UNIVERSITÄT HAMBURG

Change-point tests for the error distribution in nonparametric regression

Natalie Neumeyer & Ingrid Van Keilegom

Preprint–No. 2006-06 August 2006

FACHBEREICH MATHEMATIK

SCHWERPUNKT MATHEMATISCHE STATISTIK
UND STOCHASTISCHE PROZESSE

BUNDESSTR. 55, D – 20146 HAMBURG

Change-point tests for the error distribution in nonparametric regression

Natalie Neumeyer
Department Mathematik
Schwerpunkt Stochastik
Universität Hamburg
Bundesstrasse 55
20146 Hamburg, Germany
e-mail: neumeyer@math.uni-hamburg.de

Ingrid Van Keilegom
Institute of Statistics
Université catholique de Louvain
Voie du Roman Pays, 20
1348 Louvain-la-Neuve, Belgium
e-mail: vankeilegom@stat.ucl.ac.be

August 30, 2006

Abstract

In this paper several testing procedures are proposed that can detect change-points in the error distribution of nonparametric regression models. Different settings are considered where the change-point either occurs at some time point or at some value of the covariate. Fixed as well as random covariates are considered. Weak convergence of the suggested difference of sequential empirical processes based on nonparametrically estimated residuals to a Gaussian process is proved under the null hypothesis of no change point. In the case of testing for a change in the error distribution that occurs with increasing time in a model with random covariates the test statistic is asymptotically distribution-free and the asymptotic quantiles can be used for the test. This special test statistic can also detect a change in the regression function. In all other cases the asymptotic distribution depends on unknown features of the data generating process and a bootstrap procedure is proposed in these cases. The small sample performances of the proposed tests are investigated by means of a simulation study, and the tests are applied to a data example.

AMS Classification: 62G08, 62G10
Keywords and Phrases: bootstrap, change-point, nonparametric regression, residuals
Running title: Change-point tests for regression error

1 Introduction

In classical change-point problems one obverves a sample $\varepsilon_1, \dots, \varepsilon_n$ of random variables and one is interested in testing the null hypothesis that $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed (i. i. d.) versus the alternative that there exists a change-point $\theta \in (0,1)$, such that $\varepsilon_1, \dots, \varepsilon_{\lfloor n\theta \rfloor}$ are i. i. d. with some distribution function F and $\varepsilon_{\lfloor n\theta \rfloor+1}, \dots, \varepsilon_n$ are i. i. d. with some distribution function $G \neq F$. Test statistics are often based on the difference of the sequential empirical processes built from the first $\lfloor ns \rfloor$ and the last $n - \lfloor ns \rfloor$ observations ($0 \leq s \leq 1$), respectively. An asymptotically distribution-free test statistic is, for instance,

$$\sup_{s \in [0,1], y \in \mathbb{R}} \Big| \sqrt{n} \frac{\lfloor ns \rfloor}{n} (1 - \frac{\lfloor ns \rfloor}{n}) \Big(\frac{1}{\lfloor ns \rfloor} \sum_{i=1}^{\lfloor ns \rfloor} I\{\varepsilon_i \leq y\} - \frac{1}{n - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor+1}^{n} I\{\varepsilon_i \leq y\} \Big) \Big|,$$

see Shorack and Wellner (1986, p. 131). Picard (1985) considers this test statistic in a time series context whereas Carlstein (1988) proposes an estimator for the change-point θ. A procedure based on sequential quantile processes was suggested by Csörgö and Szyszkowicz (2000). Other change-point tests as well as estimators for the change-point were considered by Ritov (1990), Dümbgen (1991), Ferger and Stute (1992), and Ferger (1994, 1996), among others. A lot of attention has also been paid to the analogous testing problem where the random variables $\varepsilon_1, \dots, \varepsilon_n$ are not observed directly but are error variables in parametric time series models. In a time series setting with independent errors $\varepsilon_1, \dots, \varepsilon_n$, sequential empirical processes based on estimated residuals $\hat{\varepsilon}_i$ and corresponding change-point tests have been considered by Bai (1994) in the context of ARMA-models, by Koul (1996) in the context of nonlinear time series and by Ling (1998) for nonstationary autoregressive models. Apart from these papers, a vast literature on change-point tests in the time series context is available, see Giraitis, Leipus and Surgailis (1996), Inoue (2001), Horvath, Kokoszka and Teyssiere (2001), Boldin (2002), Lee and Na (2004), among others.

In this paper we consider a nonparametric regression model of the form $Y_i = m(X_i) + \varepsilon_i$ with independent observations, and construct change-point tests for the error distribution. These tests are able to detect change points in the error variance, but allow also to detect any other changes in the error distribution, like e.g. changes in skewness or kurtosis. Tests for changes in the error variance have been studied by Oyet and Sutradhar (2003) and Huh and Kang (2005), among others. The former paper tests whether the variance changes over time, whereas the latter paper considers the problem of testing whether the error variance jumps at some value of the covariate.

The tests proposed in this paper have many practical applications. Changes over time are of interest e.g. when analyzing stock market returns (where it is often assumed that the rates of return are independent), or when the measuring instrument has been cleaned or adjusted in the period during which the measurements have been taken in chronological order. Also, testing for changes in the error distribution in relation to the level of the covariate may be important when it is suspected that the accuracy of the measurement instrument differs for higher levels of the covariate (e. g., increasing temperature, humidity or dose of some substance), or in experimental toxicology studies on humans or animals, where it is often believed that there is a certain threshold from which on the toxic substance starts reacting.

The test statistics we are going to consider are based on sequential empirical processes of nonparametrically estimated residuals, $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, i. e.

$$\frac{1}{\lfloor ns \rfloor} \sum_{i=1}^{\lfloor ns \rfloor} I\{\hat{\varepsilon}_i \leq y\}.$$

The corresponding residual based empirical distribution function ($s = 1$) was considered by Akritas and Van Keilegom (2001) and Cheng (2002). Müller, Schick and Wefelmeyer (2004) suggested a smooth version of this empirical distribution function. Einmahl and Van Keilegom (2006) constructed tests for the independence of ε_i and X_i in the model $Y_i = m(X_i) + \varepsilon_i$. The tests for changes in the covariate developed in this paper can be viewed as a special case of their test, in the sense that the dependence of ε_i on X_i is assumed to be given by a change-point model.

The test statistic we propose for testing for a change in time in a setting with random covariates is asymptotically distribution-free, and, moreover, can detect changes in the regression function as well. Tests or estimators for change-points in nonparametric regression functions or their derivatives were proposed by Müller (1992), Loader (1996), Antoniadis and Gijbels (2002), Grégoire and Hamrouni (2002), Horvath and Kokoszka (2002) and Gijbels and Goderniaux (2004), among others.

The paper is organized as follows. In section 2 we consider the regression model under the assumption of random covariates and propose test statistics for testing for a change in the error distribution, where the change either occurs at some time point or at some value of the covariate. In section 3 analogous results are presented for a model with fixed covariates. Section 4 and 5 contain simulations and a real data example. All proofs are deferred to an appendix.

2 The random design case

2.1 Model and assumptions

In this section we consider a nonparametric regression model with independent observations,

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where ε_i is independent of X_i, and all ε_i are identically distributed with $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2 \in (0, \infty)$ (the index i represents the order in which the observations have been collected). We assume the univariate covariates $X_1, \ldots, X_n$ to be independent and identically distributed with distribution function F_X on compact support, say $[0, 1]$. Further we assume that F_X has a twice continuously differentiable density f_X such that $\sup_{x \in [0,1]} f_X(x) > 0$. To estimate the error distribution we build nonparametric residuals $\hat{\varepsilon}_i = Y_i - \hat{m}(X_i)$, $i = 1, \ldots, n$, where $\hat{m}$ denotes the Nadaraya–Watson estimator (Nadaraya, 1964, Watson, 1964), that is

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\Big(\frac{X_i - x}{h}\Big) Y_i \frac{1}{\hat{f}_X(x)} \tag{2.2}$$

and $\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K(\frac{X_i - x}{h})$ denotes the kernel density estimator of the covariates. Here we assume the regression function m to be twice continuously differentiable. K denotes a symmetric, twice continuously differentiable density with compact support, say $[-1, 1]$, and $\int uK(u)\,du = 0$, and $h = h_n$ denotes a sequence of bandwidths such that $nh^4 = o(1)$, $nh^{3+2\delta}(\log(h^{-1}))^{-1} \to \infty$ for $n \to \infty$ (for some $\delta > 0$). Finally, we define the following class of distribution functions,

$$\begin{aligned} \mathcal{D} = \Big\{ F : \mathbb{R} \to [0,1] \,\Big|\, & F \text{ is a twice continuously differentiable distribution function with} \\ & \text{density } f \text{ such that } \sup_{y\in\mathbb{R}} f(y) < \infty,\ \sup_{y\in\mathbb{R}} |f'(y)| < \infty \text{ and } \inf_{y\in\mathbb{R}} f(y) > 0 \Big\}. \end{aligned} \tag{2.3}$$

2.2 Testing for a change in time

We want to test the null hypothesis of independent and identically distributed errors versus the alternative

$$\exists \theta_0 \in (0,1) \text{ such that } \varepsilon_1, \ldots, \varepsilon_{\lfloor n\theta_0 \rfloor} \text{ i.i.d.} \sim F,\ \varepsilon_{\lfloor n\theta_0 \rfloor + 1}, \ldots, \varepsilon_n \text{ i.i.d.} \sim G \neq F. \tag{2.4}$$

For this aim we define a Kolomogorov-Smirnov type test statistic, $\sup\limits_{s\in[0,1]} \sup\limits_{y\in\mathbb{R}} |\hat{T}_{n1}(s,y)|$, where

$$\hat{T}_{n1}(s,y) = \sqrt{n} \frac{\lfloor ns \rfloor}{n}\Big(1 - \frac{\lfloor ns \rfloor}{n}\Big) \Big(\hat{F}_{\lfloor ns \rfloor}(y) - \hat{F}^*_{n-\lfloor ns \rfloor}(y) \Big)$$

is based on the sequential empirical processes of the first $\lfloor ns \rfloor$ and last $n - \lfloor ns \rfloor$ nonparametric residuals,

$$\begin{aligned} \hat{F}_{\lfloor ns \rfloor}(y) &= \frac{1}{\lfloor ns \rfloor} \sum_{i=1}^{\lfloor ns \rfloor} I\{\hat{\varepsilon}_i \le y\} \\ \hat{F}^*_{n-\lfloor ns \rfloor}(y) &= \frac{1}{n - \lfloor ns \rfloor} \sum_{i=\lfloor ns \rfloor + 1}^{n} I\{\hat{\varepsilon}_i \le y\}. \end{aligned}$$

Let $T_{n1}(s,y)$, $F_{\lfloor ns \rfloor}(y)$ and $F^*_{n-\lfloor ns \rfloor}(y)$ be defined analogously, but replacing the residuals $\hat{\varepsilon}_i$ by true errors ε_i. Under the null hypothesis we have the following asymptotic representation and weak convergence result.

Proposition 2.1 *Assume that model (2.1) is valid under the assumptions stated in section 2.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then,*

$$\frac{\lfloor ns \rfloor}{n}\left(\hat{F}_{\lfloor ns \rfloor}(y) - F(y)\right) \;=\; \frac{1}{n}\sum_{i=1}^{\lfloor ns \rfloor}(I\{\varepsilon_i \le y\} - F(y)) + \frac{\lfloor ns \rfloor}{n} f(y)\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i + o_p(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in \mathbb{R}$ and $s \in [0,1]$.

Theorem 2.2 *Assume that model (2.1) is valid under the assumptions stated in section 2.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then it holds that*

$$\sup_{s\in[0,1]}\sup_{y\in\mathbb{R}}|\hat{T}_{n1}(s,y) - T_{n1}(s,y)| \;=\; o_p(1)$$

and therefore $\hat{T}_{n1}(s,y)$ converges weakly to a mean zero Gaussian process $G_1(s,F(y))$ with covariance $\mathrm{Cov}(G_1(s,x), G_1(t,z)) = (s \wedge t - st)(x \wedge z - xz)$.

