Bootstrap tests for the error distribution in linear and nonparametric regression models

Natalie Neumeyer, Holger Dette, Eva-Renate Nagel

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

email: holger.dette@ruhr-uni-bochum.de

June 4, 2004

Abstract

In this paper we investigate several tests for the hypothesis of a parametric form of the error distribution in the common linear and nonparametric regression model, which are based on empirical processes of residuals. It is well known that tests in this context are not asymptotically distribution-free and the parametric bootstrap is applied to deal with this problem. The performance of the resulting bootstrap test is investigated from an asymptotic point of view and by means of a simulation study. The results demonstrate that even for moderate sample sizes the parametric bootstrap provides a reliable and easy accessible solution to the problem of goodness-of-fit testing of assumptions regarding the error distribution in linear and nonparametric regression models.

AMS Classification: 62J05, 62J10, 62G10

Keywords and Phrases: goodness-of-fit, residual process, parametric bootstrap, linear model, analysis of variance, M-estimation, nonparametric regression

1 Introduction

Parametric and nonparametric models are widely used in subject areas such as biology, chemistry or economics. The general model is usually written in the form

$$Y = m(x) + \varepsilon, \tag{1.1}$$

where m is the regression function, the error ε satisfies $E[\varepsilon] = 0$, $E[\varepsilon^2] > 0$ and x is an explanatory variable. Throughout this paper we assume that the errors corresponding to different observations are independent identically distributed. The most popular model is certainly the linear model where $m(x)$ can be represented in the form $x^T\beta$, β is an unknown parameter and classical inference in this model assumes a normal distribution for the error ε. However, occasionally statisticians are confronted with results arising from an experiment with data, which do not necessarily satisfy the assumption of normality. In some cases a transformation of the observations allows to apply classical techniques. Although this approach may lead to a satisfactory analysis, the results may be questionable in many situations. For this reason several authors have proposed procedures for analyzing the data with the linear model without the assumption of a normal distribution for the error [see for example Randles and Wolfe (1979), Denker (1985), Thompson (1991), Brunner and Neumann (1987), Marden and Muyot (1995) among many others]. On the other hand, statistical analysis in the linear model under the assumption of normality is still attractive to practioners due to its simplicity and efficiency (if the assumption of a normal distributed error is met). Therefore many authors point out that it is important to apply a goodness-of-fit test for the error distribution before analyzing the data with the linear model under a specific assumption for the error distribution.
Moreover, even in the general nonparametric regression model (1.1) additional knowledge about the error distribution can improve the efficiency of the statistical analysis. For example under the additional assumption of a Gaussian error exact or optimal tests can be derived in many cases. A typical example is the problem of goodness-of-fit testing regarding the regression function m [see Eubank and Hart (1992), Milhem (2003) or Young and Bowman (1995)]. Furthermore, the nonparametric regression model with normal errors is asymptotically equivalent to a Gaussian White Noise model [see Brown and Low (1996) and Brown, Cai, Low and Zhang (2002)]. In this context a huge amount of literature about optimal statistical procedures is available [see among others Lepski and Tsybakov (2000) or Birgé and Massart (2001)].

It is the purpose of the present paper to investigate the performance of bootstrap tests for a parametric hypothesis regarding the error distribution in models of the form (1.1). To be precise let

$$\mathcal{F} = \{F_\vartheta \mid \vartheta \in \Theta\} \tag{1.2}$$

denote a parametric class of distribution functions F_ϑ with densities f_ϑ, where $\vartheta = (\vartheta_1, \ldots, \vartheta_k)^T \in \Theta \subset I\!R^k$. For reasons of identifiability we assume $F_\vartheta \neq F_{\vartheta'}$ whenever $\vartheta \neq \vartheta'$. If F denotes the error distribution in the model (1.1) we are interested in a test for the hypotheses

$$H_0 : F \in \mathcal{F} \quad \text{vs} \quad H_1 : F \notin \mathcal{F}. \tag{1.3}$$

Since the early work of Darling (1955) much effort has been spent to the problem of testing the parametric form of a distribution function [see for example Sukhatme (1972), Durbin (1973), Pierce and Kopecky (1979), Loynes (1980), Beran (1982), Koul and Lahiri (1994) or Koul (2002) among many others]. Most of the work concentrates on the linear regression model and compares the empirical distribution function $\widehat{F}_n$ of the residuals obtained from an appropriate estimator of the parameters of the mean with the function $F_{\widehat{\vartheta}}$, where $\widehat{\vartheta}$ is an estimator of ϑ usually also based on the residuals. However, the application of the proposed tests based on a norm of the process

$$\sqrt{n}\left(\widehat{F}_n(\cdot) - F_{\widehat{\vartheta}}(\cdot)\right) \tag{1.4}$$

is limited, because in general these test statistics are not asymptotically distribution-free under the null hypothesis H_0. The asymptotic distribution depends on the structure of the design and on the type of statistic used for the estimation of the parameters β and ϑ. Recently, in a remarkable paper, Khmaladze and Koul (2004) proposed martingale transformations of residual empirical processes in order to obtain an asymptotically distribution-free test for the hypotheses (1.3) in linear regression models. However, these transformations have to be calculated separately for each testing problem, which makes their direct implementation sometimes difficult.

One purpose of the present paper is to demonstrate that the residual based bootstrap provides a simple and reliable alternative for goodness-of-fit testing regarding the error distribution in linear models. Although the application of the bootstrap appears naturally in the present context, to the knowledge of the authors this topic has not found too much attention in the literature. The work most similar in spirit with the results of the present paper is the paper of Koul and Lahiri (1994), which discusses a smoothed bootstrap procedure for the linear model, where the class $\mathcal{F}$ consists of a singleton. The second purpose of the paper is to provide an omnibus test for the hypotheses (1.3) in the general nonparametric regression model (1.1). To the knowledge of the authors this problem has not been considered so far in the literature. A Pearson type statistic based on the two cells $(-\infty, 0)$, $(0, \infty)$ for the problem of testing for a standard normal distribution has been recently considered by Akritas and van Keilegom (2001). Obviously this test is only consistent against particular alternatives. In the second part of the paper we demonstrate that a bootstrap test based on an empirical process of the form (1.4) with nonparametric residuals yields a consistent test for the hypotheses (1.3).

The remaining part of this paper is organized as follows. In Section 2 we discuss some preliminary assumptions and properties of M-estimators in linear regression models. We introduce the empirical process $\widehat{F}_n$ based on the estimated residuals, compare this process with the distribution function $F_{\widehat{\vartheta}}$ obtained under the parametric assumption $F \in \mathcal{F}$ and establish weak convergence of the difference process (1.4). In order to obtain critical values for test statistics based on real valued functionals of this process we propose a parametric bootstrap procedure and prove its consistency. Section 3 deals with the problem of testing the hypotheses (1.3) in the general nonparametric regression model (1.1). In Section 4, we illustrate the finite sample performance of the bootstrap test by means of a simulation study considering an ANOVA, a linear regression and a nonparametric regression example. Our results indicate that tests based on empirical processes of residuals and the parametric bootstrap are a powerful tool for analyzing the error distribution in parametric and nonparametric regression models. These tests are asymptotically distribution-free, easily implementable in practice and have reasonable accuracy, even for moderate sample sizes. Finally, some of the proofs and more technical details are deferred to the Appendix.

2 Testing for a parametric class of error distributions in linear models

2.1 M-estimators in linear models

We begin our investigations with the classical linear model with a fixed design

$$Y_{ni} \;=\; x_{ni}^T\beta + \varepsilon_{ni} \quad (i = 1, \ldots, n) \tag{2.1}$$

where $\beta \in I\!R^p$ denotes an unknown parameter and the errors $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ are assumed to be independent identically distributed. Note that any linear regression model of the form (1.1) can be written in this form [see Pukelsheim (1993), chapter 1], and that the notation used in (2.1) allows us to treat parametric regression and ANOVA models simultaneously. Throughout this section let $X_n \in \mathbb{R}^{n\times p}$ denote the design matrix in the linear model (2.1), where the vector $x_{ni}^T = (x_{ni1}, \ldots, x_{nip})$ corresponds to the ith row of matrix X_n. We consider the model (2.1) under the following assumptions.

[**A 1**] The density f of the error distribution exists, is uniformly continuous, positive and we denote with F the corresponding distribution function.

[**A 2**] The design matrix $X_n \in I\!R^{n\times p}$ is not random, of rank $p \le n$ and satisfies the following regularity assumptions,

$$\max_{i=1,\ldots,n} x_{ni}^T (X_n^T X_n)^{-1} x_{ni} = O(\frac{1}{n}),$$
$$\lim_{n\to\infty} \frac{1}{n} X_n^T X_n = \Sigma \in I\!R^{p\times p},$$
$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n x_{ni} = m \in I\!R^p,$$

where the matrix Σ is assumed to be positive definite.

