)}\,dy \to_{\mathscr{D}} \text{Gauss} - h_k(z) \quad \text{as defined;}$$

(b) $$\sqrt{n_k}\int_a^z \frac{(\hat{\alpha}_k(y) - \alpha_k(y))}{\hat{\alpha}_k(y)}\frac{X_k(y)}{\alpha_k(y)}\,dy \to_p 0 \quad \text{uniformly in } z;$$

(c) $$\sqrt{n_k}\int_a^z \frac{D_k^\star(y)}{\alpha_k(y)}\,dy \to_p 0 \quad \text{uniformly in } z;$$

(d) $$\sqrt{n_k}\int_a^z \frac{(\hat{\alpha}_k(y) - \alpha_k(y))}{\hat{\alpha}_k(y)}\frac{D_k^\star(y)}{\alpha_k(y)}\,dy \to_p 0 \quad \text{uniformly in } z,$$

where Gauss $- h_k(z)$ is a Gaussian Martingale with mean zero and variance function $h_k(z)$.

(b) follows from (a), Proposition 2 and that $\alpha_k(\cdot)$ is bounded away from zero, from the continuous mapping theorem.

Namely, define

$$Z_1(y) = \frac{\hat{\alpha}_k(y) - \alpha_k(y)}{\hat{\alpha}_k(y)}$$

and

$$Z_2(y) = \int_a^y \frac{X_k(v)}{\alpha_k(v)}\,dv$$

and consider the mapping $\phi(x, y)(\cdot) = \int_a^{(\cdot)} x\,dy$, see Andersen, Borgan, Gill and Keiding [(1993), page 113] for details (the mapping is continuous on the space of continuous functions). It now follows that

$$\begin{aligned}\phi(Z_1, \sqrt{n_k}Z_2)(z) &= \sqrt{n_k}\int_a^z Z_1(y)\,dZ_2(y)\\ &= \sqrt{n_k}\int_a^z Z_1(y)\frac{X_k(y)}{\alpha_k(y)}\,dy \to_{\mathscr{D}} 0\end{aligned}$$

uniformly in z. Similarly, (d) follows from (c) and Proposition 2.

We, thus, need to show (a) and (c). We start with (c) and prove that

$$\sqrt{n_k}\int_a^z \frac{D_k^\star(y)}{\alpha_k(y)}\,dy \to_p 0.$$

From the proof of Proposition 2 it follows that (c) is satisfied. This term's numerical value can be bounded by

$$\begin{aligned}&\left|\sqrt{n_k}\int_a^z \frac{D_k^\star(y)}{\alpha_k(y)}\,dy\right|\\ &\quad\le \left|\frac{1}{\sqrt{n_k}}\sum_{i=1}^{n_k}\int_a^z\int_0^t \frac{m_k(V_s^{k,i}) - m_k(y)}{\alpha_k(y)}\frac{1}{|b|}K(y - V_s^{k,i}, b)\,\lambda_s^{k,i}\,dy\,ds\right|\\ &\quad\le d_K\frac{C}{\sqrt{n_k}}\sum_{i=1}^{n_k}\int_0^t (b_1^2 + \cdots + b_d^2)\,\lambda_s^{k,i}\,ds,\end{aligned}$$

since $e_K = 0$. This last term converges to 0 if $\max(b_1, \dots, b_d)^2 \sqrt{n_k} = o(1)$, that is, $\max(b_1, \dots, b_d) = o(n_k^{-1/4})$. Note that this rate may be improved by bias correction; this will, however, complicate the proof quite a bit. This condition interplays with the obtained rate of convergence for $\hat{\alpha}_k(y)$ and limits the proof to covariates with dimension less than or equal to 3.