The limit of our test statistic is distribution-free,

$$\sup_{s\in[0,1]}\sup_{y\in\mathbb{R}}|\hat{T}_{n1}(s,y)| = \sup_{s\in[0,1]}\sup_{z\in[0,1]}|\hat{T}_{n1}(s,F^{-1}(z))| \xrightarrow{d} \sup_{s\in[0,1]}\sup_{z\in[0,1]}|G_1(s,z)|,$$

provided $f(y) > 0$ for all y, and the asymptotic quantiles tabled by Picard (1985) can be used for the test. This is an extraordinary case in the context of test statistics based on the empirical distribution function of nonparametric residuals as usual the asymptotic distribution of such tests depends heavily on the unknown error distribution F as well as on the density f, see for instance Pardo-Fernández, Van Keilegom and González-Manteiga (2006) or Neumeyer and Dette (2006) and the asymptotic results in the following sections 2.3, 3.2 and 3.3.

Remark 2.3 The considered test statistic $\sup_{s,y}|\hat{T}_{n1}(s,y)|$ can additionally detect changes that occur in the regression function m at some time point i. We consider to this end model (2.1) under the assumptions of section 2.1 with identically distributed errors, but under the alternative we assume the existence of some $\theta \in (0,1)$ such that for the regression function the representation

$$m(X_i) \;=\; m_1(X_i)I\{i \le \lfloor n\theta \rfloor\} + m_2(X_i)I\{i > \lfloor n\theta \rfloor\} \tag{2.5}$$

is valid, where $m_1 \neq m_2$, but m_1 and m_2 are both continuous. Under this alternative the kernel estimator

$$\hat{m}(x) = \frac{1}{n}\sum_{i=1}^{\lfloor n\theta \rfloor}\frac{1}{h}K(\frac{X_i - x}{h})(m_1(X_i) + \varepsilon_i)\frac{1}{\hat{f}_X(x)} + \frac{1}{n}\sum_{i=\lfloor n\theta \rfloor+1}^{n}\frac{1}{h}K(\frac{X_i - x}{h})(m_2(X_i) + \varepsilon_i)\frac{1}{\hat{f}_X(x)}$$

converges in probability to $\theta m_1(x)+(1-\theta)m_2(x)$. Hence, $\hat{F}_{\lfloor n\theta\rfloor}(y)$ estimates $F_1(y) = \int F(y+(1-\theta)\{m_2(x)-m_1(x)\})dF_X(x)$ and $\hat{F}^*_{n-\lfloor n\theta\rfloor}(y)$ estimates $F_2(y) = \int F(y+\theta\{m_1(x)-m_2(x)\})dF_X(x)$ in a consistent way. To show the consistency of the test against the fixed alternative (2.5), let us assume that $F_1 \equiv F_2$. Then, under our assumptions and by an application of Theorem 1 in Pardo-Fernández, Van Keilegom and González-Manteiga (2006), it follows that the functions $(1-\theta)(m_2 - m_1)$ and $\theta(m_1 - m_2)$ are equal, and hence m_1 equals m_2. This shows that $F_1 \equiv F_2$ if and only if $m_1 \equiv m_2$, from which the consistency can be deduced.

Remark 2.4 The result of Theorem 2.2 remains valid in models without covariates. Here residuals are defined as $\hat{\varepsilon}_i = Y_i - \frac{1}{n}\sum_{j=1}^n Y_j$ and all other notations are as before. The aim here is to test for a change in the distribution of independent random variables with the same constant mean. An important application of this type of tests comes from the analysis of stock-market prices, where it is often assumed that the rates of return (i.e. $R_i = (P_{i+1}-P_i)/P_i$, where P_i is a certain index value at time i) are independent. It is then of interest to test whether, after normalization to zero mean, the distribution of the R_i's shows any change in time, caused by e.g. a major economical or political event. A nice example of this type of data is analyzed in Hsu (1979), among others.

Remark 2.5 The above results can be used to establish the consistency of an estimator of the change point θ_0. First, note that

$$\theta_0 = \arg\max_{s\in[0,1]}\Big(\sup_{y\in\mathbb{R}} |A(s,y)|\Big),$$

where

$$\begin{aligned} A(s,y) &= \{s(1-s)\}^{1/2}\Big[\frac{1}{s}\{(s\wedge\theta_0)F(y) + (s-\theta_0)_+G(y)\} \\ &\quad -\frac{1}{1-s}\{(\theta_0 - s)_+F(y) + (1-s\vee\theta_0)G(y)\}\Big] \\ &= \Big\{\frac{\lfloor ns\rfloor}{n}(1-\frac{\lfloor ns\rfloor}{n})\Big\}^{1/2}\Big\{\frac{1}{\lfloor ns\rfloor}\sum_{i=1}^{\lfloor ns\rfloor} P(\varepsilon_i \le y) - \frac{1}{n-\lfloor ns\rfloor}\sum_{i=\lfloor ns\rfloor+1}^{n} P(\varepsilon_i\le y)\Big\} + o_p(1). \end{aligned}$$

Hence, a natural estimator of θ_0 is given by

$$\hat{\theta}_n = \arg\max_{s\in J_n}\Big(\sup_{y\in\mathbb{R}} |\hat{A}_n(s,y)|\Big),$$

where $\hat{A}_n(s,y) = \{\frac{\lfloor ns\rfloor}{n}(1-\frac{\lfloor ns\rfloor}{n})\}^{1/2}\{\hat{F}_{\lfloor ns\rfloor}(y) - \hat{F}^*_{n-\lfloor ns\rfloor}(y)\}$ and $J_n = \{k/n : 1 \le k \le n-1\}$. It can be easily shown that

$$\begin{aligned} &\sup_s \Big| \sup_y |\hat{A}_n(s,y)| - \sup_y |A(s,y)| \Big| \\ &\le \sup_{s,y} |\hat{A}_n(s,y) - A(s,y)| = o_p(1), \end{aligned}$$

where the convergence in probability follows from Proposition 2.1. It can now be proved, using Theorem 5.7 in van der Vaart (1998, page 45), that $\hat{\theta}_n - \theta_0 = o_p(1)$. Similar remarks apply to the results in Theorems 2.7, 3.2 and 3.4.

2.3 Testing for a change in covariate

In this section we want to test whether there is a change in the error distribution connected with or caused by increasing covariate values, i.e. we want to test the null hypothesis of independence of ε_i and X_i for all i, versus the alternative that there exists a $0 < x_0 < 1$ such that

$$\varepsilon | X = x \sim F \text{ for all } x \leq x_0, \text{ and } \varepsilon | X = x \sim G \neq F \text{ for all } x > x_0. \tag{2.6}$$

Define the test statistic now as $\sup_{s\in[0,1]} \sup_{y\in\mathbb{R}} |\hat{T}_{n2}(s,y)|$, where

$$\hat{T}_{n2}(s,y) = \sqrt{n}\hat{F}_X(s)(1-\hat{F}_X(s)) \left(\frac{1}{\hat{F}_X(s)} \hat{F}_n(s,y) - \frac{1}{1-\hat{F}_X(s)} \hat{F}_n^*(s,y) \right)$$

is based on the modified sequential residual based empirical processes,

$$\hat{F}_n(s,y) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq s\} I\{\hat{\varepsilon}_i \leq y\} \tag{2.7}$$

$$\hat{F}_n^*(s,y) = \frac{1}{n} \sum_{i=1}^n I\{X_i > s\} I\{\hat{\varepsilon}_i \leq y\}, \tag{2.8}$$

and where $\hat{F}_X$ denotes the empirical distribution function of the covariates, i. e.

$$\hat{F}_X(s) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq s\}.$$

In this setting the test statistic is no longer asymptotically distribution-free. We have the following asymptotic results.

Proposition 2.6 *Assume that model (2.1) is valid under the assumptions stated in section 2.1 with errors $\varepsilon_1, \dots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then,*

$$\hat{F}_n(s,y) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq s\} I\{\varepsilon_i \leq y\} + f(y) \frac{1}{n} \sum_{i=1}^n \varepsilon_i \int_0^s \frac{1}{h} K(\frac{X_i - x}{h})\, dx + o_p(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in \mathbb{R}$ and $s \in [0,1]$.

Theorem 2.7 *Assume that model (2.1) is valid under the assumptions stated in section 2.1 with errors $\varepsilon_1, \dots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then it holds that*

$$\begin{aligned} \hat{T}_{n2}(s,y) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(& [I\{\varepsilon_i \leq y\} - F(y)][I\{X_i \leq s\} - F_X(s)] \\ & + f(y)\varepsilon_i \Big[\int_0^s \frac{1}{h} K(\frac{X_i - x}{h})\, dx - F_X(s) \Big] \Big) + o_p(1) \end{aligned}$$

uniformly with respect to $y \in \mathbb{R}$, $s \in [0,1]$. $\hat{T}_{n2}$ *converges weakly to a mean zero Gaussian process* G_2 *with covariance*

$$\begin{aligned}\mathrm{Cov}(G_2(s,y), G_2(t,z)) &= (F_X(s) \wedge F_X(t) - F_X(s)F_X(t))\Big(F(y \wedge z) - F(y)F(z) + \sigma^2 f(y) f(z) \\ &\quad + f(y)E[\varepsilon_1 I\{\varepsilon_1 \le z\}] + f(z)E[\varepsilon_1 I\{\varepsilon_1 \le y\}]\Big).\end{aligned}$$

The asymptotic covariances of the process $\hat{T}_{n2}$ depend on the distributions of covariates and errors, but the asymptotic distribution of the test statistic,

$$\sup_{y \in \mathbb{R}} \sup_{s \in [0,1]} |\hat{T}_{n2}(s,y)| = \sup_{y \in \mathbb{R}} \sup_{t \in [0,1]} |\hat{T}_{n2}(F_X^{-1}(t), y)|,$$

depends only on the error distribution.

Remark 2.8 When the distribution F_X would be known, we do not need to estimate it in the formula of $\hat{T}_{n2}(s,y)$. As a consequence the term $-F(y)F(z)$ in the asymptotic covariance of the process $G_2(s,y)$ vanishes.

Remark 2.9 The following smooth residual bootstrap can be applied to approximate the critical values of the test. Based on the original sample $(X_1, Y_1), \dots, (X_n, Y_n)$ build bootstrap observations

$$Y_i^B = \hat{m}(X_i) + \varepsilon_i^B + aZ_i, \quad i = 1, \dots, n,$$

where ε_i^B is drawn randomly with replacement from centered residuals $\hat{\varepsilon}_k - \frac{1}{n}\sum_{j=1}^n \hat{\varepsilon}_j$, $k = 1, \dots, n$, a is a small smoothing parameter and $Z_1, \dots, Z_n$ are independent centered random variables, independent from the original sample. Denote by $\hat{F}_n^B$, $\hat{F}_n^{*B}$ and $\hat{T}_{n2}^B$ the analogues of $\hat{F}_n$, $\hat{F}_n^*$ and $\hat{T}_{n2}$, respectively, calculated from the bootstrap sample $(X_1, Y_1^B), \dots, (X_n, Y_n^B)$. Neumeyer (2006) shows that under regularity conditions under the null hypothesis as well as under fixed alternatives the process $\sqrt{n}(\hat{F}_n^B(1,y) - \hat{F}_n(1,y))$, $y \in \mathbb{R}$, converges conditionally on the original sample in probability to the limit distribution (under the null hypothesis) of the process $\sqrt{n}(\hat{F}_n(1,y) - F(y))$, $y \in \mathbb{R}$. Combining the methods in Neumeyer's (2006) proofs with the proofs given in the current work it can be shown that a similar result holds for the sequential empirical processes. Hence, the critical values of the test statistic $\sup_{s,y} |\hat{T}_{n2}(s,y)|$ can be approximated by the quantiles of the conditional distribution of $\sup_{s,y} |\hat{T}_{n2}^B(s,y)|$ given the original sample. These quantiles are estimated by resampling b times from the original sample for some large b and calculating the corresponding order statistics. In a similar way it can be shown that the critical values of the test statistic $\sup_{s,y} |\hat{T}_{n1}(s,y)|$ can be approximated by the above smooth residual bootstrap.