Throughout this section let $\widehat{\beta}_n$ denote an M-estimator for the parameter β. More precisely, let $\psi : I\!R \to I\!R$ denote a right continuous function with left limits, nondecreasing, with bounded total variation, such that $\int F\, d\psi < \infty$, $E[\psi^2(\varepsilon_{ni})] = \int \psi^2(t) f(t)\, dt < \infty$ and

$$E[\psi(\varepsilon_{ni})] = \int \psi\, dF \;=\; 0,$$

then the M-estimator $\widehat{\beta}_n$ of β is defined as a solution of the system of equations

$$X_n^T \begin{pmatrix} \psi(Y_{n1} - x_{n1}^T \widehat{\beta}_n) \\ \vdots \\ \psi(Y_{nn} - x_{nn}^T \widehat{\beta}_n) \end{pmatrix} = \sum_{i=1}^n x_{ni} \psi(Y_{ni} - x_{ni}^T \widehat{\beta}_n) = 0.$$

For example, with the choice $\psi(x) = x$ we obtain the least squares estimator

$$\widehat{\beta}_n \ = \ (X_n^T X_n)^{-1} X_n^T Y_n \ = \ \beta + (X_n^T X_n)^{-1} X_n^T \varepsilon_n$$

(with the notations $Y_n = (Y_{n1}, \dots, Y_{nn})^T, \varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nn})^T$) for which the above conditions can easily be verified if the second moment of the error distribution exists. A further important class of M-estimators is obtained by the function $\psi(x) = \text{sign}(x)$, which corresponds to the median [see e. g. Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. Other classes are the Huber estimators and quantiles [see e. g. van der Vaart (1998), p. 43]. The results of this paper are applicable for a broad class of M-estimators in linear models, but in the examples presented below we concentrate on the least squares and median case for the sake of brevity.

The important properties of M-estimators in linear models with fixed design are carefully described in the nice monograph of Koul (2002, p. 103) and briefly mentioned here for the sake of completeness. In particular we have under the assumptions stated above

$$||(X_n^T X_n)^{1/2} (\widehat{\beta}_n - \beta)|| \ = \ O_P(1), \tag{2.2}$$

and the stochastic expansion

$$(X_n^T X_n)^{1/2} (\widehat{\beta}_n - \beta) \ = \ \frac{1}{\int f \, d\psi} (X_n^T X_n)^{-1/2} \sum_{i=1}^n x_{ni} \psi(\varepsilon_{ni}) + o_P(1) \tag{2.3}$$

is valid. In order to investigate the properties of the error distribution we introduce the residuals

$$\widehat{\varepsilon}_{ni} \ = \ Y_{ni} - x_{ni}^T \widehat{\beta}_n \ = \ \varepsilon_{ni} - x_{ni}^T (\widehat{\beta}_n - \beta)$$

and the corresponding empirical distribution function

$$\widehat{F}_n(y) \ = \ \frac{1}{n} \sum_{i=1}^n I\{\widehat{\varepsilon}_{ni} \le y\}.$$

It is intuitively clear that – under the assumption that the linear model is correctly specified – this function is an approximation of the empirical distribution function of the unobservable errors,

$$F_n(y) \ = \ \frac{1}{n} \sum_{i=1}^n I\{\varepsilon_{ni} \le y\}.$$

This statement has been made precise in Theorem 6.2.1 in Koul (2002, p. 232).

2.2 The testing procedure and its asymptotic analysis

The special case, where $\mathcal{F}$ is a scale family ($k = 1$)

$$\mathcal{F} = \{F(\cdot/\vartheta) \mid \vartheta > 0\}$$

with a known distribution function F is certainly the most important situation from a practical point of view and several test statistics have been proposed by Koul (2002) for this situation. As can be seen in Theorem 2.1 below the asymptotic distribution of the process (1.4) depends in a complicated manner on the structure of the design and on the type of statistic used for the estimation of the parameters β and ϑ. In the following we will demonstrate that a residual based parametric bootstrap provides a simple and reliable solution of these difficulties, especially for moderate sample sizes. For our theoretical investigations we require several additional regularity assumptions under the null hypothesis $H_0 : F \in \mathcal{F}$, where $\mathcal{F}$ is defined in (1.2). Let $E_\vartheta[\cdot]$, $\mathrm{Var}_\vartheta(\cdot)$ denote the expectation resp. variance under the distribution associated with F_ϑ.

[**A 3**] The density $f_\vartheta(y)$ is uniformly continuous with respect to y and continuous with respect to ϑ.

[**A 4**] The function ψ defining the M-estimator $\widehat{\beta}_n$ satisfies

$$\begin{aligned} E_\vartheta[\psi(\varepsilon_{ni})] &= \int \psi \, dF_\vartheta = 0 \text{ for all } \vartheta \in \Theta, \\ E_\vartheta[\psi^2(\varepsilon_{ni})] &= \int \psi^2 \, dF_\vartheta < \infty \text{ for all } \vartheta \in \Theta \ . \end{aligned}$$

[**A 5**] Each $F_\vartheta \in \mathcal{F}$ is continuously differentiable with respect to ϑ_i ($i = 1, \ldots, k$) such that for all $\vartheta \in \Theta$ the partial derivatives

$$\frac{\partial F_\vartheta(y)}{\partial \vartheta_i}$$

are bounded and uniformly continuous with respect to y.

Throughout this section let $\widehat{\vartheta} = (\widehat{\vartheta}_1, \ldots, \widehat{\vartheta}_k)^T$ denote an estimator for the parameter $\vartheta \in \Theta$ (usually based on the residuals $\widehat{\varepsilon}_{n1}, \ldots, \widehat{\varepsilon}_{nn}$ from the parametric fit).

[**A 6**] We assume that the following linear expansion of $\widehat{\vartheta}_j$ ($j = 1, \ldots, k$) with respect to the residuals $\widehat{\varepsilon}_{n1}, \ldots, \widehat{\varepsilon}_{nn}$, is valid,

$$\begin{aligned} \widehat{\vartheta}_j - \vartheta_j &= \frac{1}{n} \sum_{i=1}^n h_{\vartheta_j}(\widehat{\varepsilon}_{ni}) + o_P(\frac{1}{\sqrt{n}}) \\ &= \frac{1}{n} \sum_{i=1}^n h_{\vartheta_j}(\varepsilon_{ni}) - \frac{1}{n} \sum_{i=1}^n h'_{\vartheta_j}(\varepsilon_{ni}) x_{ni}^T (\widehat{\beta}_n - \beta) + o_P(\frac{1}{\sqrt{n}}), \end{aligned} \quad (2.4)$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_k)^T$ denotes the 'true parameter'. Here h_{ϑ_j} is a continuously differentiable 'influence' function such that

$$\begin{aligned} E_\vartheta[h_{\vartheta_j}(\varepsilon_{ni})] &= \int h_{\vartheta_j}(t) f_\vartheta(t)\,dt = 0, \\ E_\vartheta[h^2_{\vartheta_j}(\varepsilon_{ni})] &= \int h^2_{\vartheta_j}(t) f_\vartheta(t)\,dt < \infty \end{aligned}$$

for all $\vartheta \in \Theta$, where the functions $\vartheta \mapsto h_\vartheta(\cdot)$ and $\vartheta \mapsto h'_\vartheta(\cdot)$ are assumed to be continuous.

We remark that if the estimator $\widehat{\vartheta}$ is efficient, then an expansion of the form (2.4) is typically satisfied, where $(h_{\vartheta_1}, \ldots, h_{\vartheta_k})^T$ is the score vector times the inverse information matrix [see van der Vaart (1998)]. The proof of the following theorem is deferred to the Appendix.