To show (a) we prove that the finite-dimensional distribution of $\tilde{I}_k(z)$ converges to the appropriate Gaussian distribution using the results of Rebolledo (1980) and that it is tight. We start by considering the finite-dimensional distributions

$$\tilde{I}_k(z) = \sqrt{n_k} \int_a^z \frac{X(y)}{\alpha_k(y)}\, dy.$$

Defining

$$g_{k,i}(s, w, z) = \int_a^z \frac{H_{k,i}(s, w, y)}{\alpha_k(y)}\, dy,$$

where

$$H_{k,i}(s, w, y) = (w - m_k(y)) \frac{1}{|b|} K\big(y - V_s^{k,i}, b\big).$$

We have that $\tilde{I}_k(z)$ is a square-integrable Martingale (in t) from Proposition 1. The predictable variation is thus

$$\begin{aligned}
\langle \tilde{I}_k(z) \rangle_t &= \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \int_{\Re^d} g_{k,i}^2(s, w, z) \lambda_s^{k,i}\, dF_s^{k,i}(w - m_k(V_s^{k,i}))\, ds \\
&= \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \sigma_k^2(V_s^{k,i}) \bigg[\int_a^z \frac{1}{\alpha_k(y)} \frac{1}{|b|} K(y - V_s^{k,i}, b)\, dy \bigg]^2 \lambda_s^{k,i}\, ds \\
&\quad + \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \bigg[\int_a^z \frac{m_k(V_s^{k,i}) - m_k(y)}{\alpha_k(y)} \frac{1}{|b|} K(y - V_s^{k,i}, b)\, du \bigg]^2 \lambda_s^{k,i}\, ds \\
&\to_p h_k(z).
\end{aligned}$$

The second term converges to 0 in probability uniformly since the regression function m_k has bounded second derivative, $d_K = O(1)$ and $e_K = 0$, and from assumption (i),

$$h_k(z) = \lim_{n_k \to \infty} \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \sigma_k^2(V_s^{k,i}) \bigg[\int_a^z \frac{1}{\alpha_k(y)} \frac{1}{|b|} K(y - V_s^{k,i}, b)\, dy \bigg]^2 \lambda_s^{k,i}\, ds.$$

Next, we check the Lindeberg condition,

$$\begin{aligned}
L_n(z) &= \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \int_{\Re^d} g_{k,i}^2(s, w, z) I\big(|g_{k,i}(s, w, z)| > \sqrt{n_k}\varepsilon\big) \\
&\qquad \times \lambda_s^{k,i}\, dF_s^{k,i}(w - m_k(V_s^{k,i}))\, ds \quad \text{as defined} \\
&\le \frac{1}{n_k} \sum_{i=1}^{n_k} \int_0^t \int_{\Re^d} \big(w - m_k(V_s^{k,i})\big)^2 I\big(|w - m_k(V_s^{k,i})| > \sqrt{n_k} c\varepsilon\big) \\
&\qquad \times \lambda_s^{k,i}\, dF_s^{k,i}(w - m_k(V_s^{k,i}))\, ds + o_p(1).
\end{aligned}$$

This follows by the simple inequality

$$|a-b|^2 I\big(|a-b|>\varepsilon\big) \le 4|a|^2 I\left\{|a|>\frac{\varepsilon}{2}\right\} + 4|b|^2 I\left\{|b|>\frac{\varepsilon}{2}\right\}.$$

The first term converges to 0 since the variance function $\sigma_k^2(\cdot)$ is bounded.

This shows that the finite-dimensional distribution of $\tilde{I}_k(z)$ converge toward the Gaussian distribution that was claimed. We complete the proof by showing that $\tilde{I}_k(z)$ is tight.

We show that

$$E\big(\tilde{I}_k(z_1)-\tilde{I}_k(z_2)\big)^2 \le C \parallel z_1 - z_2 \parallel^2 + o(1),$$

where $\parallel \cdot \parallel$ is the product norm. First, note that

$$D_t \;=\; \tilde{I}_k(z_1)-\tilde{I}_k(z_2) \quad \text{as defined}$$

is a square-integrable Martingale according to Proposition 1. It follows that

$$\begin{aligned} &E\big(\tilde{I}_k(z_1)-\tilde{I}_k(z_2)\big)^2 \\ &\quad = E\left\{\frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t \sigma_k^2(V_s^{k,i})\big[J_{k,i}(s,z_1)-J_{k,i}(s,z_2)\big]^2 \lambda_s^{k,i}\,ds\right\} + o_p(1), \end{aligned}$$

where

$$J_{k,i}(s,z) = \int_a^z \frac{1}{\alpha_k(y)}\frac{1}{|b|}K(y-V_s^{k,i},b)\,dy.$$

Now, using that $\alpha_k(y)$ is bounded below yields the desired result. This completes the proof of Proposition 4.