3 The fixed design case

3.1 Model and assumptions

In this section we consider a nonparametric regression model with independent observations

$$Y_i = m(x_{ni}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where the distribution of ε_i is the same for all i, with $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2 \in (0, \infty)$. The covariates $x_{n1}, \ldots, x_{nn}$ are not random. We assume the existence of a strictly increasing known distribution function F_X with support $[0,1]$ and density f_X, such that $F_X(x_{ni}) = g(\frac{i}{n})$, $i = 1, \ldots, n$, for some one-to-one known permutation function g from $[0,1]$ onto $[0,1]$ that is piecewise linear with slopes 1 or -1 and right-continuous. The function g allows the experimenter to select designs others than the one that starts with the smallest value at the first time point and which increases the value of the covariate at each time point. Indeed, in order to distinguish the effect of time from the effect of the covariate, it is better not to take this simple design, but to use instead a non-monotone design, like e.g. the one corresponding to the function $g(s) = sI(s \le 0.25) + (s + 0.25)I(0.25 < s \le 0.5) + (s + 0.25)I(0.5 < s \le 0.75) + (s - 0.5)I(0.75 < s)$, which we will use in the simulations.
Assumptions on the errors, the design density f_X, the regression function m, also assumptions on K and h are as in section 2.1. The estimators $\hat{f}_X$ (now deterministic), $\hat{m}$ and $\hat{\varepsilon}_i$ are as before, but replacing the random design points X_i by the fixed x_{ni}.

3.2 Testing for a change in time

The definitions of the null and alternative hypotheses, and of the sequential empirical processes $\hat{F}_{\lfloor ns \rfloor}$, $\hat{F}^*_{n-\lfloor ns \rfloor}$, $F_{\lfloor ns \rfloor}$ and $F^*_{n-\lfloor ns \rfloor}$ are the same as in section 2.2. The test statistic is given by $\hat{T}_{n3}(s, y) = \hat{T}_{n1}(s, y)$. In contrast to the random design case, the asymptotic distribution of the test statistic will in this case not be distribution-free, but depend on the error distribution.

Proposition 3.1 *Assume that model (3.1) is valid under the assumptions stated in section 3.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then,*

$$\frac{\lfloor ns \rfloor}{n}\Big(\hat{F}_{\lfloor ns \rfloor}(y) - F(y)\Big) = \frac{\lfloor ns \rfloor}{n}\big(F_{\lfloor ns \rfloor}(y) - F(y)\big) + f(y)\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i \int_{S(s)} \frac{1}{h} K(\frac{x_{ni} - x}{h})\, dx + o_p(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in \mathbb{R}$ and $s \in [0,1]$, where $S(s) = \{F_X^{-1}(g(t)) : 0 \le t \le s\}$.

Theorem 3.2 *Assume that model (3.1) is valid under the assumptions stated in section 3.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then it holds that*

$$\begin{aligned}\hat{T}_{n3}(s,y) \;=\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(I\{\varepsilon_i \le y\}\Big[I\{\frac{i}{n} \le s\} - \frac{\lfloor ns \rfloor}{n}\Big] \\ & + f(y)\varepsilon_i\Big[\int_{S(s)}\frac{1}{h}K(\frac{x_{ni}-x}{h})dx - \frac{\lfloor ns \rfloor}{n}\Big]\Big) + o_p(1)\end{aligned}$$

uniformly with respect to $y \in \mathbb{R}$, $s \in [0,1]$. $\hat{T}_{n3}$ converges weakly to a mean zero Gaussian process G_3 with covariance

$$\begin{aligned}\operatorname{Cov}(G_3(s,y), G_3(t,z)) \;=\; & [s \wedge t - st]\Big(F(y \wedge z) + \sigma^2 f(y)f(z) + f(y)E[\varepsilon_1 I\{\varepsilon_1 \le z\}] \\ & + f(z)E[\varepsilon_1 I\{\varepsilon_1 \le y\}]\Big).\end{aligned}$$

3.3 Testing for a change in covariate

The null hypothesis of i. i. d. errors $\varepsilon_1, \ldots, \varepsilon_n$ will now be tested against the alternative that there exists a $0 < x_0 < 1$ such that

$$\varepsilon_i \sim F \text{ for all } x_{ni} \le x_0, \text{ and } \varepsilon_i \sim G \neq F \text{ for all } x_{ni} > x_0. \tag{3.2}$$

We can use a process similar to $\hat{T}_{n2}$ defined in section 2.3 but we do not have to estimate F_X. Therefore we define

$$\hat{T}_{n4}(s,y) \;=\; \sqrt{n}F_X(s)(1-F_X(s))\left(\frac{1}{F_X(s)}\hat{F}_n(s,y) - \frac{1}{1-F_X(s)}\hat{F}_n^*(s,y)\right),$$

where $\hat{F}_n$ and $\hat{F}_n^*$ are as in (2.7) and (2.8), respectively (replace X_i by x_{ni}).

The proofs of the following results are omitted, as they can be obtained in a somewhat similar way as those of Propositions 2.6 and 3.1 and Theorems 2.7 and 3.2.

Proposition 3.3 *Assume that model (3.1) is valid under the assumptions stated in section 3.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then,*

$$\hat{F}_n(s,y) \;=\; \frac{1}{n}\sum_{i=1}^{n} I\{x_{ni} \le s\}I\{\varepsilon_i \le y\} + f(y)\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\int_0^s \frac{1}{h}K(\frac{x_{ni}-x}{h})\,dx + o_p(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in \mathbb{R}$ and $s \in [0,1]$.

Theorem 3.4 *Assume that model (3.1) is valid under the assumptions stated in section 3.1 with errors $\varepsilon_1, \ldots, \varepsilon_n$ that are independent and identically distributed with distribution function $F \in \mathcal{D}$ defined in (2.3). Then it holds that*

$$\begin{aligned}\hat{T}_{n4}(s, y) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{\varepsilon_i \leq y\}[I\{x_{ni} \leq s\} - F_X(s)] \\ &\quad + f(y)\varepsilon_i \Big[\int_0^s \frac{1}{h} K(\frac{x_{ni} - x}{h})\, dx - F_X(s) \Big] \Big) + o_p(1)\end{aligned}$$

uniformly with respect to $y \in \mathbb{R}$, $s \in [0, 1]$. $\hat{T}_{n4}$ converges weakly to a mean zero Gaussian process G_4 with covariance

$$\begin{aligned}\mathrm{Cov}(G_4(s, y), G_4(t, z)) &= (F_X(s) \wedge F_X(t) - F_X(s)F_X(t)) \Big(F(y \wedge z) + \sigma^2 f(y) f(z) \\ &\quad + f(y)E[\varepsilon_1 I\{\varepsilon_1 \leq z\}] + f(z)E[\varepsilon_1 I\{\varepsilon_1 \leq y\}] \Big).\end{aligned}$$

Our test statistic is then

$$\sup_{y\in\mathbb{R}} \sup_{s\in[0,1]} |\hat{T}_{n4}(s, y)| = \sup_{y\in\mathbb{R}} \sup_{t\in[0,1]} |\hat{T}_{n4}(F_X^{-1}(t), y)| = \sup_{y\in\mathbb{R}} \sup_{t\in[0,1]} |\hat{T}_{n3}(t, y)|$$

and we have the same limit as in Theorem 3.2.

Remark 3.5 In a similar way as for the random design case, we can approximate the critical values of the test statistics $\hat{T}_{n3}$ and $\hat{T}_{n4}$ by using a smooth bootstrap procedure based on resampling the residuals. As for the random design case, the proof of the asymptotic validity of this bootstrap procedure can be done by generalizing results in Neumeyer (2006) for sequential empirical processes.

4 Simulations

Under the null hypothesis of no change-point, we consider the following model : when the design is random, let

$$Y_i = m(X_i) + \varepsilon_i \tag{4.1}$$

$(i = 1, \ldots, n)$, where X_i follows a uniform distribution on $[0, 1]$, $m(X_i) = X_i + 1$ and ε_i follows a zero mean normal distribution with variance 0.5^2. For the fixed design case, we replace X_i by $x_{ni} = g(i/n)$, where

$$\begin{aligned}g(s) &= sI(s \leq 0.25) + (s + 0.25)I(0.25 < s \leq 0.5) \\ &\quad + (s + 0.25)I(0.5 < s \leq 0.75) + (s - 0.5)I(0.75 < s).\end{aligned}$$

Next, we specify the distributions F and G, which together with the values of $x_0 = 0.5$ and $\theta_0 = 0.5$ completely determine the alternative hypotheses (2.4), (2.6) and (3.2) :

$$\begin{aligned}
&\text{Model 1}: F = N(0, 0.5^2), && G = N(0, (0.5 + \delta)^2), \\
&\text{Model 2}: F = N(0, 0.5^2), && G = \frac{1}{2}N(-2\delta, 0.5^2) + \frac{1}{2}N(2\delta, 0.5^2),
\end{aligned}$$

where $\delta > 0$. The simulations are carried out for samples of size 50, 100 and 200. A total of 1000 samples are selected at random, and for each sample $b = 1000$ random resamples are drawn (except for $n = 200$, for which we take 500 samples and $b = 500$ resamples). To estimate the regression function $m(\cdot)$ we use a biquadratic kernel function $K(u) = (15/16)(1-u^2)^2 I(|u| \leq 1)$ and we determine the bandwidth h by using a cross-validation procedure, i.e. we select the value of h that minimizes

$$\sum_{i=1}^{n} [Y_i - \hat{m}_{(i),h}(X_i)]^2,$$

where $\hat{m}_{(i),h}(\cdot)$ is the estimator defined in (2.2), but without using observation i. The critical values of our test statistics $\sup_{y,s} |\hat{T}_{nj}(s, y)|$ ($j = 1, 2, 3, 4$) are obtained my means of the smoothed bootstrap procedure, proposed in Remarks 2.9 and 3.5. The pilot bandwidth a used to smooth the bootstrap distribution equals 0.1 and the Z_i are standard normally distributed. For the test statistic $\sup_{y,s} |\hat{T}_{n1}(s, y)|$ we calculate in addition the critical values based on its asymptotic distribution.

The results are shown in Figures 1 and 2. The figures show that the level is well respected for all sample sizes, all four setups and both the asymptotic and the bootstrap procedure (take $\delta = 0$). The exact values of the rejection probabilities under the null hypothesis are given in Table 1. Under the alternative hypothesis, both model 1 and 2 demonstrate that the power increases with increasing sample size and increasing value of δ. Only small differences are observed between the powers under the four setups. Finally, for the test $\sup_{y,s} |\hat{T}_{n1}(s, y)|$ the power obtained from the asymptotic procedure seems to work equally well as the one obtained from the bootstrap method.

n	Random design			Fixed design	
	Time		Covar	Time	Covar.
	Asym.	Boot.			
50	.037	.059	.037	.053	.047
100	.049	.060	.043	.047	.058
200	.052	.058	.060	.048	.046

Table 1: *Rejection probabilities under the null hypothesis of no change-point.*

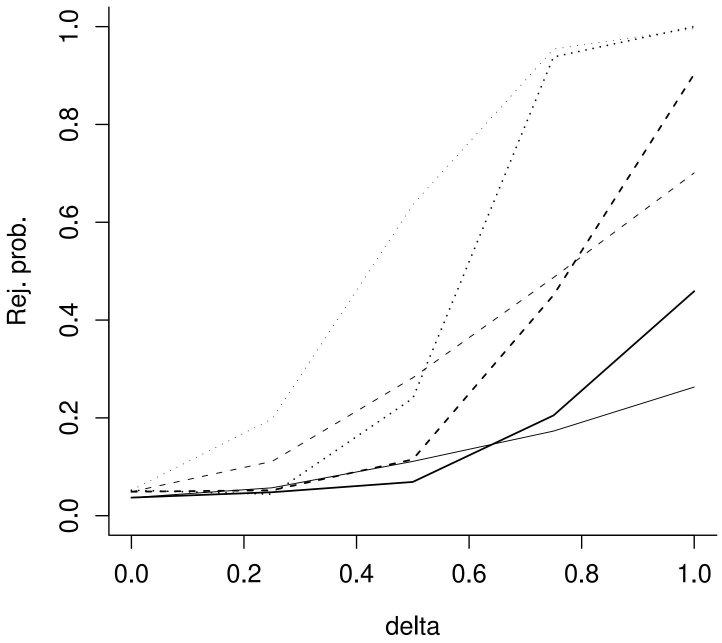

Figure 1: *Rejection probabilities obtained from the asymptotic distribution of the test statistic* $\sup_{y,s} |\hat{T}_{n1}(s, y)|$ *for* $n = 50$ *(solid curve),* $n = 100$ *(dashed curve) and* $n = 200$ *(dotted curve). The thin curves represent the results for model 1, the thick curves the results for model 2.*

5 Real data example

We now analyze data concerning the so-called LIDAR technique (which stands for LIght Detection And Ranging), that uses the reflection of laser-emitted light to detect chemical compounds in the atmosphere. We consider as response the logarithm of the ratio of received light from two laser sources, whereas the covariate is the distance traveled before the light is reflected back to its source. A scatterplot of the data set, consisting of 221 observations, is shown in Figure 3, together with a kernel estimator of the regression function (with bandwidth $h = 50$ obtained from cross-validation). See Ruppert, Wand and Carroll (2003) for more details on this data set. As is mentioned in Huh and Kang (2005), the scatterplot suggests that a change-point in the error distribution might exist near $x = 600$. The application of the proposed testing procedure gives a p-value of 0 (obtained from 1000 bootstrap samples), which strongly suggests the presence of a change-point. An estimator of the change-point (see Remark 2.5 for details) is given by $x = 586.5$. The change-point splits the data set in two new data sets, consisting of those data points for which the covariate is either smaller or larger than the change-point. We apply the testing procedure on each of these new data sets, and find non-significant p-values, indicating that there is only one change-point in the data.