Theorem 2.1 *Assume model (2.1) under the assumptions [A 1]–[A 6]. Under the null hypothesis $H_0 : F = F_\vartheta$ for some $\vartheta \in \Theta$ the estimated empirical process*

$$\sqrt{n}\left(\widehat{F}_n(F_\vartheta^{-1}(t))) - F_{\widehat{\vartheta}}(F_\vartheta^{-1}(t))\right); \quad t \in [0,1],$$

converges weakly to a Gaussian process G with covariance kernel

$$\begin{aligned} \mathrm{Cov}_\vartheta(G(s), G(t)) &= s \wedge t - st + \sum_{j=1}^k \sum_{\ell=1}^k \frac{\partial F_\vartheta}{\partial \vartheta_j}(F_\vartheta^{-1}(s)) \frac{\partial F_\vartheta}{\partial \vartheta_\ell}(F_\vartheta^{-1}(t)) \int h_{\vartheta_j}(y) h_{\vartheta_\ell}(y) f_\vartheta(y)\,dy \\ &+ m^T \Sigma^{-1} m \int \psi^2(y) f_\vartheta(y)\,dy\, c_\vartheta(F_\vartheta^{-1}(s)) c_\vartheta(F_\vartheta^{-1}(t)) \\ &- \sum_{j=1}^k \frac{\partial F_\vartheta}{\partial \vartheta_j}(F_\vartheta^{-1}(t)) \int_0^s h_{\vartheta_j}(F_\vartheta^{-1}(x))\,dx \\ &- \sum_{j=1}^k \frac{\partial F_\vartheta}{\partial \vartheta_j}(F_\vartheta^{-1}(s)) \int_0^t h_{\vartheta_j}(F_\vartheta^{-1}(x))\,dx \\ &+ m^T \Sigma^{-1} m \Big(c_\vartheta(F_\vartheta^{-1}(t)) \int_0^s \psi(F_\vartheta^{-1}(x))\,dx + c_\vartheta(F_\vartheta^{-1}(s)) \int_0^t \psi(F_\vartheta^{-1}(x))\,dx \Big) \\ &- m^T \Sigma^{-1} m \sum_{j=1}^k \Big(c_\vartheta(F_\vartheta^{-1}(t)) \frac{\partial F_\vartheta}{\partial \vartheta_j}(F_\vartheta^{-1}(s)) + c_\vartheta(F_\vartheta^{-1}(s)) \frac{\partial F_\vartheta}{\partial \vartheta_j}(F_\vartheta^{-1}(t)) \Big) \\ &\times \int h_{\vartheta_j}(x) \psi(x) f_\vartheta(x)\,dx, \end{aligned}$$

where the matrix Σ and the vector m are defined in assumption [A 2] and

$$c_\vartheta(y) = \frac{1}{\int f_\vartheta\,d\psi} \Big(f_\vartheta(y) + \sum_{j=1}^k \frac{\partial F_\vartheta(y)}{\partial \vartheta_j} \int h'_{\vartheta_j}(t) f_\vartheta(t)\,dt \Big), \quad y \in I\!R.$$

For a construction of a test for the hypotheses (1.3) we propose to compare the estimators of the distribution function of the errors under the null hypothesis and alternative. The null hypothesis is rejected for large values of a continuous functional of the process

$$\widehat{S}_n(y) \;=\; \sqrt{n}(\widehat{F}_n - F_{\widehat{\vartheta}})(y), \quad y \in I\!R, \tag{2.5}$$

where $\widehat{\vartheta}$ is an appropriate estimate for the parameter ϑ satisfying the assumption [A 6]. Consider, for instance, the Kolmogorov-Smirnov statistic

$$T_n \;=\; \sup_{y\in I\!R} |\widehat{S}_n(y)| \;=\; \sup_{t\in[0,1]} \left| \sqrt{n}\left(\widehat{F}_n(F_{\vartheta}^{-1}(t)) - F_{\widehat{\vartheta}}(F_{\vartheta}^{-1}(t)) \right) \right| \tag{2.6}$$

or the Cramér-von Mises statistic

$$Y_n \;=\; \int \widehat{S}_n^2(x) d\widehat{F}_n(x) \;=\; \int \left[\sqrt{n}\left(\widehat{F}_n(F_{\vartheta}^{-1}(t)) - F_{\widehat{\vartheta}}(F_{\vartheta}^{-1}(t)) \right) \right]^2 dt + o_P(1).$$

Theorem 2.1 and the application of the Continuous Mapping Theorem specify the asymptotic behaviour of these statistics under the null hypothesis $F = F_\vartheta$. The limiting distributions depend in a complicated way on the design, the parametric family $\mathcal{F}$ and the function ψ used for the M-estimation, and as a consequence tests based on the empirical process (2.5) are difficult to implement in practice. For this reason we investigate the performance of the parametric bootstrap in this problem.

2.3 A parametric bootstrap procedure

Throughout this section let $\mathcal{Y}_n = \{Y_{n1}, \ldots, Y_{nn}\}$ denote the sample of observations and $\widehat{\vartheta}$ an estimator for the parameter ϑ based on $\mathcal{Y}_n$. For a sample of independent identically uniformly distributed random variables $U_1, \ldots, U_n \sim U[0,1]$ we define

$$\varepsilon_{ni}^* = F_{\widehat{\vartheta}}^{-1}(U_i) \sim F_{\widehat{\vartheta}}, \quad i = 1, \ldots, n,$$

and bootstrap observations

$$Y_{ni}^* \;=\; x_{ni}^T \widehat{\beta}_n + \varepsilon_{ni}^* \quad (i = 1, \ldots, n),$$

where $\widehat{\beta}_n$ is an M-estimator for the parameter β based on the sample $\mathcal{Y}_n$. Let $\widehat{\beta}_n^*$ denote the analogue of $\widehat{\beta}_n$, based on bootstrap observations, defined by the equation

$$\sum_{i=1}^n x_{ni}\psi(Y_{ni}^* - x_{ni}^T\widehat{\beta}_n^*) \;=\; 0, \tag{2.7}$$

then the bootstrap residuals are given by

$$\widehat{\varepsilon}_{ni}^* \;=\; Y_{ni}^* - x_{ni}^T\widehat{\beta}_n^* \;=\; \varepsilon_{ni}^* - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n).$$

The analogues of the empirical distribution functions $\widehat{F}_n$ and F_n in the bootstrap setting are the processes

$$\begin{aligned} \widehat{F}_n^*(y) &= \frac{1}{n}\sum_{i=1}^n I\{\widehat{\varepsilon}_{ni}^* \le y\}, \\ F_n^*(y) &= \frac{1}{n}\sum_{i=1}^n I\{\varepsilon_{ni}^* \le y\}, \end{aligned}$$

respectively. In the following let $\widehat{\vartheta}^* = (\widehat{\vartheta}_1^*, \dots, \widehat{\vartheta}_k^*)^T$ denote the estimator analogous to $\widehat{\vartheta} = (\widehat{\vartheta}_1, \dots, \widehat{\vartheta}_k)^T$ based on the bootstrap observations Y_{ni}^*, $i = 1, \dots, n$. Some regularity conditions are required to establish consistency of this bootstrap approach.

[**A 7**] The M-estimate $\widehat{\beta}_n^*$ in the bootstrap setting has the following stochastic expansions,

$$||(X_n^T X_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n)|| = O_P(1), \tag{2.8}$$

$$(X_n^T X_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n) = \frac{1}{\int f_{\widehat{\vartheta}}\, d\psi}(X_n^T X_n)^{-1/2}\sum_{i=1}^n x_{ni}\psi(\varepsilon_{ni}^*) + o_{P_n^*}(1) \tag{2.9}$$

in probability, where the notation $Z_n = o_{P_n^*}(1)$ in probability means that for all $\delta > 0$ $P(|Z_n| > \delta \mid \mathcal{Y}_n) = o_P(1)$.

[**A 8**] We also assume convergence in probability of the estimator $\widehat{\vartheta}$ to some value $\vartheta_0 \in \Theta$, where

$$\vartheta_0 = \arg\inf_{\vartheta\in\Theta} d(F, F_\vartheta)$$

for some metric d. Under the null hypothesis $H_0 : F \in \mathcal{F}$, ϑ_0 is the true parameter, while under the alternative ϑ_0 corresponds to the best approximation F_{ϑ_0} of F in the class $\mathcal{F}$ with respect to the metric d.

[**A 9**] We assume the following uniform convergence,

$$\sup_{x\in I\!R} |f_{\widehat{\vartheta}}(x) - f_{\vartheta_0}(x)| = o_P(1),$$

$$\sup_{x\in I\!R}\Big|\frac{\partial F_\vartheta(x)}{\partial\vartheta}|_{\vartheta=\widehat{\vartheta}} - \frac{\partial F_\vartheta(x)}{\partial\vartheta}|_{\vartheta=\vartheta_0}\Big| = o_P(1).$$

[**A 10**] For the estimator $\widehat{\vartheta}^* = (\widehat{\vartheta}_1^*, \dots, \widehat{\vartheta}_k^*)^T$ from the bootstrap sample $Y_{n1}^*, \dots, Y_{nn}^*$ we assume that the stochastic expansion

$$\widehat{\vartheta}_j^* - \widehat{\vartheta}_j = \frac{1}{n}\sum_{i=1}^n h_{\widehat{\vartheta}_j}(\varepsilon_{ni}^*) - \frac{1}{n}\sum_{i=1}^n E[h'_{\widehat{\vartheta}_j}(\varepsilon_{ni}^*) \mid \mathcal{Y}_n] x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) + o_{P_n^*}(\frac{1}{\sqrt{n}})$$

is valid in probability (see also assumption [A 6] and Proposition A.1 in the Appendix), where

$$E[h_{\widehat{\vartheta}_j}(\varepsilon_{ni}^*) \mid \mathcal{Y}_n] = \int h_{\widehat{\vartheta}_j}(t) f_{\widehat{\vartheta}}(t)\, dt = 0.$$

Remark 2.2 **(a)** Condition (2.9) is obviously fulfilled for least squares estimators, where

$$\widehat{\beta}_n^* - \widehat{\beta}_n = (X_n^T X_n)^{-1} X_n^T Y_n^* - \widehat{\beta}_n = (X_n^T X_n)^{-1} \sum_{i=1}^n x_{ni} \varepsilon_{ni}^*.$$

Condition (2.8) can then be concluded under the additional assumption $E[\int x^2 f_{\widehat{\vartheta}}(x) dx] = O(1)$.