Proof of Proposition 5. It suffices to show that

$$H_k^\star(z) = \frac{1}{n_k}\sum_{i=1}^{n_k}\int_0^t\int_{\Re} H_{k,i}\big(s,w,z,\alpha_k(\cdot),m_k(\cdot)\big)P^{k,i}(ds\times dw)$$

and $\widehat{H_k(z)}$ are asymptotically equivalent. It follows from Proposition 2 since $\widehat{\alpha_k}(\cdot)$ and $\widehat{m_k}(\cdot)$ are uniformly consistent, and α_k is bounded away from zero.

In the i.i.d. case when $G_s^{k,i}(dv) = f_s^{k,i}(v)\,dl_d$,

$$\begin{aligned} E(\alpha_k^\star(y)) \;&=\; E\left(\int_0^t \frac{1}{|b|}K(y-V_s^{k,i},b)\lambda_s^{k,i}\,ds\right) \\ &\to E\left(\int_0^t f_s^{k,i}(y)\lambda_s^{k,i}\,ds\right) = \alpha_k(y), \end{aligned}$$

and it follows that

$$\begin{aligned} E(H_k^\star(z)) \;&=\; E\left(\int_0^t\int_{a-b}^{z+b}\sigma_k^2(v)\left[\int_a^z \frac{1}{\alpha_k(y)}\frac{1}{|b|}K(y-v,b)\,dy\right]^2 f_s^{k,i}(v)\,dv\,\lambda_s^{k,i}\,ds\right) \\ &\to \int_a^z \frac{\sigma_k^2(y)}{\alpha_k(y)}\,dy. \end{aligned}$$

When $G_s^{k,i}(dv) = \varepsilon_s \otimes f_s^{k,i}(v_1, \dots, v_{d-1})\, dl_{d-1}$,

$$E(\alpha_k^\star(y)) \to E\big(f_s^{k,i}(y)\lambda_s^{k,i}\big) = \alpha_k(y),$$

and the result follows similarly. □

Acknowledgments. We thank the Department of Growth and Reproduction at the University Hospital in Copenhagen for letting us use their data. We also would like to thank two referees for their helpful comments.

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.

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

BICKEL, P. J. and WICHURA, M. J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42** 1656–1670.

BOEL, R., VARAIYA, P. and WONG, E. (1975a). Martingales on jump processes. I: Representation results. *Siam J. Control* **13** 999–1022.

BOEL, R., VARAIYA, P. and WONG, E. (1975b). Martingales on jump processes. II: Applications. *Siam J. Control* **13** 1023–1061.

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

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

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.

FUSARO, R. E., NIELSEN, J. P. and SCHEIKE, T. H. (1993). Marker-dependent hazard estimation: An application to AIDS. *Statistics in Medicine* **12** 843–865.

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

JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes*. Springer, New York.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

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

MCKEAGUE, I. W. and UTIKAL, K. J. (1990a). Inference for a nonlinear counting process regression model. *Ann. Statist.* **18** 1172–1187.

MCKEAGUE, I. W. and UTIKAL, K. J. (1990b). Identifying nonlinear covariate effects in semimartingale regression models. *Probab. Theory Related Fields* **87** 1–25.

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. (1995). A central limit theorem for local Martingales with applications to the analysis of longitudinal data. *Scand. J. Statist.* **22** 279–294.

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

NIELSEN, J. P. and LINTON, O. B. (1995). Kernel estimation in a nonparametric marker dependent hazard model. *Ann. Statist.* **23** 1735–1749.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebeite* **51** 269–286.

SCHEIKE, T. H. (1993). Statistical analysis of tessellations and nonparametric kernel regression with biological applications. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

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

SCHEIKE, T. H. (1996). Nonparametric kernel regression when the regressor follows a counting process. *J. Nonparametr. Statist.* **6** 337–353.

WATSON, G. (1964). Smooth regression analysis. *Sankhyā Ser. A* **26** 359–372.

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

DIVISION OF BIOSTATISTICS
MEDICAL COLLEGE OF WISCONSIN
8701 WATERTOWN PLANK ROAD
MILWAUKEE, WISCONSIN 53226
E-MAIL: meijie@biostat.mcw.edu