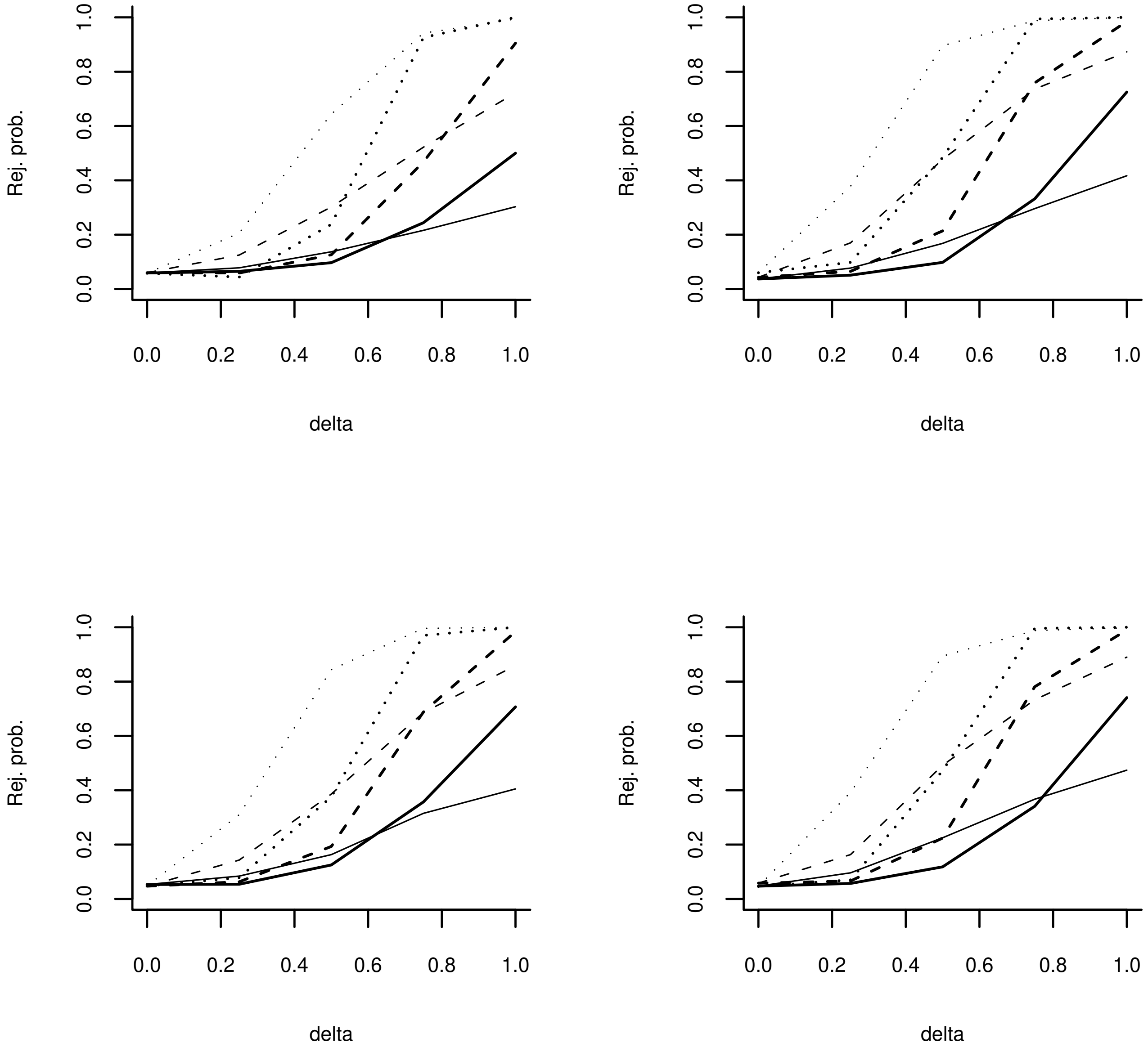

Figure 2: *Rejection probabilities obtained from the bootstrap distribution of the test statistics* $\sup_{y,s} |\hat{T}_{nj}(s,y)|$ *(*$j = 1, 2, 3, 4$*) for* $n = 50$ *(solid curve),* $n = 100$ *(dashed curve) and* $n = 200$ *(dotted curve). The thin curves represent the results for model 1, the thick curves the results for model 2. The random (fixed) design case is shown in the first (second) row; the results for a change in time (covariate) are in the left (right) column.*

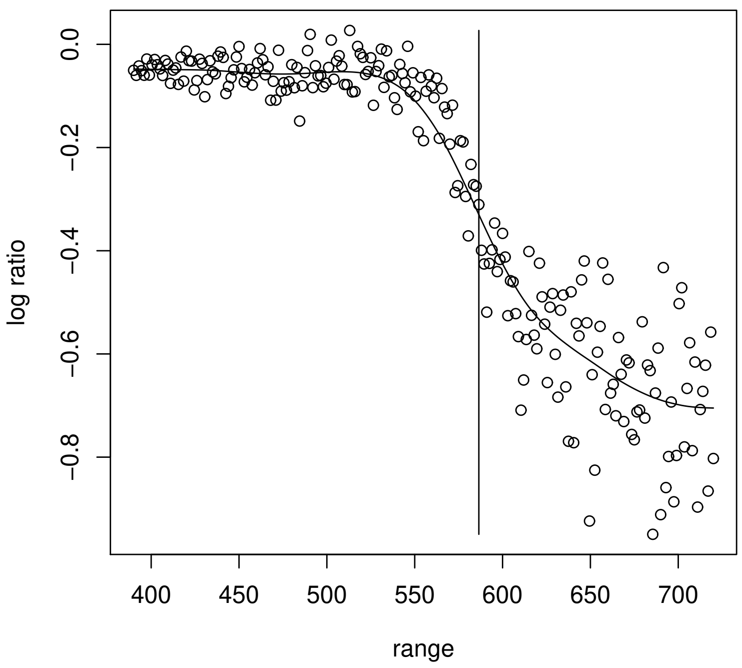

Figure 3: *Scatterplot of the LIDAR data, together with a kernel estimator of the regression function and the location of the change-point.*

A Appendix: Proofs

Proof of Proposition 2.1

Lemma A.1 *Under the assumptions of Proposition 2.1 we have*

$$\sup_{s\in[0,1]}\sup_{y\in\mathbb{R}}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(I\{\hat{\varepsilon}_i\leq y\}-I\{\varepsilon_i\leq y\}-\int_0^1 F(y+\hat{m}(x)-m(x))f_X(x)\,dx+F(y)\Big)I\{\frac{i}{n}\leq s\}\Big|$$
$$=o_p(\frac{1}{\sqrt{n}}).$$

Proof: The proof is similar to but less complicated than the proof of Lemma A.3 in the fixed design case and is therefore omitted. □

The assertion of Proposition 2.1 now follows by Taylor's expansion similar to the proof of Proposition 2.6. □

Proof of Theorem 2.2

Under the null hypothesis from Lemma A.1 and the analogous result for $\hat{F}^*_{n-\lfloor ns\rfloor}$ we obtain uniformly in $s \in [0,1]$ and $y \in \mathbb{R}$

$$\begin{aligned}
\hat{T}_{n1}(s,y) &= \sqrt{n}(1-\frac{\lfloor ns\rfloor}{n})\frac{\lfloor ns\rfloor}{n}(\hat{F}_{\lfloor ns\rfloor}(y)-F(y)) - \sqrt{n}\frac{\lfloor ns\rfloor}{n}(1-\frac{\lfloor ns\rfloor}{n})(\hat{F}^*_{n-\lfloor ns\rfloor}(y)-F(y)) \\
&= \sqrt{n}(1-\frac{\lfloor ns\rfloor}{n})\frac{\lfloor ns\rfloor}{n}\Big(F_{\lfloor ns\rfloor}(y)-F(y)+\int_0^1 (F(y+\hat{m}(x)-m(x))-F(y))f_X(x)\,dx\Big) \\
&\quad - \sqrt{n}\frac{\lfloor ns\rfloor}{n}(1-\frac{\lfloor ns\rfloor}{n})\Big(F^*_{n-\lfloor ns\rfloor}(y)-F(y)+\int_0^1 (F(y+\hat{m}(x)-m(x))-F(y))f_X(x)\,dx\Big) \\
&\quad + o_p(1) \\
&= T_{n1}(s,y)+o_p(1).
\end{aligned}$$

$\square$

Proof of Proposition 2.6

Lemma A.2 *Under the assumptions of Proposition 2.6 we have*

$$\begin{aligned}
\sup_{s\in[0,1]}\sup_{y\in\mathbb{R}}\Big|\frac{1}{n}\sum_{i=1}^n & I\{\hat{\varepsilon}_i \le y\}I\{X_i \le s\} - I\{\varepsilon_i \le y\}I\{X_i\le s\} \\
&-\int_0^s F(y+\hat{m}(x)-m(x))f_X(x)\,dx + F(y)F_X(s)\Big| \;=\; o_p(\frac{1}{\sqrt{n}}).
\end{aligned}$$

Proof: Note that $I\{\hat{\varepsilon}_i \le y\} = I\{\varepsilon_i \le y+\hat{m}(X_i)-m(X_i)\}$ and consider the empirical process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\Big(\varphi(X_i,\varepsilon_i)-E[\varphi(X_i,\varepsilon_i)]\Big) \quad (\varphi\in\mathcal{F})$$

indexed by the following class of functions,

$$\mathcal{F} \;=\; \Big\{(X,\varepsilon)\mapsto I\{\varepsilon\le y+d(X)\}I\{X\le s\}-I\{\varepsilon\le y\}I\{X\le s\}\;\Big|\; y\in\mathbb{R}, s\in[0,1], d\in C_1^{1+\delta}[0,1]\Big\}$$

(see van der Vaart and Wellner, 1996, p. 154, for the definition of $C_1^{1+\delta}[0,1]$). $\mathcal{F}$ is a product of the uniformly bounded Donsker classes $\{(X,\varepsilon)\mapsto I\{X\le s\}|s\in[0,1]\}$ and $\{(X,\varepsilon)\mapsto I\{\varepsilon \le y+d(X)\}-I\{\varepsilon\le y\}|y\in\mathbb{R}, d\in C_1^{1+\delta}[0,1]\}$ (see Akritas and Van Keilegom, 2001, Lemma B.1) and is therefore Donsker as well (Ex. 2.10.8, van der Vaart and Wellner, 1996, p. 192). The remaining part of the proof follows exactly the lines of the end of the proof of Lemma B.1, Akritas and Van Keilegom (2001, p. 567), using the observation that

$$\mathrm{Var}\Big(I\{\varepsilon\le y+d(X)\}I\{X\le s\}-I\{\varepsilon\le y\}I\{X\le s\}\Big) \;\le\; E\Big[\Big(I\{\varepsilon\le y+d(X)\}-I\{\varepsilon\le y\}\Big)^2\Big].$$