(b) If ψ is differentiable with Lipschitz continuous derivative, assumption (2.9) can be deduced analogously to Proposition 2.1 in Koul and Lahiri (1994).

(c) In definition (2.7) of the M-estimator in the bootstrap setting it is not necessary to center like in Koul and Lahiri (1994) because

$$E[\psi(\varepsilon_{ni}^*) \mid \mathcal{Y}_n] \;=\; \int \psi(x) f_{\widehat{\vartheta}}(x)\, dx \;=\; 0.$$

Theorem 2.3 *Assume model (2.1) under assumptions [A 1]–[A 10]. Under the null hypothesis and under fixed alternatives the process*

$$\sqrt{n}\left(\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t)\right); \quad t \in [0,1],$$

conditionally on $\mathcal{Y}_n$, converges weakly to a Gaussian process in probability. The asymptotic covariance structure is the same as defined in Theorem 2.1, where the parameter ϑ has to be replaced by the value ϑ_0 defined in assumption [A 8].

The proof of Theorem 2.3 is deferred to the Appendix. From the theorem the consistency of a bootstrap test for the hypothesis (1.3) in the linear model (2.1) can be deduced as follows. Let T_n denote the test statistic based on a continuous functional of the process $\widehat{S}_n$ defined in (2.5) and let T_n^* denote the corresponding bootstrap statistic based on

$$\widehat{S}_n^*(y) \;=\; \sqrt{n}(\widehat{F}_n^*(y) - F_{\widehat{\vartheta}^*}(y)), \quad y \in I\!R.$$

Consider, for instance, the Kolmogorov–Smirnov statistic

$$T_n^* \;=\; \sup_{y \in I\!R} |\widehat{S}_n^*(y)| \;=\; \sup_{t\in[0,1]} |\sqrt{n}(\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t)))|.$$

If t_n is the realization of the test statistic T_n defined in (2.6) based on the sample $\mathcal{Y}_n$ then a level α–test is obtained by rejecting the null hypothesis whenever $t_n > c_{1-\alpha}$, where $P_{H_0}(T_n > c_{1-\alpha}) = \alpha$. The quantile $c_{1-\alpha}$ can now be approximated by the bootstrap quantile $c_{1-\alpha}^*$ defined by

$$P(T_n^* > c_{1-\alpha}^* \mid \mathcal{Y}_n) \;=\; \alpha.$$

From Theorem 2.3 and the Continuous Mapping Theorem we obtain a consistent asymptotic level α–test by rejecting the null hypothesis if $t_n > c_{1-\alpha}^*$. We will illustrate the finite sample properties of this approach by means of a simulation study in Section 4.

Remark 2.4 Note that we restricted our investigations to the linear model with a fixed design, which has mainly technical reasons (most of the results of Koul (2002) were formulated for this design). Under certain changes in the assumptions and in the derivation of the expansion of the process $\widehat{F}_n$ in the proof of Theorem 2.1 the testing procedure can be generalized to the random design case in linear regression models as well.

3 Goodness-of-fit testing in nonparametric regression

We now consider the problem of testing parametric hypotheses regarding the error distribution in nonparametric regression models. As pointed out in the introduction additional knowledge on the error distribution can improve the efficiency of the statistical analysis and can lead to optimal procedures from an asymptotic point of view. Again the fixed and random design have to be treated separately and for the sake of brevity and transparency we restrict ourselves to the case of a random design, for which some of the technical arguments are slightly simpler. Throughout this section we assume that independent observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ from the nonparametric regression model

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

are available. We are interested in testing the hypothesis $H_0 : F \in \mathcal{F}$, where the parametric class $\mathcal{F}$ of distribution functions is defined in (1.2). In order to derive bootstrap tests in the general nonparametric regression model and to prove consistency we require the following assumptions.

[A 11] The design points $X_1, \ldots, X_n$ are independent identically distributed random variables, independent from the errors, with density f_X, which has compact support, say $[0, 1]$, such that f_X is bounded away from zero. The design density f_X and the regression function m are assumed to be twice continuously differentiable on the interval $[0, 1]$.

[A 12] The random variables $\varepsilon_1, \ldots, \varepsilon_n$ are independent with distribution function F such that $E[\varepsilon_1] = 0$, $\mathrm{Var}(\varepsilon_1) < \infty$.

[A 13] (Compare with assumption [A 2]) The distribution function F of the errors is twice continuously differentiable with uniformly continuous density f, such that f' is bounded (for all $F \in \mathcal{F}$).

[A 14] Let K be a symmetric twice continuously differentiable density with compact support and vanishing first moment $\int uK(u)\,du = 0$ and $h = h_n$ denote a sequence of bandwidths converging to zero for an increasing sample size $n \to \infty$ such that $nh^4 = o(1)$ and $nh^{3+\delta}/\log(1/h) \to \infty$ for some $\delta > 0$.

For the estimation of the residuals nonparametrically we use the well known Nadaraya–Watson kernel estimator for the unknown regression function $m(\cdot)$, which is defined by

$$\widehat{m}(x) \;=\; \frac{\sum_{i=1}^{n} K(\frac{X_i-x}{h})Y_i}{\sum_{j=1}^{n} K(\frac{X_j-x}{h})}\;. \tag{3.2}$$

The residuals from the nonparametric fit are given by

$$\widehat{\varepsilon}_{ni} \;=\; Y_i - \widehat{m}(X_i) \quad (i=1,\ldots,n).$$

The choice of this estimator is mainly motivated to keep the technical arguments simple and the results can easily be extended to more advanced kernel estimators as the Gasser-Müller or local polynomial estimator. Let $\widehat{F}_n$ and F_n denote the empirical distribution functions based on the nonparametric residuals $\widehat{\varepsilon}_{n1},\ldots,\widehat{\varepsilon}_{nn}$ and on true errors $\varepsilon_1,\ldots,\varepsilon_n$, respectively. The weak convergence of the process $\sqrt{n}(\widehat{F}_n - F)$ in the nonparametric setting was established by Akritas and van Keilegom (2001). Let $\widehat{\vartheta} = (\widehat{\vartheta}_1,\ldots,\widehat{\vartheta}_k)^T$ denote an estimator for the parameter $\vartheta \in \Theta$ based on the residuals $\widehat{\varepsilon}_{n1},\ldots,\widehat{\varepsilon}_{nn}$.

[A 15] (Compare with assumption [A 6]) We assume that the following stochastic expansion for the estimates $\widehat{\vartheta}_j$ $(j=1,\ldots,k)$ with respect to the errors is valid,

$$\widehat{\vartheta}_j - \vartheta_j \;=\; \frac{1}{n}\sum_{i=1}^{n} h_{\vartheta_j}(\varepsilon_i) - \frac{1}{n}\sum_{i=1}^{n} h'_{\vartheta_j}(\varepsilon_i)(\widehat{m}(X_i) - m(X_i)) + o_P(\frac{1}{\sqrt{n}}),$$

where h_{ϑ_j} is continuously differentiable such that $E_\vartheta[h_{\vartheta_j}(\varepsilon_i)] = 0$, $E_\vartheta[h^2_{\vartheta_j}(\varepsilon_i)] < \infty$ for all $\vartheta \in \Theta$. Moreover, the functions $\vartheta \mapsto h_\vartheta(\cdot)$ and $\vartheta \mapsto h'_\vartheta(\cdot)$ are assumed to be continuous.

Theorem 3.1 *Assume model (3.1) under the assumptions [A 3], [A 5] and [A 11]–[A 15]. Under the null hypothesis $H_0 : F = F_\vartheta$ for some $\vartheta \in \Theta$ the estimated empirical process*

$$\sqrt{n}\left(\widehat{F}_n(F_\vartheta^{-1}(t)) - F_{\widehat{\vartheta}}(F_\vartheta^{-1}(t))\right); \quad t \in [0,1]$$

converges weakly to a Gaussian process G with covariance kernel $\mathrm{Cov}_\vartheta(G(s),G(t))$ from Theorem 2.1, where $m^T\Sigma^{-1}m$ has to be replaced by 1 and $\psi(x) = x$.