$\square$

From Lemma A.2 above we have the following expansion, uniformly with respect to $s \in [0,1]$ and $y \in \mathbb{R}$,

$$\begin{aligned}\hat{F}_n(s,y) &= F_n(s,y) + \int_0^s \left(F(y+\hat{m}(x)-m(x)) - F(y)\right) f_X(x)\,dx + o_p(\frac{1}{\sqrt{n}}) \\ &= F_n(s,y) + f(y)\int_0^s (\hat{m}(x)-m(x))f_X(x)\,dx + o_p(\frac{1}{\sqrt{n}})\end{aligned}$$

by Taylor's expansion, the assumption $\sup_y |f'(y)| < \infty$ and the fact that

$$\sup_{x\in[0,1]} (\hat{m}(x)-m(x))^2 = O_p(\frac{\log h^{-1}}{nh}) = o_p(n^{-1/2})$$

by standard results about kernel regression estimators and the bandwidth assumptions. Using the definition of $\hat{m}$ we further obtain

$$\hat{F}_n(s,y) = F_n(s,y) + f(y)\frac{1}{n}\sum_{i=1}^n \varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})\frac{f_X(x)}{\hat{f}_X(x)}\,dx \tag{A.1}$$

$$+ f(y)\frac{1}{n}\sum_{i=1}^n \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})(m(X_i)-m(x))\frac{f_X(x)}{\hat{f}_X(x)}\,dx + o_p(\frac{1}{\sqrt{n}}). \tag{A.2}$$

Next, we want to replace the random denominator $\hat{f}_X$ by the true density. To this end we write

$$\begin{aligned}&\sup_{s\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})\frac{f_X(x)}{\hat{f}_X(x)}\,dx - \frac{1}{n}\sum_{i=1}^n \varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})\,dx\Big| \\ &\le \int_0^1 |\hat{f}_X(x) - f_X(x)|\,dx \sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \frac{1}{h}K(\frac{X_i-x}{h})\frac{1}{\hat{f}_X(x)}\Big|.\end{aligned}$$

The second factor is of the type $\sup_x |\hat{g}(x) - g(x)|$ where $\hat{g}$ denotes the Nadaraya-Watson estimator in a special regression model $y_i = g(X_i)+\varepsilon_i$ with regression function $g \equiv 0$. Therefore for the whole product term we obtain the rate $O_p(\log(h^{-1})n^{-1}h^{-1}) = o_p(n^{-1/2})$. In a similar way we replace the random denominator in (A.2) by the true density. Next, we will prove

$$\sup_{s\in[0,1]}\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})(m(X_i)-m(x))\,dx\Big| = o_p(1). \tag{A.3}$$

One can show similar to but less complicated than the proof of Theorem 2.7 that the process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n(\varphi_{n,s}(X_i) - E[\varphi_{n,s}(X_i)])\ ,\ s\in[0,1],$$

converges weakly to a Gaussian process, where $\varphi_{n,s}(X_i) = \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})(m(X_i)-m(x))\,dx$. To prove (A.3) it remains to show that for fixed $s \in [0,1]$, $E[\varphi_{n,s}(X_i)] = o(n^{-1/2})$ and $\mathrm{Var}(\varphi_{n,s}(X_i)) = o(1)$. The details are omitted for the sake of brevity. We now obtain

$$\hat{F}_n(s,y) = F_n(s,y) + f(y)\frac{1}{n}\sum_{i=1}^n \varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i-x}{h})\,dx + o_p(\frac{1}{\sqrt{n}})$$

uniformly with respect to $s \in [0,1]$ and $y \in \mathbb{R}$. □

Proof of Theorem 2.7

Using the definition of $\hat{T}_{n2}$ and Proposition 2.6 we have the following expansion, uniformly with respect to $s \in [0,1]$, $y \in \mathbb{R}$:

$$\begin{aligned}
\hat{T}_{n2}(s,y) &= (1-\hat{F}_X(s))\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{X_i \le s\}I\{\varepsilon_i \le y\} + f(y)\varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i - x}{h})\,dx\Big) \\
&\quad - \hat{F}_X(s)\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{X_i > s\}I\{\varepsilon_i \le y\} + f(y)\varepsilon_i \int_s^1 \frac{1}{h}K(\frac{X_i - x}{h})\,dx\Big) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i \le y\}[(1-F_X(s))I\{X_i \le s\} - F_X(s)I\{X_i > s\}] \\
&\quad + f(y)\varepsilon_i\Big[(1-F_X(s))\int_0^s \frac{1}{h}K(\frac{X_i - x}{h})\,dx - F_X(s)\int_s^1 \frac{1}{h}K(\frac{X_i - x}{h})\,dx\Big]\Big) \\
&\quad + \sqrt{n}\Big(F_X(s) - \hat{F}_X(s)\Big)\Big(\frac{1}{n}\sum_{i=1}^n \Big[I\{X_i \le s\}I\{\varepsilon_i \le y\} + I\{X_i > s\}I\{\varepsilon_i \le y\} \\
&\quad + f(y)\varepsilon_i \int_0^s \frac{1}{h}K(\frac{X_i - x}{h})\,dx + f(y)\varepsilon_i \int_s^1 \frac{1}{h}K(\frac{X_i - x}{h})\,dx\Big]\Big) + o_p(1).
\end{aligned}$$

Inserting the definition of $\hat{F}_X(s)$ we further obtain

$$\begin{aligned}
\hat{T}_{n2}(s,y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\varepsilon_i \le y\}[I\{X_i \le s\} - F_X(s)] + f(y)\varepsilon_i\Big[\int_0^s \frac{1}{h}K(\frac{X_i - x}{h})\,dx - F_X(s)\Big] \\
&\quad - \Big(\frac{1}{n}\sum_{j=1}^n I\{\varepsilon_j \le y\} + f(y)\frac{1}{n}\sum_{j=1}^n \varepsilon_j\Big)[I\{X_i \le s\} - F_X(s)]\Big) + o_p(1).
\end{aligned}$$

Furthermore, one can show that $n^{-1}\sum_{j=1}^n I\{\varepsilon_j \le y\} + f(y)n^{-1}\sum_{j=1}^n \varepsilon_j$ can be replaced by its expectation $F(y)$, and therefore,

$$\begin{aligned}
\hat{T}_{n2}(s,y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big([I\{\varepsilon_i \le y\} - F(y)][I\{X_i \le s\} - F_X(s)] \\
&\quad + f(y)\varepsilon_i\Big[\int_0^s \frac{1}{h}K(\frac{X_i - x}{h})\,dx - F_X(s)\Big] + o_p(1).
\end{aligned}$$

To show weak convergence of the process we write the main term of $\hat{T}_{n2}(s,y)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(\varphi_{n,s,y,f(y)}(X_i,\varepsilon_i) - E[\varphi_{n,s,y,f(y)}(X_i,\varepsilon_i)]\Big),$$

where the functions are defined by

$$\varphi_{n,s,y,z}(X,\varepsilon) = (I\{\varepsilon \le y\} - F(y))(I\{X \le s\} - F_X(s)) + z\varepsilon\Big(\int_0^s \frac{1}{h}K(\frac{X - x}{h})\,dx - F_X(s)\Big)$$

(the n-dependence comes from $h = h_n$). The error density is bounded by assumption. Let $B > 0$ satisfy $f(y) \leq B$ for all $y \in \mathbb{R}$. We apply Theorem 2.11.23 in van der Vaart and Wellner (1996, p. 221) to the sequence of function classes

$$\mathcal{F}_n \;=\; \{\varphi_{n,s,y,z} \mid s \in [0,1], y \in \mathbb{R}, z \in [0,B]\}.$$

The envelope Φ_n of the class $\mathcal{F}_n$ is uniformly bounded by a function $\Phi_{env}(X,\varepsilon) = 1 + c\varepsilon$ for some constant c. The first two conditions of (2.11.21) in van der Vaart and Wellner (1996, p. 220) are fulfilled because $E[(\Phi_{env}(X_1,\varepsilon_1))^2] < \infty$. For the last condition in (2.11.21) we calculate

$$\begin{aligned}
&E\Big[(\varphi_{n,s,y,z}(X_1,\varepsilon_1) - \varphi_{n,s',y',z'}(X_1,\varepsilon_1))^2\Big]\\
=\;& E\Big[\Big((I\{\varepsilon_1 \leq y\} - F(y) - I\{\varepsilon_1 \leq y'\} + F(y'))(I\{X_1 \leq s\} - F_X(s))\\
&+ (I\{\varepsilon_1 \leq y'\} - F(y'))(I\{X_1 \leq s\} - F_X(s) - I\{X_1 \leq s'\} + F_X(s'))\\
&+ \varepsilon_1(z - z')\Big[\int_0^s \frac{1}{h}K(\frac{X_1 - x}{h})\,dx - F_X(s)\Big]\\
&+ \varepsilon_1 z'\Big[\int_0^s \frac{1}{h}K(\frac{X_1 - x}{h})\,dx - F_X(s) - \int_0^{s'} \frac{1}{h}K(\frac{X_1 - x}{h})\,dx + F_X(s')\Big]\Big)^2\Big].
\end{aligned}$$

For $s \leq s'$ we have

$$\int_0^1 \Big(\int_s^{s'} \frac{1}{h}K(\frac{u-x}{h})\,dx\Big)^2 du \;\leq\; \int_s^{s'}\int_0^1 \frac{1}{h}K(\frac{u-x}{h})\,du\,dx \;=\; O(1)(s'-s),$$

and therefore we can write

$$E\Big[(\varphi_{n,s,y,z}(X_1,\varepsilon_1) - \varphi_{n,s',y',z'}(X_1,\varepsilon_1))^2\Big] \;\leq\; C\rho((s,y,z),(s',y',z')),$$

where C is some constant independent of n, s, y, z, s', y', z' and we define the semimetric ρ on $\mathbb{T} = [0,1] \times \mathbb{R} \times [0,B]$ by

$$\rho((s,y,z),(s',y',z')) \;=\; |s - s'| + |F(y) - F(y')| + |z - z'|. \tag{A.4}$$

$\mathbb{T}$ is a totally bounded semimetric space with respect to ρ and condition (2.11.21) is satisfied. Next we show that the bracketing condition given in Theorem 2.11.23 in van der Vaart and Wellner (1996) is fulfilled, that is $\int_0^{\delta_n}(\log N_{[]}(\epsilon, \mathcal{F}_n, L_2(P)))^{1/2}d\epsilon$ converges to zero for every $\delta_n \searrow 0$. To this end, let $\epsilon > 0$. We choose grids $-\infty = y_1 < y_2 < \ldots < y_m = \infty$, $0 = s_1 < s_2 < \ldots < s_M = 1$ and $0 = z_1 < z_2 < \ldots < z_L = B$ such that $F(y_j) - F(y_{j-1}) < \epsilon^2$, $s_k - s_{k-1} < \epsilon^2$ and $z_l - z_{l-1} < \epsilon^2$. The number of grid points are $m = O(\epsilon^{-2})$, $M = O(\epsilon^{-2})$ and $L = O(\epsilon^{-2})$. We define $N_{[]}(\epsilon, \mathcal{F}_n, L_2(P)) = O(\epsilon^{-6})$ brackets for the class $\mathcal{F}_n$ by $[F^L_{j,k,l}, F^U_{j,k,l}]$ where

$$F^L_{j,k,l}(X,\varepsilon) \;=\; (I\{\varepsilon \leq y_{j-1}\} - F(y_j))(I\{X \leq s_{j-1}\} - F_X(s_j))$$

$$
\begin{aligned}
&\quad + z_{l-1}\varepsilon\Big(\int_0^{s_{j-1}} \frac{1}{h}K(\frac{X-x}{h})\,dx - F_X(s_j)\Big)\\
F^U_{j,k,l}(X,\varepsilon) &= (I\{\varepsilon \le y_j\} - F(y_{j-1}))(I\{X \le s_j\} - F_X(s_{j-1}))\\
&\quad + z_l\varepsilon\Big(\int_0^{s_j} \frac{1}{h}K(\frac{X-x}{h})\,dx - F_X(s_{j-1})\Big).
\end{aligned}
$$

Each bracket has L_2–norm smaller than ϵ which can be shown analogously to the verification of condition (2.11.21) above. The brackets cover $\mathcal{F}_n$ by construction. Note that the brackets depend on n but the number of brackets does not. Therefore the bracketing integral converges to zero for every $\delta_n \searrow 0$. From Theorem 2.11.23 we can deduce weak convergence of the process provided the sequence of covariances

$$
\begin{aligned}
& E\Big[\varphi_{n,s,y,z}(X_1,\varepsilon_1)\varphi_{n,s',y',z'}(X_1,\varepsilon_1)\Big] - E\Big[\varphi_{n,s,y,z}(X_1,\varepsilon_1)\Big]E\Big[\varphi_{n,s',y',z'}(X_1,\varepsilon_1)\Big]\\
= {} & E\Big[(I\{\varepsilon_1 \le y\} - F(y))(I\{\varepsilon_1 \le y'\} - F(y'))(I\{X_1 \le s\} - F_X(s))(I\{X_1 \le s'\} - F_X(s'))\Big]\\
& + \sigma^2 z z' E\Big[\Big(\int_0^s \frac{1}{h}K(\frac{X_1-x}{h})\,dx - F_X(s)\Big)\Big(\int_0^{s'} \frac{1}{h}K(\frac{X_1-x}{h})\,dx - F_X(s')\Big)\Big]\\
& + zE[\varepsilon_1(I\{\varepsilon_1 \le y'\} - F(y'))]E\Big[\Big(\int_0^s \frac{1}{h}K(\frac{X_1-x}{h})\,dx - F_X(s)\Big)(I\{X_1 \le s'\} - F_X(s'))\Big]\\
& + z'E[\varepsilon_1(I\{\varepsilon_1 \le y\} - F(y))]E\Big[\Big(\int_0^{s'} \frac{1}{h}K(\frac{X_1-x}{h})\,dx - F_X(s')\Big)(I\{X_1 \le s\} - F_X(s))\Big]
\end{aligned}
$$

converges pointwise (for $n \to \infty, h \to 0$). This is fulfilled because, for example,

$$
\begin{aligned}
E\Big[\int_0^s \frac{1}{h}K(\frac{X_1-x}{h})\,dx\Big] &= \int_0^1\int_0^s \frac{1}{h}K(\frac{u-x}{h})\,dx\,du\\
&= \int_h^{s\wedge(1-h)}\int_{-x/h}^{(1-x)/h} K(v)\,dv\,dx + O(h) = \int_0^s dx + o(1) = s + o(1),
\end{aligned}
$$

because for $h \to 0$ we have uniformly in $x \in (h, s\wedge(1-h))$ that $-x/h \le -1$ and $(1-x)/h \ge 1$. Note that in the special case $z = f(y)$, $z' = f(y')$ the sequence of covariances converges to $\mathrm{Cov}(G_2(s,y), G_2(s',y'))$ defined in Theorem 2.7. Weak convergence of the process $\hat{T}_{n2}(s,y)$ can be deduced from considering the subclass of $\mathcal{F}_n$ that is defined by $z = f(y)$. $\Box$

Proof of Proposition 3.1

Lemma A.3 *Under the assumptions of Proposition 3.1 we have*

$$
\sup_{s\in[0,1]}\sup_{y\in\mathbb{R}}\Big|\frac{1}{n}\sum_{i=1}^n \Big(I\{\hat{\varepsilon}_i \le y\} - I\{\varepsilon_i \le y\} - F(y + \hat{m}(x_{ni}) - m(x_{ni})) + F(y)\Big)I\{\frac{i}{n} \le s\}\Big| = o_p(\frac{1}{\sqrt{n}}).
$$

Proof: We are going to apply Theorem 2.11.9 in van der Vaart and Wellner (1996, p. 211) to the process

$$
\sum_{i=1}^n Z^o_{ni}(\varphi) = \sum_{i=1}^n (Z_{ni}(\varphi) - E[Z_{ni}(\varphi)]), \quad \varphi = (y,d,s) \in \mathcal{F},
$$

where $\mathcal{F} = \{(y, d, s) \mid y \in \mathbb{R}, d \in C_1^{1+\delta}[0,1], s \in [0,1]\}$ and

$$Z_{ni}(\varphi) \;=\; n^{-1/2}\Big(I\{\varepsilon_i \le y + d(x_{ni})\} - I\{\varepsilon_i \le y\}\Big) I\{\frac{i}{n} \le s\}.$$

We have $\hat{m} - m \in C_1^{1+\delta}[0,1]$ with probability converging to one by Proposition 3 and Lemma B.1 in Akritas and Van Keilegom (2001). First note that $\mathcal{F}$ is a totally bounded semimetric space with semimetric $\overline{\rho}$ defined by

$$\overline{\rho}((y,d,s),(y',d',s')) \;=\; \max\Big\{D(y,y'), \sup_{x\in[0,1]} |d(x) - d'(x)|, |s-s'|\Big\}, \tag{A.5}$$

where

$$D(y,y') \;=\; \sup_{x\in[0,1]} \sup_{d \in C_1^{1+\delta}[0,1]} |F(y + d(x)) - F(y' + d(x))|.$$

We postpone the verification that $(\mathcal{F}, \overline{\rho})$ is a totally bounded semimetric space to the end of this proof.

Because $\sup_{f\in\mathcal{F}} |Z_{ni}(f)| \le n^{-1/2}$ the first condition in Theorem 2.11.9 in van der Vaart and Wellner (1996, p. 211) is fulfilled, that is

$$\sum_{i=1}^n E\Big[\sup_{f\in\mathcal{F}} |Z_{ni}(f)| I\{\sup_{f\in\mathcal{F}} |Z_{ni}(f)| > \eta\}\Big] \;=\; o(1)$$

for all $\eta > 0$. For the verification of the second condition we assume $\delta_n = o(1)$, $\varphi = (y, d, s)$, $\varphi' = (y', d', s')$ and write

$$\begin{aligned}
&\sup_{\overline{\rho}(\varphi,\varphi')<\delta_n} \sum_{i=1}^n (Z_{ni}(\varphi) - Z_{ni}(\varphi'))^2 \\
\le\; & \sup_{\overline{\rho}(\varphi,\varphi')<\delta_n} \frac{1}{n}\sum_{i=1}^n \Big(8E[(I\{\varepsilon_i \le y + d(x_{ni})\} - I\{\varepsilon_i \le y + d'(x_{ni})\})^2] \\
&+ 8E[(I\{\varepsilon_i \le y + d'(x_{ni})\} - I\{\varepsilon_i \le y' + d'(x_{ni})\})^2] \\
&+ 4E[(I\{\varepsilon_i \le y\} - I\{\varepsilon_i \le y'\})^2] + 8E[(I\{\frac{i}{n} \le s\} - I\{\frac{i}{n} \le s'\})^2]\Big) \\
\le\; & 8 \sup_{\overline{\rho}(\varphi,\varphi')<\delta_n} \frac{1}{n}\sum_{i=1}^n \Big(|F(y + d(x_{ni})) - F(y + d'(x_{ni}))| + |F(y + d'(x_{ni})) - F(y' + d'(x_{ni}))| \\
&+ |F(y) - F(y')| + \frac{\lfloor ns \rfloor - \lfloor ns' \rfloor}{n}\Big) \\
\le\; & 8\Big(\sup_{y\in\mathbb{R}} f(y) \sup_{\overline{\rho}(\varphi,\varphi')<\delta_n} \frac{1}{n}\sum_{i=1}^n |d(x_{ni}) - d'(x_{ni})| + 2\delta_n + \delta_n + \frac{1}{n}\Big) \\
=\; & O(\delta_n + \frac{1}{n}) \;=\; o(1).
\end{aligned}$$

Next we verify the third condition in the aforementioned theorem, that is,

$$\int_0^{\delta_n} (\log N_{[\,]}(\epsilon, \mathcal{F}, L_2^n))^{1/2} d\epsilon \longrightarrow 0 \text{ for all } \delta_n \searrow 0. \tag{A.6}$$

In order to calculate the bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_2^n)$ for a fixed $\epsilon > 0$ we partition $\mathcal{F}$ in the following way into subregions $\mathcal{F}_{mkj}$ such that

$$\sum_{i=1}^n E\Big[\sup_{\varphi,\varphi' \in \mathcal{F}_{mkj}} |Z_{ni}(\varphi) - Z_{ni}(\varphi')|^2\Big]$$

$$\text{(A.7)} \leq 2n^{-1}\sum_{i=1}^n E\Big[\sup_{\varphi,\varphi' \in \mathcal{F}_{mkj}} \Big| I\{\varepsilon_i \leq y + d(x_{ni})\} I\{\frac{i}{n} \leq s\} - I\{\varepsilon_i \leq y' + d'(x_{ni})\} I\{\frac{i}{n} \leq s'\}\Big|^2\Big]$$

$$\text{(A.8)} \quad + 2n^{-1}\sum_{i=1}^n E\Big[\sup_{\varphi,\varphi' \in \mathcal{F}_{mkj}} \Big| I\{\varepsilon_i \leq y\} I\{\frac{i}{n} \leq s\} - I\{\varepsilon_i \leq y'\} I\{\frac{i}{n} \leq s'\}\Big|^2\Big]$$

$$\leq C\epsilon^2$$

for some constant C [where $\varphi = (y, d, s)$, $\varphi' = (y', d', s')$]. In the following we concentrate on (A.7) [(A.8) is treated similarly]. To this end let $d_m^L \leq d_m^U$, $m = 1, \ldots, M$, be $M = O(\exp(\kappa\epsilon^{-2/(1+\delta)}))$ brackets for $C_1^{1+\delta}[0,1]$ with length ϵ^2 with respect to the supremum norm (see van der Vaart and Wellner, 1996, Th. 2.7.1 and Cor. 2.7.2). Then define for each $m = 1, \ldots, M$, $F_m^L(y) = n^{-1}\sum_{i=1}^n P(\varepsilon_1 \leq y + d_m^L(x_{ni}))$ and let y_{mk}^L ($k = 1, \ldots, K = O(\epsilon^{-2})$) partition the line in segments such that $F_m^L(y_{mk}^L) - F_m^L(y_{m,k-1}^L) < \epsilon^2$. Define y_{mk}^U in a similar way using $F_m^U(y) = n^{-1}\sum_{i=1}^n P(\varepsilon_1 \leq y + d_m^U(x_{ni}))$. For the brackets of $[0,1]$ we have to consider two cases. First let $\epsilon^2 \geq n^{-1}$. Then we divide $[0,1]$ into $J = O(\epsilon^{-2})$ subintervals $[s_j, s_{j+1})$ of length less or equal to ϵ^2. Now define

$$\mathcal{F}_{mkj} = \{(y, d, s) \mid \tilde{y}_{mk}^L \leq y \leq \tilde{y}_{mk}^U, d_m^L \leq d \leq d_m^U, s_j \leq s < s_{j+1}\}$$

where $\tilde{y}_{mk}^L = y_{mk}^L$ and $\tilde{y}_{mk}^U$ is the smallest of the y_{mk}^U which is larger than or equal to $y_{m,k+1}^L$. We obtain as upper bound for (A.7),

$$\begin{aligned}
& 2n^{-1}\sum_{i=1}^n E\Big[I\{\varepsilon_i \leq \tilde{y}_{mk}^U + d_m^U(x_{ni})\} - I\{\varepsilon_i \leq \tilde{y}_{mk}^L + d_m^L(x_{ni})\}\Big] \\
& \quad + 2n^{-1}\sum_{i=1}^n I\{\frac{i}{n} < s_{j+1}\} - I\{\frac{i}{n} \leq s_j\} \\
&= 2(F_m^U(\tilde{y}_{mk}^U) - F_m^L(\tilde{y}_{mk}^L)) + 2(n^{-1} + \epsilon^2) \\
&\leq 2\Big(|F_m^U(\tilde{y}_{mk}^U) - F_m^U(\tilde{y}_{m,k+1}^L)| + |F_m^U(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{m,k+1}^L)| \\
& \quad + |F_m^L(\tilde{y}_{m,k+1}^L) - F_m^L(\tilde{y}_{mk}^L)| + (n^{-1} + \epsilon^2)\Big) \\
&= O(\epsilon^2)
\end{aligned}$$

where the last equality follows from $n^{-1} \leq \epsilon^2$, the definitions of $\tilde{y}_{mk}^L$, $\tilde{y}_{mk}^U$ and the following consideration,

$$\begin{aligned}
\sup_{y\in\mathbb{R}} |F_m^U(y) - F_m^L(y)| &\leq \sup_{y\in\mathbb{R}} n^{-1}\sum_{i=1}^n |F(y + d_m^U(x_{ni})) - F(y + d_m^L(x_{ni}))| \\
&\leq \sup_{y\in\mathbb{R}} f(y) \sup_{x\in[0,1]} |d_m^U(x) - d_m^L(x)| = O(\epsilon^2).
\end{aligned}$$