For a bootstrap procedure based on a sample $\mathcal{Y}_n = \{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ and independent identically uniformly distributed random variables $U_1,\ldots,U_n \sim U[0,1]$ we consider random variables $\varepsilon^*_{ni} = F^{-1}_{\widehat{\vartheta}}(U_i)$ $(i=1,\ldots,n)$ with distribution function $F_{\widehat{\vartheta}}$. Analogous to the linear case bootstrap observations are then defined by

$$Y^*_{ni} \;=\; \widehat{m}(X_i) + \varepsilon^*_{ni} \quad (i=1,\ldots,n).$$

Let $\widehat{m}^*$ denote the regression estimate (3.2) based on the bootstrap sample $(X_1, Y^*_{n1}), \ldots, (X_n, Y^*_{nn})$, then the bootstrap residuals are given by

$$\widehat{\varepsilon}^*_{ni} \; = \; Y^*_{ni} - \widehat{m}^*(X_i).$$

Finally we denote the empirical distribution functions based on bootstrap residuals $\widehat{\varepsilon}^*_{n1}, \ldots, \widehat{\varepsilon}^*_{nn}$ and bootstrap errors $\varepsilon^*_{n1}, \ldots, \varepsilon^*_{nn}$ by $\widehat{F}^*_n$ and F^*_n, respectively. In order to prove consistency of the parametric bootstrap in the nonparametric regression model we require one more assumption regarding a stochastic expansion of the bootstrap estimate under the null hypothesis.

[A 16] (Compare with assumption [A 10]) For the estimator $\widehat{\vartheta}^* = (\widehat{\vartheta}^*_1, \ldots, \widehat{\vartheta}^*_k)^T$ based on the bootstrap sample $(X_1, Y^*_{n1}), \ldots, (X_n, Y^*_{nn})$ we assume that the stochastic expansion

$$\widehat{\vartheta}^*_j - \widehat{\vartheta}_j \; = \; \frac{1}{n}\sum_{i=1}^{n} h_{\widehat{\vartheta}_j}(\varepsilon^*_{ni}) - \frac{1}{n}\sum_{i=1}^{n} E[h'_{\widehat{\vartheta}_j}(\varepsilon^*_{n1}) \mid \mathcal{Y}_n]\varepsilon^*_{ni} + o_{P^*_n}(\frac{1}{\sqrt{n}})$$

is valid in probability, where $E[h_{\widehat{\vartheta}_j}(\varepsilon^*_{n1}) \mid \mathcal{Y}_n] = 0$.

Theorem 3.2 *Assume model (3.1) under the assumptions [A 1], [A 5], [A 8], [A 9], [A 11]–[A 16]. Under the null hypothesis and under fixed alternatives the process*

$$\sqrt{n}\left(\widehat{F}^*_n(F^{-1}_{\widehat{\vartheta}}(t)) - F_{\widehat{\vartheta}^*}(F^{-1}_{\widehat{\vartheta}}(t))\right); \quad t \in [0,1],$$

conditionally on $\mathcal{Y}_n$, converges weakly to a Gaussian process in probability. The asymptotic covariance structure is the same as defined in Theorem 3.1, where the parameter ϑ has to be replaced by the value ϑ_0 defined in assumption [A 8].

4 Finite sample properties

In this section we investigate the finite sample properties of the bootstrap test proposed in Sections 2 and 3. We consider the classical one-way-layout in analysis of variance, a quadratic regression and a nonparametric regression model on the interval $[0,1]$. If T_n is a continuous functional of the empirical proces (2.5), then the null hypothesis is rejected at level α if

$$T_n > T^*_{n(\lfloor B(1-\alpha)\rfloor)}, \tag{4.1}$$

where $T^*_{n(1)} < \ldots < T^*_{n(B)}$ denote the order statistics of the bootstrap sample $T_n^{*(1)}, \ldots, T_n^{*(B)}$, where the bootstrap is described in Section 2 and 3, respectively. We use $B = 200$ replications for the resampling procedure and 2000 simulation runs for the estimation of the rejection probabilities.

4.1 Testing for normality in a regression model

In this example we investigate the finite sample performance of the bootstrap procedure for testing for a normal error distribution in a homoscedastic quadratic regression model. The model under consideration is

$$Y_{ni} = \beta_1 x_{ni} + \beta_2 x_{ni}^2 + \varepsilon_{ni}, \quad i = 1, \dots, n, \tag{4.2}$$

with $x_{ni} = i/n$ ($i = 1, \dots, n$) corresponding to the uniform design. The hypothesis of interest is

$$H_0 : F(x) = \Phi\Big(\frac{x}{\sigma}\Big) \tag{4.3}$$

for some $\sigma > 0$, where Φ denotes the cumulative distribution function of the standard normal distribution, and F is the distribution function of the random variables ε_{ni}. The parameters β_1 and β_2 are estimated by least squares technique, while the scale factor σ is estimated by

$$\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{j=1}^{n} \widehat{\varepsilon}_{nj}^2, \tag{4.4}$$

where $\widehat{\varepsilon}_{nj} = Y_{nj} - \widehat{\beta}_1 x_{nj} - \widehat{\beta}_2 x_{nj}^2$ denote the residuals from the least squares fit. The test statistic is given by the Kolmogorov-Smirnov statistic

$$T_n = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - \Phi\Big(\frac{x}{\widehat{\sigma}}\Big)|, \tag{4.5}$$

and the null hypothesis is rejected if the inequality (4.1) is satisfied. In the simulation the parameters are chosen as $\beta_1 = \beta_2 = 1, \sigma^2 = 0.5$ and for the error distribution five cases are considered, that is

$$\begin{aligned} \varepsilon_{ni} &\sim \mathcal{N}(0, 0.5) \quad \text{(null hypothesis)}, \\ \varepsilon_{ni} &\sim (\mathcal{X}_k^2 - k)/\sqrt{4k}, \quad k = 1, 2, 3, 4. \end{aligned} \tag{4.6}$$

n	25			50			75			100		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.024	0.044	0.087	0.027	0.054	0.105	0.031	0.050	0.094	0.030	0.051	0.096
df_1	0.719	0.800	0.886	0.982	0.994	0.999	1.000	1.000	1.000	1.000	1.000	1.000
df_2	0.410	0.507	0.615	0.845	0.906	0.951	0.980	0.992	0.998	0.997	0.998	0.999
df_3	0.281	0.371	0.485	0.673	0.767	0.851	0.871	0.924	0.963	0.965	0.984	0.995
df_4	0.228	0.317	0.412	0.507	0.616	0.732	0.740	0.828	0.904	0.888	0.941	0.969

Table 4.1: *Simulated rejection probabilities of the test (4.1) for various error distributions. The model is given by (4.2) and the error distribution is normal (df_0) and chi-square with k degrees of freedom (df_k) such that $E[\varepsilon_{ni}] = 0$ and $E[\varepsilon_{ni}^2] = 0.5$.*

The results are depicted in Table 4.1 for the sample sizes $n = 25, 50, 75, 100$ and levels $\alpha =$ 2.5%, 5% and 10 %. We observe a good approximation of the nominal even if the sample size is $n = 25$. The alternatives are detected with reasonable probabilities, which increase with the sample size and decrease with increasing degrees of freedom.

4.2 Testing for normality in ANOVA models

In this example we investigate the properties of the bootstrap procedure in the classical ANOVA model of a one-way-layout, that is

$$Y_{nij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, \ldots, p; j = 1, \ldots, n_i. \tag{4.7}$$

Classical theory of analysis of variance assumes a normal distribution for the iid errors, that is

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, p; j = 1, \ldots, n_i.$$

We have applied the bootstrap test with the statistic

$$T_n = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - \Phi(\frac{x}{\widehat{\sigma}})|,$$

where

$$\widehat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{nij} - \bar{Y}_{ni\cdot})^2$$

is the common least squares estimate of the variance. We considered the case of $p = 4$ factors and $n = 20$ and $n = 40$ total observations, where the sample sizes for the different groups are

$$\begin{aligned} A &: n_1 = n_2 = n_3 = n_4 = \frac{n}{4}; \\ B &: n_1 = \frac{3n}{20}; n_2 = \frac{n}{4}; n_3 = \frac{n}{4}; n_4 = \frac{7n}{20}; \end{aligned} \tag{4.8}$$

the variance is given by $\sigma^2 = 0.5$ and the means are $\mu_i = i; i = 1, 2, 3, 4$. The rejection probabilities of the bootstrap test (4.1) are displayed in Table 4.2 for the null hypothesis of a normal distribution and the four alternatives presented in (4.6).