In the second case, i. e. $\epsilon^2 < n^{-1}$ we consider brackets $[\tilde{s}_j, \tilde{s}_{j+1})$ where $\tilde{s}_j < \tilde{s}_{j+1}$ and $\{\tilde{s}_j \mid j = 1, \ldots, \tilde{J}\} = \{\tilde{s}_j \mid j = 1, \ldots, \tilde{J}\} \cup \{\frac{k}{n} \mid k = 1, \ldots, n\}$. In this case we have $\tilde{J} = O(\epsilon^{-2} + n) = O(\epsilon^{-2})$ subintervals and

$$n^{-1} \sum_{i=1}^{n} I\{\frac{i}{n} < s_{j+1}\} - I\{\frac{i}{n} \leq s_j\} = 0$$

always. The rest of the derivation for (A.7) is as above. In either case the bracketing number MKJ resp. $MK\tilde{J}$ does not depend on n and is of the order $O(\epsilon^{-4} \exp(\kappa \epsilon^{-2/(1+\delta)}))$. Therefore, (A.6) is fulfilled.
The convergence of the marginal distribution can be shown easily using Cramér-Wold's device and Lindeberg's condition. From the proof of Theorem 2.11.9 in van der Vaart and Wellner (1996, p. 220) follows asymptotic equicontinuity, that is

$$P\Big(\sup_{\overline{\rho}(\varphi,\varphi')<\delta_n} \Big| \sum_{i=1}^{n} (Z_{ni}^o(\varphi) - Z_{ni}^o(\varphi')) \Big| > \epsilon \Big) \longrightarrow 0$$

for all $\delta_n \searrow 0$. Our assertion can be deduced for $\varphi = (y, \hat{m} - m, s)$, $\varphi' = (y, 0, s)$.
It remains to show that $(\mathcal{F}, \overline{\rho})$ is a totally bounded semimetric space. It is easy to see that $\overline{\rho}$ defined in (A.5) is a semimetric. To prove that $\mathcal{F}$ is totally bounded we show that for all $\epsilon > 0$ the bracketing number $N_{[\,]}(\epsilon, \mathcal{F}, \overline{\rho})$ is finite (see van der Vaart and Wellner, 1996, p. 84). To this end we choose similar as before subintervals $[s_j, s_{j+1}]$, $j = 1, \ldots, J = O(\epsilon^{-1})$, of length less or equal to ϵ for $[0, 1]$ and brackets $d_m^L \leq d_m^U$, $m = 1, \ldots, M = O(\exp(\kappa \epsilon^{-1/(1+\delta)}))$, of length ϵ with respect to the supremum norm. Then, for fixed j and m let y_{jmk}, $k = 1, \ldots, K = O(\epsilon^{-1})$, partition the line in intervals of probability less or equal to ϵ with respect to the probability measure $F_{jm}(y) = P(\varepsilon_1 \leq y + d_m^L(s_j))$. Now, let $y_0, y_1, \ldots, y_L$, $L = O(\epsilon^{-2} \exp(\kappa \epsilon^{-1/(1+\delta)}))$, denote all ordered y_{jmk} values with $y_0 = -\infty$, $y_L = \infty$. Then, for all $\ell = 0, \ldots, L-1$,

$$\begin{aligned} D(y_\ell, y_{\ell+1}) &= \sup_{x\in[0,1]} \sup_{d\in C_1^{1+\delta}[0,1]} |F(y_{\ell+1} + d(x)) - F(y_\ell + d(x))| \\ &\leq \max_{j=1,\ldots,J-1} \sup_{x\in[s_j,s_{j+1}]} \max_{m=1,\ldots,M-1} \sup_{d_m^L \leq d \leq d_m^U} \Big\{ |F(y_{\ell+1} + d(x)) - F(y_{\ell+1} + d(s_j))| \\ &\quad + |F(y_{\ell+1} + d(s_j)) - F(y_{\ell+1} + d_m^L(s_j))| + |F(y_{\ell+1} + d_m^L(s_j)) - F(y_\ell + d_m^L(s_j))| \\ &\quad + |F(y_\ell + d_m^L(s_j)) - F(y_\ell + d(s_j))| + |F(y_\ell + d(s_j)) - F(y_\ell + d(x))| \Big\}. \end{aligned}$$

For some k, y_ℓ and $y_{\ell+1}$ belong to the same interval $[y_{jmk}, y_{jm,k+1}]$ and therefore, $|F(y_{\ell+1} + d_m^L(s_j)) - F(y_\ell + d_m^L(s_j))| \leq \epsilon$. Further, $\sup_{x\in[0,1]} |d'(x)| \leq 1$ by definition of $C_1^{1+\delta}[0, 1]$. We obtain

$$\begin{aligned} D(y_\ell, y_{\ell+1}) &\leq \max_{j=1,\ldots,J-1} \sup_{x\in[s_j,s_{j+1}]} \max_{m=1,\ldots,M-1} \sup_{d_m^L \leq d \leq d_m^U} \Big\{ 2 \sup_{y\in\mathbb{R}} f(y) |x - s_j| \\ &\quad + 2 \sup_{y\in\mathbb{R}} f(y) \sup_{x\in[0,1]} |d(x) - d_m^L(x)| + \epsilon \Big\} \\ &\leq c\epsilon \end{aligned}$$

for some constant c. This proves the bracketing number to be bounded. $\square$

From the expansion in Lemma A.3 we obtain by a Taylor expansion and the definition of $\hat{m}$ similar to the proof of Proposition 2.6,

$$\begin{aligned}\frac{\lfloor ns\rfloor}{n}\Big(\hat{F}_{\lfloor ns\rfloor}(y)-F(y)\Big) &= \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(I\{\varepsilon_i\le y\}-F(y))+\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(F(y+\hat{m}(x_{ni})-m(x_{ni}))-F(y))\\ &\quad+o_p(\frac{1}{\sqrt{n}})\\ &= \frac{\lfloor ns\rfloor}{n}\big(F_{\lfloor ns\rfloor}(y)-F(y)\big)+f(y)\frac{1}{n}\sum_{j=1}^{n}\varepsilon_j\Big(\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}\frac{1}{h}K(\frac{x_{ni}-x_{nj}}{h})\frac{1}{f_X(x_{ni})}\Big)\\ &\quad+o_p(\frac{1}{\sqrt{n}})\end{aligned}$$

uniformly in $y\in\mathbb{R}$ and $s\in[0,1]$. For the deterministic sum we obtain via a Riemann-sum-approximation

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}\frac{1}{h}K(\frac{x_{ni}-x}{h})\frac{1}{f_X(x_{ni})} &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\Big(\frac{F_X^{-1}(g(\frac{i}{n}))-x}{h}\Big)\frac{1}{f_X(F_X^{-1}(g(\frac{i}{n})))}I\{\frac{i}{n}\le s\}\\ &= \int_0^1\frac{1}{h}K\Big(\frac{F_X^{-1}(g(t))-x}{h}\Big)\frac{1}{f_X(F_X^{-1}(g(t)))}I\{t\le s\}\,dt\,(1+o(1))\\ &= \int_{S(s)}\frac{1}{h}K\Big(\frac{u-x}{h}\Big)\,du\,(1+o(1)),\end{aligned}$$

uniformly with respect to $s\in[0,1]$ and the assertion of Proposition 3.1 follows. $\square$

Proof of Theorem 3.2

The expansion of $\hat{T}_{n3}$ follows from Proposition 3.1 and the analogous result for $\hat{F}^*_{n-\lfloor ns\rfloor}$. To show weak convergence we write the process as $\hat{T}_{n3}(s,y)=\sum_{i=1}^n Z_{ni}(s,y)$ where

$$Z_{ni}(s,y) = \frac{1}{\sqrt{n}}\Big(I\{\varepsilon_i\le y\}\Big[I\{\frac{i}{n}\le s\}-\frac{\lfloor ns\rfloor}{n}\Big]+f(y)\varepsilon_i\Big[\int_{S(s)}\frac{1}{h}K(\frac{x_{ni}-x}{h})dx-\frac{\lfloor ns\rfloor}{n}\Big]\Big)$$

but similar to the proof of Theorem 2.7 we first consider the process $\sum_{i=1}^n Z^*_{ni}(s,y,z)$ where

$$Z^*_{ni}(s,y,z) = \frac{1}{\sqrt{n}}\Big(I\{\varepsilon_i\le y\}\Big[I\{\frac{i}{n}\le s\}-\frac{\lfloor ns\rfloor}{n}\Big]+z\varepsilon_i\Big[\int_{S(s)}\frac{1}{h}K(\frac{x_{ni}-x}{h})dx-\frac{\lfloor ns\rfloor}{n}\Big]\Big)$$

and $(s,y,z)\in\mathbb{T}=[0,1]\times\mathbb{R}\times[0,B]$. $\mathbb{T}$ is a totally bounded semimetric space with respect to the semimetric ρ defined by

$$\rho((s,y,z),(s',y',z')) = \max(|s-s'|,|g(s)-g(s')|)+|F(y)-F(y')|+|z-z'|. \tag{A.9}$$

We apply Theorem 2.11.9 in van der Vaart and Wellner (1996, p. 211). Because we have the bound

$$|Z^*_{ni}(s,y,z)| \;\leq\; n^{-1/2}(1+c\varepsilon_i)$$

(uniformly in n, s, y, z) the first condition of this theorem, that is

$$\sum_{i=1}^n E\Big[\sup_{(s,y,z)\in\mathbb{T}} |Z^*_{ni}(s,y,z)| I\{\sup_{(s,y,z)\in\mathbb{T}} |Z^*_{ni}(s,y,z)| > \eta\}\Big] \longrightarrow 0 \text{ for every } \eta > 0,$$

follows from

$$\begin{aligned} n^{1/2}E\left[|\varepsilon_1| I\{|\varepsilon_1| > \eta n^{1/2}\}\right] \;&=\; \frac{1}{\eta}E\left[\eta n^{1/2}|\varepsilon_1| I\{|\varepsilon_1| > \eta n^{1/2}\}\right] \\ &\leq\; \frac{1}{\eta}E\left[\varepsilon_1^2 I\{|\varepsilon_1| > \eta n^{1/2}\}\right]. \end{aligned}$$

The last term converges to zero for $n\to\infty$ for all $\eta>0$ because $E[\varepsilon_1^2]<\infty$. For the second condition in the aforementioned theorem we bound for some constant $C>0$

$$\begin{aligned} &\sup_{\delta_n} \sum_{i=1}^n E\left[(Z^*_{ni}(s,y,z) - Z^*_{ni}(s',y',z'))^2\right] \\ &\leq C \sup_{\delta_n, s>s'} \Big[|F(y)-F(y')| + \frac{1}{n}\sum_{i=1}^n \Big| I\{\frac{i}{n}\leq s\} - I\{\frac{i}{n}\leq s'\}\Big| \\ &\quad + \Big(\frac{\lfloor ns\rfloor}{n} - \frac{\lfloor ns'\rfloor}{n}\Big)^2 + (z-z')^2 + \frac{1}{n}\sum_{i=1}^n \Big(\int_{U(s,s')} \frac{1}{h}K(\frac{x_{ni}-x}{h})\,dx\Big)^2\Big] \\ &\leq O(\delta_n) + C\sup_{\delta_n, s>s'} \Big[\frac{|\lfloor ns\rfloor - \lfloor ns'\rfloor|}{n} + \frac{1}{n}\sum_{i=1}^n \Big(\int_{U(s,s')} \frac{1}{h}K(\frac{x_{ni}-x}{h})\,dx\Big)^2\Big] \end{aligned}$$

where $U(s,s') = \{F_X^{-1}(g(t)) : s' \leq t \leq s\}$ for $s>s'$, $\delta_n \searrow 0$ and we used the notation $\sup_{\delta_n} = \sup_{\rho((s,y,z),(s',y',z'))<\delta_n}$. Take δ_n small enough, namely smaller than one half of the smallest jump size of the function g, to make sure that the function g does not make any jump in between s and s'. Restrict attention for simplicity to the cases where s and s' belong to a line segment with slope 1 (the case of slope -1 can be considered analogously). In that case $s>s'$ implies $g(s)>g(s')$. We further have $\sup_{|s-s'|<\delta_n} |\lfloor ns\rfloor - \lfloor ns'\rfloor|/n = O(\max(\delta_n, n^{-1}))$ and by a change of variable in the integral,