We observe a reasonable approximation of the nominal level for the sample size $n = 40$, while for the sample size $n = 20$ the bootstrap test is conservative. Deviations from the null hypothesis are detected with a probability which decreases with increasing degrees of freedom of the chi-square distribution. It is interesting to note that for the sample size $n = 40$ the unbalanced design B yields a slightly better approximation of the nominal level than the balanced design A. However, for more unbalanced designs the approximation of the nominal level is slightly less accurate compared to the balanced design [these results are not depicted for the sake of brevity].

	$n = 20$						$n = 40$					
	A			B			A			B		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.018	0.037	0.074	0.015	0.038	0.069	0.021	0.045	0.091	0.024	0.049	0.092
df_1	0.431	0.525	0.645	0.432	0.532	0.643	0.922	0.953	0.974	0.919	0.948	0.975
df_2	0.210	0.280	0.370	0.190	0.264	0.365	0.663	0.775	0.843	0.640	0.735	0.832
df_3	0.122	0.173	0.262	0.134	0.193	0.274	0.472	0.564	0.670	0.462	0.565	0.681
df_4	0.094	0.140	0.207	0.097	0.145	0.214	0.345	0.438	0.553	0.361	0.460	0.578

Table 4.2: *Simulated rejection probabilities of the bootstrap test (4.1) for normality in the one-way-layout (4.7) with four factors for various distributions defined in (4.6) and designs given in (4.8).*

4.3 Different M-estimators

In this example we briefly investigate the effect of the function ψ on the performance of the bootstrap tests for the error distribution. We consider the location-scale model

$$Y_{ni} = \mu + \varepsilon_{ni}, \quad i = 1, \ldots, n, \tag{4.9}$$

where $\mu \in \mathbb{R}, E[\varepsilon_{ni}] = 0$ and $E[\varepsilon_{ni}^2] = \sigma^2$. For the function ψ two cases are considered, that is

$$\psi(x) = x; \quad \psi(x) = \operatorname{sign}(x);$$

and the corresponding M-estimators $\widehat{\mu}$ are the sample mean and median, respectively. The variance of the error distribution has been estimated by the residual sum of squares in both cases, that is

$$\widehat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_{ni} - \widehat{\mu})^2.$$

We consider the performance of the test (4.1) for a normal distribution based on the Kolmogorov-Smirnov statistic, where the resampling scheme has been described in Section 3. The rejection probabilities of the bootstrap test are presented in Table 4.3 for sample sizes $n = 25, 50$ and the error distributions given in (4.6). The results are very similar, where the function $\psi(x) = \operatorname{sign}(x)$ yields better results for larger deviations from the null hypothesis (df_1, df_2). However, if the alternative is closer to the null (df_3, df_4) the test based on the least squares estimator is more powerful. This corresponds to intuition, because it is known that for a normal distribution the M-estimator corresponding to the function $\psi(x) = -\frac{f'(x)}{f(x)} = x$ is the best M-estimator [see Serfling (1980)].

$\psi(x) = x$							$\psi(x) = \text{sign}(x)$					
	n=25			n=50			n=25			n=50		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.029	0.057	0.107	0.029	0.058	0.100	0.026	0.052	0.101	0.029	0.057	0.107
df_1	0.921	0.955	0.975	0.999	1.000	1.000	0.965	0.978	0.985	1.000	1.000	1.000
df_2	0.590	0.687	0.802	0.929	0.963	0.985	0.683	0.747	0.814	0.964	0.974	0.984
df_3	0.427	0.532	0.654	0.740	0.822	0.897	0.427	0.507	0.601	0.784	0.833	0.887
df_4	0.316	0.418	0.540	0.585	0.697	0.810	0.297	0.368	0.473	0.559	0.639	0.730

Table 4.3. *Simulated rejected probabilities of the test (4.1) for a normal distribution in the model (4.9) and two different M-estimators corresponding to the functions $\psi(x) = x$ and $\psi(x) = \text{sign}(x)$. The error distributions are given by (4.6), where df_0 corresponds to a normal distribution and df_k $(k = 1, 2, 3, 4)$ to a $\mathcal{X}_k^2$-distribution standardized such that $E[\varepsilon_{ni}] = 0$ and $E[\varepsilon_{ni}^2] = 0.5$.*

4.4 Nonparametric regression

Our final example considers the problem of testing for a normal distribution (4.3) in a nonparametric regression model. For this we simulated data according to the regression model (4.2), where the variance of the error is given by 0.5. Note that we consider a fixed design in order to compare the results with the situation described in Section 4.1. The results for the random design are very similar and for this reason not displayed. We have applied the bootstrap test based on the Cramér-von Mises statistic

$$Y_n = \int \Big\{ \widehat{F}_n(x) - \Phi\Big(\frac{x}{\widetilde{\sigma}}\Big) \Big\}^2 d\widehat{F}_n(x), \tag{4.10}$$

and the Kolmogorov-Smirnov statistic (4.5), where $\widehat{F}_n$ is the empirical distribution function from the nonparametric residuals $\widehat{\varepsilon}_{ni} = Y_i - \widehat{m}(X_i)$ and $\widetilde{\sigma}^2$ the common estimator of the variance in the nonparametric regression model (1.1), that is

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_{ni}^2.$$

For the estimation of the regression function we used a local linear estimator [see Fan and Gijbels (1996)] because of its better performance at the boundary of the design space. The results are shown in Tabel 4.4 and 4.5 corresponding to the L^2- and sup-norm. The bandwidth is chosen as

$$\widehat{h} = \Big(\frac{\widehat{\sigma}^2}{n}\Big)^{1/5}$$

where $\widehat{\sigma}^2$ is the nonparametric estimator of Rice (1984). We observe a reasonable approximation of the nominal level in all cases under consideration. The Cramér-von Mises test usually yields

larger rejection probabilities than the Kolmogorov-Smirnov test, but in all cases the deviation from normality is detected with reasonable probabilities.
It is also of interest to compare the results in Table 4.5 with the corresponding values in Table 4.1, where the hypothesis of normality has been tested under the additional assumption of a linear regression model. The tests using the knowledge of the parametric form of the regression function yield a better approximation of the nominal level than the purely nonparametric procedures. As expected the tests of normality of this section are usually less powerful than the tests constructed in Section 4.1, which use the specific information of the structure of the regression function. Nevertheless, the results of this section demonstrate, that the bootstrap tests based on empirical processes yield also reliable procedures for goodness-of-fit testing in nonparametric regression models.

n	25			50			75			100		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.037	0.052	0.103	0.033	0.055	0.103	0.038	0.057	0.101	0.035	0.049	0.099
df_1	0.752	0.808	0.885	0.991	0.993	0.997	1.000	1.000	1.000	1.000	1.000	1.000
df_2	0.547	0.631	0.734	0.947	0.961	0.983	0.995	0.996	0.998	1.000	1.000	1.000
df_3	0.432	0.516	0.613	0.819	0.869	0.927	0.962	0.977	0.987	0.994	0.998	0.998
df_4	0.332	0.394	0.499	0.734	0.794	0.880	0.917	0.943	0.964	0.977	0.984	0.996

Table 4.4: *Simulated rejection probabilities of the bootstrap test (4.1) for a normal distribution in the nonparametric regression model (3.1). The test is based on the Cramér-von Mises statistic (4.10) and the error distribution is normal (df_0) and chi-square with k degrees of freedom (df_k) such that $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = 0.5$.*

n	25			50			75			100		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.040	0.062	0.100	0.041	0.060	0.108	0.038	0.060	0.104	0.032	0.054	0.107
df_1	0.567	0.631	0.728	0.860	0.898	0.935	0.942	0.957	0.974	0.968	0.979	0.999
df_2	0.444	0.514	0.638	0.806	0.853	0.914	0.944	0.963	0.984	0.965	0.977	0.994
df_3	0.314	0.389	0.504	0.677	0.742	0.821	0.858	0.896	0.934	0.961	0.978	0.988
df_4	0.256	0.323	0.452	0.545	0.636	0.753	0.792	0.848	0.915	0.914	0.948	0.971

Table 4.5: *Simulated rejection probabilities of the bootstrap test (4.1) for a normal distribution in the nonparametric regression model (3.1). The test is based on the Kolmogorov-Smirnov statistic (4.5) and the error distribution is normal (df_0) and chi-square with k degrees of freedom (df_k) such that $E[\varepsilon_i] = 0$ and $E[\varepsilon_i^2] = 0.5$.*

A Appendix: Proofs

A.1 An auxiliary result

In order to prove Theorem 2.1 we need the following proposition.