$$\begin{aligned} &\sup_{\delta_n, s>s'} \frac{1}{n}\sum_{i=1}^n \Big(\int_{U(s,s')} \frac{1}{h}K(\frac{x_{ni}-x}{h})\,dx\Big)^2 \\ \leq\; &\sup_{\delta_n, s>s'} \frac{1}{n}\sum_{i=1}^n \Big(\sup_{u\in[-1,1]}|K(u)|\Big[\Big\{\frac{F_X^{-1}(g(s))-x_{ni}}{h}\wedge 1\Big\} - \Big\{\frac{F_X^{-1}(g(s'))-x_{ni}}{h}\vee(-1)\Big\}\Big] \\ &\times I\{F_X^{-1}(g(s)) - x_{ni} > -h\} I\{F_X^{-1}(g(s')) - x_{ni} < h\}\Big)^2 \end{aligned}$$

$$= O(1) \sup_{\delta_n, s>s'} \frac{1}{n} \sum_{i=1}^n \Big[\frac{F_X^{-1}(g(s)) - F_X^{-1}(g(s'))}{h} \wedge 2 \Big]^2$$

$$\times I\{F_X(F_X^{-1}(g(s')) - h) < g(\frac{i}{n}) < F_X(F_X^{-1}(g(s)) + h)\}$$

$$= O(\min(\frac{\delta_n}{h}, 1)^2 \max(\delta_n, h, n^{-1})) \ = \ O(\min(\frac{\delta_n}{h}, 1) \max(\delta_n, h)) \ = \ O(\delta_n),$$

where the last line follows from a first order Taylor expansion and our assumption that the density f_X is bounded and bounded away from zero. All in all the second condition in Theorem 2.11.9, van der Vaart and Wellner (1996, p. 211), that is

$$\sup_{\delta_n} \sum_{i=1}^n E\left[(Z_{ni}^*(s,y,z) - Z_{ni}^*(s',y',z'))^2 \right] \longrightarrow 0 \text{ for all } \delta_n \searrow 0,$$

is satisfied. It remains to show the third condition, that is,

$$\text{(A.10)} \qquad \int_0^{\delta_n} (\log N_{[\,]}(\epsilon, \mathbb{T}, L_2^n))^{1/2} d\epsilon \longrightarrow 0 \text{ for all } \delta_n \searrow 0.$$

In order to calculate the bracketing number $N_{[\,]}(\epsilon, \mathbb{T}, L_2^n)$ for a fixed $\epsilon > 0$ we partition $\mathbb{T}$ in the following way into subregions $\mathbb{T}_{jkl} = [s_j, s_{j+1}) \times [y_k, y_{k+1}) \times [z_l, z_{l+1})$ such that

$$\sum_{i=1}^n E\Big[\sup_{\substack{(s,y,z), \\ (s',y',z') \in \mathbb{T}_{jkl}}} |Z_{ni}^*(s,y,z) - Z_{ni}^*(s',y',z')|^2 \Big] \ \le \ \epsilon^2.$$

To this end we have to consider two cases. First let $\epsilon^2 \ge n^{-1}$. Then we divide $[0,1]$ into $J = O(\epsilon^{-2})$ subintervals $[s_j, s_{j+1})$ such that $\max(|s_j - s_{j+1}|, |g(s_j) - g(s_{j+1}-)|) \le \epsilon^2$, divide $[0,B]$ into $O(\epsilon^{-2})$ subintervals $[z_l, z_{l+1})$ of length less or equal to ϵ^2 and divide $\mathbb{R}$ into $O(\epsilon^{-2})$ subintervals $[y_k, y_{k+1})$ such that $|F(y_k) - F(y_{k+1})| \le \epsilon^2$. With this definition we have $\rho((s,y,z),(s',y',z')) \le \epsilon^2$ for all $(s,y,z),(s',y',z') \in \mathbb{T}_{jkl}$. It is then possible to use the monotonicity of each term, for instance,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n E\Big[\sup_{y_k \le y, y' \le y_{k+1}} |I\{\varepsilon_i \le y\} - I\{\varepsilon_i \le y'\}| \Big] \ &= \ \frac{1}{n} \sum_{i=1}^n E[I\{\varepsilon_i \le y_{k+1}\} - I\{\varepsilon_i \le y_k\}] \\ &= \ F(y_{k+1}) - F(y_k) \ \le \ \epsilon^2, \end{aligned}$$

to yield with analogous calculations as above (for the verification of the second condition)

$$\sum_{i=1}^n E\Big[\sup_{\substack{(s,y,z), \\ (s',y',z') \in \mathbb{T}_{jkl}}} |Z_{ni}^*(s,y,z) - Z_{ni}^*(s',y',z')|^2 \Big] \ = \ O(\epsilon^2) + O(\epsilon^2 + n^{-1}) \ = \ O(\epsilon^2).$$

For the second case, that is $\epsilon^2 < n^{-1}$, we choose subintervals $[z_l, z_{l+1})$ of $[0,B]$ and $[y_k, y_{k+1})$ of $\mathbb{R}$ as before, but divide $[0,1]$ into $\tilde{J} = O(\epsilon^{-2}) + n = O(\epsilon^{-2})$ subintervals $[\tilde{s}_j, \tilde{s}_{j+1})$, where

$\{\tilde{s}_j \mid 0 \le j \le \tilde{J}\} = \{s_j \mid 0 \le j \le J\} \cup \{\frac{k}{n} \mid 1 \le k \le n\}$. With this definition we obtain, for instance,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E\Big[\sup_{\tilde{s}_j \le s, s' < \tilde{s}_{j+1}} |I\{\frac{i}{n} \le s\} - I\{\frac{i}{n} \le s'\}|^2\Big] \;&\le\; \frac{1}{n}\sum_{i=1}^{n} I\{i < n\tilde{s}_{j+1}\} - I\{i \le n\tilde{s}_j\} \\ &=\; O(\epsilon^2).\end{aligned}$$

In both cases the bracketing number $N_{[]}(\epsilon, \mathbb{T}, L_2^n) = O(\epsilon^{-6})$ does not depend on n and condition (A.10) is valid. To obtain the weak convergence of the process by Theorem 2.11.9, van der Vaart and Wellner (1996, p. 211), it remains to verify the convergence of the marginals. Applying Cramér Wold's device we consider the random sequence $Z_i = \sum_{j=1}^{k} a_j Z_{ni}(s_j, y_j)$ for constants $a_j \in \mathbb{R}$, $s_j \in [0,1]$, $y_j \in \mathbb{R}$ $(j = 1, \dots, k)$ and show Lindeberg's condition, that is

$$\sum_{i=1}^{n} E\Big[Z_i^2 I\{|Z_i| > \delta\}\Big] \;\le\; (\sum_{j=1}^{k} |a_j|)^2 E\Big[(1 + c\varepsilon_1)^2 I\Big\{|1 + c\varepsilon_1| > \sqrt{n}\delta / \sum_{j=1}^{k} |a_j|\Big\}\Big] \;=\; o(1).$$

The proof is finished by a straightforward calculation of the covariances by using Riemann-sum-approximations like in the proof of Proposition 3.1, and by noting that

$$\begin{aligned}\int_{S(s)} \frac{1}{h} K(\frac{x_{ni} - x}{h})\, dx \;&=\; I(x_{ni} \in S(s)) + o(1) \\ &=\; I\Big(g(\frac{i}{n}) = g(t) \text{ for some } t \le s\Big) + o(1) = I\Big(\frac{i}{n} \le s\Big) + o(1).\end{aligned}$$

□

Acknowledgements. Part of this paper was written while the first author visited the Australian National University in Canberra and this author would like to thank the members of the Mathematical Sciences Institute for their hospitality. The financial support by the Deutsche Forschungsgemeinschaft (SFB 475 and research grant Ne1189/1-1) is gratefully acknowledged. The research of the second author was supported by IAP research network grant nr. P5/24 of the Belgian government (Belgian Science Policy).

References

M. Akritas and I. Van Keilegom (2000). Nonparametric estimation of the residual distribution. *Scand. J. Statist.* 28, 549–567.

A. Antoniadis and I. Gijbels (2002). Detecting abrupt changes by wavelet methods. *J. Nonparametric Stat.* 14, 7–29.

J. Bai (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models. *Ann. Statist.* 22, 2051–2061.

M. V. Boldin (2002). On sequential residual empirical processes in heteroscedastic time series. *Math. Methods Statist.* 11, 453–464.

E. Carlstein (1988). Nonparametric change-point estimation. *Ann. Statist.* 16, 188–197.

F. Cheng (2002). Consistency of error density and distribution function estimation in nonparametric regression. *Statist. Prob. Letters* 59, 257–270.

M. Csörgö, B. Szyszkowicz (2000). Weighted quantile processes and their applications to change-point analysis. Stochastic models (Ottawa, ON, 1998), 67–84, CMS Conf. Proc., 26, Amer. Math. Soc., Providence, RI.

L. Dümbgen (1991). The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.* 19, 1471–1495.

J. Einmahl, I. Van Keilegom (2006). Tests for independence in nonparametric regression (submitted).

D. Ferger (1994). On the power of nonparametric change-point-tests. *Metrika* 41, 277–292.

D. Ferger (1996). On the aymptotic behavior of change-point estimators in case of no change with applications to testing. *Statist. Decis.* 14, 137–143.

D. Ferger and W. Stute (1992). Convergence of change-point estimators. *Stoch. Process. Appl.* 42, 345–351.

I. Gijbels and A.–C. Goderniaux (2004). Bootstrap test for change-points in nonparametric regression. *J. Nonparametric Stat.* 16, 591–611.

L. Giraitis, R. Leipus and D. Surgailis (1996) The change-point problem for dependent observations. *J. Statist. Plann. Inf.* 53, 297–310.

G. Grégoire and Z. Hamrouni (2002). Change point estimation by local linear smoothing. *J. Multivariate Anal.* 83, 56–83.

L. Horvath and P. Kokoszka (2002). Change-point detection with nonparametric regression. *Statistics* 36, 9–31.

L. Horvath, P. Kokoszka and G. Teyssiere (2001). Empirical process of the squared residuals of an ARCH sequence. *Ann. Statist.* 29, 445–469.

D. A. Hsu (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *J. Amer. Statist. Assoc.* 74, 31–40.

J. Huh and K.-H. Kang (2005). Nonparametric estimation of the variance function with a change point (submitted).

A. Inoue (2001). Testing for distributional change in time series. *Economet. Theory* 17, 156–187.

H. L. Koul (1996). Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Ann. Statist.* 24, 380–404.

S. Lee and S. Na (2004). A nonparametric test for the change in the density function in strong mixing processes. *Statist. Prob. Letters* 66, 1–25.

S. Ling (1998). Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.* 26, 741–754.

C. R. Loader (1996). Change point estimation using nonparametric regression. *Ann. Statist.* 24, 1667–1678.

H.–G. Müller (1992). Change-points in nonparametric regression analysis. *Ann. Statist.* 20, 737–761.

U. U. Müller, A. Schick, W. Wefelmeyer (2004). Estimating the error distribution function in nonparametric regression. preprint.
http://www.mi.uni-koeln.de/ wefelm/files/cdfsmooth06.pdf

É. A. Nadaraya (1964). On non–parametric estimates of density functions and regression curves. *J. Probab. Appl.* 10, 186–190.

N. Neumeyer (2006). Smooth residual bootstrap for empirical processes of nonparametric regression residuals. preprint: http://www.math.uni-hamburg.de/research/ims.html

N. Neumeyer and H. Dette (2006). Testing for symmetric error distribution in nonparametric regression models. *Statist. Sinica* (to appear).

A. J. Oyet and B. Sutradhar (2003). Testing variances in wavelet regression models. *Statist. Prob. Letters* 61, 97–109.

J. C. Pardo-Fernández, I. Van Keilegom and E. González-Manteiga (2006). Comparison of regression curves based on the estimation of the error distribution. *Statist. Sinica* (to appear).

D. Picard (1985). Testing and estimating change-points in time series. *Adv. Appl. Prob.* 17, 841–867.

Y. Ritov (1990). Asymptotic efficient estimation of the change point with unknown distributions. *Ann. Statist.* 18, 1829–1939.

D. Ruppert, M. P. Wand and R. J. Carroll (2003). *Semiparametric regression.* Cambridge University Press, Cambridge.

G. R. Shorack and J. A. Wellner (1986). *Empirical processes with applications to statistics.* Wiley, New York.

A. W. van der Vaart (1996). *Asymptotic statistics.* Cambridge Univ. Press, Cambridge.

A. W. van der Vaart and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer, New York.

G. S. Watson (1964). Smooth Regression Analysis. *Sankhya A* 26, 359–372.