Proposition A.1 *Under the assumptions of Theorem 2.1 we have for all $j = 1, \ldots, k$*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta)h'_{\vartheta_j}(\varepsilon_{ni}) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta)E_\vartheta[h'_{\vartheta_j}(\varepsilon_{n1})] + o_P(1).$$

Proof: The proof follows from expansion (2.3), which gives

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta)(h'_{\vartheta_j}(\varepsilon_{ni}) - E_\vartheta[h'_{\vartheta_j}(\varepsilon_{n1})])\\
&= \frac{1}{\int f\,d\psi}\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n \psi(\varepsilon_{nj})(h'_{\vartheta_j}(\varepsilon_{ni}) - E_\vartheta[h'_{\vartheta_j}(\varepsilon_{ni})])x_{ni}^T(X_n^TX_n)^{-1}x_{nj} + o_P(1).
\end{aligned}$$

Now a calculation of the variance of the dominating term yields

$$\begin{aligned}
&\operatorname{Var}\Big(\frac{1}{\int f\,d\psi}\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^n \psi(\varepsilon_{nj})(h'_{\vartheta_j}(\varepsilon_{ni}) - E_\vartheta[h'_{\vartheta_j}(\varepsilon_{ni})])x_{ni}^T(X_n^TX_n)^{-1}x_{nj}\Big)\\
&= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{\ell=1}^n E\Big[\psi(\varepsilon_{nj})(h'_{\vartheta_j}(\varepsilon_{ni}) - E_\vartheta[h'_{\vartheta_j}(\varepsilon_{ni})])\psi(\varepsilon_{nk})(h'_{\vartheta_j}(\varepsilon_{n\ell}) - E_\vartheta[h'_{\vartheta_j}(\varepsilon_{n\ell})])\\
&\qquad\qquad \times x_{ni}^T(X_n^TX_n)^{-1}x_{nj}x_{nk}^T(X_n^TX_n)^{-1}x_{n\ell}\Big]\frac{1}{(\int f\,d\psi)^2}\\
&= O(\frac{1}{n})\Big(\sum_{i=1}^n x_{ni}^T(X_n^TX_n)^{-1}x_{ni}\Big)^2 + O(\frac{1}{n})\sum_{i=1}^n\sum_{j=1}^n x_{ni}^T(X_n^TX_n)^{-1}x_{nj}x_{nj}^T(X_n^TX_n)^{-1}x_{ni}\\
&= O(\frac{1}{n}) \;=\; o(1),
\end{aligned}$$

where the last line follows from assumption [A 2] and the representation

$$X_n^TX_n = \sum_{j=1}^n x_{nj}x_{nj}^T.$$

□

A.2 Proof of Theorem 2.1

We only state the proof for the case of a one-dimensional parameter space Θ, that is $k = 1$. The general case $k \geq 2$ is treated exactly in the same way with an additional amount of notation. An application of Theorem 6.2.1 from Koul (2002, p. 232), (2.7) and the expansion

$$F_{\widehat{\vartheta}}(y) - F_\vartheta(y) \;=\; \sum_{j=1}^k \frac{\partial F_\vartheta(y)}{\partial\vartheta_j}(\widehat{\vartheta}_j - \vartheta_j) + o_P(\frac{1}{\sqrt{n}}) \tag{A.1}$$

yield uniformly with respect to $y \in I\!R$,

$$\begin{aligned}
\sqrt{n}(\widehat{F}_n(y) - F_{\widehat{\vartheta}}(y)) &= \sqrt{n}(F_n(y) - F_\vartheta(y)) - \sqrt{n}(F_{\widehat{\vartheta}}(y) - F_\vartheta(y)) \\
&\quad + f_\vartheta(y)\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta) + o_P(1) \\
&= \sqrt{n}(F_n(y) - F_\vartheta(y)) - \frac{\partial F_\vartheta(y)}{\partial\vartheta}\frac{1}{\sqrt{n}}\sum_{i=1}^n h_\vartheta(\varepsilon_{ni}) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta)\Big(f_\vartheta(y) + \frac{\partial F_\vartheta(y)}{\partial\vartheta}h_\vartheta'(\varepsilon_{ni})\Big) + o_P(1).
\end{aligned}$$

From the expansion (2.3) of the M-estimator $\widehat{\beta}_n$, Proposition A.1 and the assumption [A 5] we obtain uniformly with respect to $t \in [0,1]$,

$$\sqrt{n}\left(\widehat{F}_n(F_\vartheta^{-1}(t)) - F_{\widehat{\vartheta}}(F_\vartheta^{-1}(t))\right) = G_n(t) + o_P(1). \tag{A.2}$$

Here the process G_n is defined by

$$G_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big\{I\{\varepsilon_{ni} \le F_\vartheta^{-1}(t)\} - t - \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(t))h_\vartheta(\varepsilon_{ni}) + \widetilde{x}_{ni}\psi(\varepsilon_{ni})c_\vartheta(F_\vartheta^{-1}(t))\Big\},$$

where we have used the notation

$$\widetilde{x}_{ni} = \sum_{j=1}^n x_{nj}^T(X_n^TX_n)^{-1}x_{ni}. \tag{A.3}$$

and $c_\vartheta(F_\vartheta^{-1}(t))$ is given in Theorem 2.1 (for the case $k=1$). Note that assumption [A 2] implies

$$\frac{1}{n}\sum_{j=1}^n \widetilde{x}_{nj}^2 = \frac{1}{n}\sum_{j=1}^n \widetilde{x}_{nj} \longrightarrow m^T\Sigma^{-1}m \quad \text{for } n\to\infty \tag{A.4}$$

$$\max_{j=1,\dots,n}\widetilde{x}_{nj} = O(1). \tag{A.5}$$

A straightforward but tedious calculation of the covariance gives

$$\begin{aligned}
&\mathrm{Cov}_\vartheta(G_n(s), G_n(t)) = E_\vartheta[G_n(s)G_n(t)] \\
&= \frac{1}{n}\sum_{i=1}^n E_\vartheta\Big[\Big\{I\{\varepsilon_{ni} \le F_\vartheta^{-1}(s)\} - s - \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(s))h_\vartheta(\varepsilon_{ni}) + \widetilde{x}_{ni}\psi(\varepsilon_{ni})c_\vartheta(F_\vartheta^{-1}(s))\Big\} \\
&\qquad\qquad \Big\{I\{\varepsilon_{ni} \le F_\vartheta^{-1}(t)\} - t - \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(t))h_\vartheta(\varepsilon_{ni}) + \widetilde{x}_{ni}\psi(\varepsilon_{ni})c_\vartheta(F_\vartheta^{-1}(t))\Big\}\Big] \\
&= E_\vartheta\Big[(I\{\varepsilon_{n1} \le F_\vartheta^{-1}(s)\} - s)(I\{\varepsilon_{n1} \le F_\vartheta^{-1}(t)\} - t)\Big] \\
&\quad + \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(s))\frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(t))E_\vartheta[h_\vartheta^2(\varepsilon_{n1})]
\end{aligned}$$

$$
\begin{aligned}
&+ c_\vartheta(F_\vartheta^{-1}(s))c_\vartheta(F_\vartheta^{-1}(t))E_\vartheta[\psi^2(\varepsilon_{n1})]\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}^2 \\
&- \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(t))E_\vartheta[I\{\varepsilon_{n1} \le F_\vartheta^{-1}(s)\}h_\vartheta(\varepsilon_{n1})] \\
&- \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(s))E_\vartheta[I\{\varepsilon_{n1} \le F_\vartheta^{-1}(t)\}h_\vartheta(\varepsilon_{n1})] \\
&+ c_\vartheta(F_\vartheta^{-1}(t))E_\vartheta[I\{\varepsilon_{n1} \le F_\vartheta^{-1}(s)\}\psi(\varepsilon_{n1})]\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni} \\
&+ c_\vartheta(F_\vartheta^{-1}(s))E_\vartheta[I\{\varepsilon_{n1} \le F_\vartheta^{-1}(t)\}\psi(\varepsilon_{n1})]\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni} \\
&- \Big\{\frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(s))c_\vartheta(F_\vartheta^{-1}(t)) + \frac{\partial F_\vartheta}{\partial\vartheta}(F_\vartheta^{-1}(t))c_\vartheta(F_\vartheta^{-1}(s))\Big\}E_\vartheta[h_\vartheta(\varepsilon_{n1})\psi(\varepsilon_{n1})]\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni} \\
= {} & \operatorname{Cov}_\vartheta(G(s), G(t)) + o(1),
\end{aligned}
$$

where $\operatorname{Cov}_\vartheta(G(s), G(t))$ is defined in Theorem 2.1, and the last equality follows with (A.4). Finally, the weak convergence of the process G_n to the Gaussian process G follows from the convergence of the finite dimensional distributions and tightness. The proof of the convergence of the finite dimensional distributions to normal distributions can be done by a straightforward application of the Cramér–Wold device and a verification of Lindeberg's condition. Tightness of the process G_n can be shown in terms of asymptotic equicontinuity, that is for all $\epsilon > 0$

$$
\lim_{\delta \searrow 0} \limsup_{n\to\infty} P\Big(\sup_{\substack{s,t\in[0,1]\\ |s-t|\le\delta}} \Big|G_n(s) - G_n(t)\Big| > \epsilon\Big) = 0,
$$

which follows from tightness of the standard empirical process and from the uniform continuity of the functions $f_\vartheta \circ F_\vartheta^{-1}$ and $\frac{\partial F_\vartheta}{\partial\vartheta} \circ F_\vartheta^{-1}$. □

A.3 Proof of Theorem 2.3

For the sake of transparency we only consider the case $k = 1$ of a one-dimensional parameter space. The general case is treated exactly in the same way with an additional amount of notation. Additionally, we assume without loss of generality the almost sure convergence of $\widehat{\vartheta}$ to ϑ_0 (note that every subsequence contains a subsequence that converges almost surely). From Lemma 3.1 in Koul and Lahiri (1994) and assumption (2.8) it follows for all $\delta > 0$

$$
P\left(\sup_{t\in[0,1]} \Big|\sqrt{n}\Big(\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)\Big)\Big| > \delta \;\Big|\; \mathcal{Y}_n\right) = o_P(1).
$$

This estimate implies equivalence, in the sense of conditional weak convergence, of the processes $\sqrt{n}(\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t)))$ and

$$
\sqrt{n}(F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))) - \sqrt{n}(F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t)))
$$

$$
\begin{aligned}
&+ f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) \\
= \;& \sqrt{n}(F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))) - \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{\sqrt{n}}\sum_{i=1}^n h_{\widehat{\vartheta}}(\varepsilon_{ni}^*) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)\Big(f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t)) + \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))E[h'_{\widehat{\vartheta}}(\varepsilon_{n1}^*) \mid \mathcal{Y}_n]\Big) + o_{P_n^*}(1) \\
= \;& G_n^*(t) + o_{P_n^*}(1)
\end{aligned}
$$

in probability, where we used (2.9), [A 10], the bootstrap analogue of (A.1) and the process G_n^* is defined by

$$
G_n^*(t) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^n e_{ni}^*(t)
$$

with

$$
e_{ni}^*(t) \;=\; \Big\{ I\{F_{\widehat{\vartheta}}(\varepsilon_{ni}^*) \le t\} - t - \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))h_{\widehat{\vartheta}}(\varepsilon_{ni}^*) + \widetilde{x}_{ni}\psi(\varepsilon_{ni}^*)c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\Big\}
$$

and

$$
c_{\widehat{\vartheta}}(y) \;=\; \frac{1}{\int f_{\widehat{\vartheta}}\, d\psi}\Big(f_{\widehat{\vartheta}}(y) + \frac{\partial F_\vartheta(y)}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}\int h'_{\widehat{\vartheta}}(t)f_{\widehat{\vartheta}}(t)\, dt\Big), \quad y \in I\!R. \tag{A.6}
$$

Noting that $E[G_n^*(t) \mid \mathcal{Y}_n] = 0$ and that $c_{\widehat{\vartheta}}(y)$ converges in probability to the quantity $c_{\vartheta_0}(y)$ defined in Theorem 2.1 we obtain for the conditional covariance of the bootstrap process

$$
\begin{aligned}
&\mathrm{Cov}(G_n^*(s), G_n^*(t) \mid \mathcal{Y}_n) = E[G_n^*(s)G_n^*(t) \mid \mathcal{Y}_n] \;=\; \frac{1}{n}\sum_{i=1}^n E[e_{ni}^*(s)e_{ni}^*(t) \mid \mathcal{Y}_n] \\
= \;& s \wedge t - st + \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s))\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\int h_{\widehat{\vartheta}}^2(x)f_{\widehat{\vartheta}}(x)\, dx \\
&+ \int \psi^2(x)f_{\widehat{\vartheta}}(x)\, dx\, c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s))c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}^2 \\
&- \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\int_0^s h_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(x))\, dx - \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s))\int_0^t h_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(x))\, dx \\
&+ \Big(c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\int_0^s \psi(F_{\widehat{\vartheta}}^{-1}(x))\, dx + c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s))\int_0^t \psi(F_{\widehat{\vartheta}}^{-1}(x))\, dx\Big)\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni} \\
&- \Big(c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s)) + c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(s))\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\Big) \\
&\times \int h_{\widehat{\vartheta}}(x)\psi(x)f_{\widehat{\vartheta}}(x)\, dx \frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}\;,
\end{aligned}
$$

which converges in probability to the covariance kernel defined in Theorem 2.1 where the parameter ϑ has to be replaced by ϑ_0.

Conditional weak convergence of the process G_n^* to a Gaussian process in probability follows from conditional convergence of the finite-dimensional distributions in probability and conditional tightness [compare the proofs of Lemma A.2 and Lemma A.3 of Stute, González Manteiga and Presedo Quindimil (1998)]. To prove convergence of the finite-dimensional distributions we fix constants $a_1, \ldots, a_\ell \in I\!R, t_1, \ldots, t_\ell \in [0,1]$ and consider the linear combination

$$\sum_{j=1}^{\ell} a_j G_n^*(t_j) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\sum_{j=1}^{\ell} a_j e_{ni}^*(t_j) \Big) \;=\; \bar{G}_{n,1}^* - \bar{G}_{n,2}^* + \bar{G}_{n,3}^*,$$

where

$$\begin{aligned}
\bar{G}_{n,1}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{\ell} a_j (I\{\varepsilon_{ni}^* \le F_{\widehat{\vartheta}}^{-1}(t_j)\} - t_j), \\
\bar{G}_{n,2}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_{\widehat{\vartheta}}(\varepsilon_{ni}^*) \Big(\sum_{j=1}^{\ell} a_j \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} F_{\widehat{\vartheta}}^{-1}(t_j) \Big), \\
\bar{G}_{n,3}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{x}_{ni} \psi(\varepsilon_{ni}^*) \Big(\sum_{j=1}^{\ell} a_j c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t_j)) \Big).
\end{aligned}$$

We show, conditionally on the sample, that Lindeberg's condition is fulfilled in probability. The verification of the Lindeberg condition for the sum $\bar{G}_{n,1}^* - \bar{G}_{n,2}^* + \bar{G}_{n,3}^*$ follows from the validity of the Lindeberg condition for each summand $\bar{G}_{n,i}^*$ $(i = 1,2,3)$ by a standard argument. For the sake of brevity we only consider $\bar{G}_{n,3}^*$, the remaining terms are treated similarly. For $\delta > 0$ we obtain for the Lindeberg condition

$$\begin{aligned}
& L_n(\delta) \\
&= \frac{1}{n} \sum_{i=1}^{n} E\left[\widetilde{x}_{ni}^2 \psi^2(\varepsilon_{ni}^*) \Big(\sum_{j=1}^{\ell} a_j c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t_j)) \Big)^2 I\Big\{ \Big| \widetilde{x}_{ni} \psi(\varepsilon_{ni}^*) \Big(\sum_{j=1}^{\ell} a_j c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t_j)) \Big) \Big| > \sqrt{n}\delta \Big\} \;\Big|\; \mathcal{Y}_n \right] \\
&\le \Big(\sum_{j=1}^{\ell} a_j c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t_j)) \Big)^2 \frac{1}{n} \sum_{i=1}^{n} \widetilde{x}_{ni}^2 \int \psi^2(t) f_{\widehat{\vartheta}}(t) I\Big\{ |\widetilde{x}_{ni}| |\psi(t)| \Big| \sum_{j=1}^{\ell} a_j c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t_j)) \Big| > \sqrt{n}\delta \Big\}\, dt \\
&= o_P(1),
\end{aligned}$$

where the last line follows with the relations (A.4) and (A.5) and the convergence of $c_{\widehat{\vartheta}}$ [defined in (A.6)] and $f_{\widehat{\vartheta}}$ in probability for $n \to \infty$.

To prove tightness, conditionally on the sample in probability, we consider the following decomposition of the process,

$$G_n^*(t) = G_{n,1}^*(t) + G_{n,2}^*(t) + G_{n,3}^*(t),$$

where

$$G_{n,1}^*(t) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I\{F_{\widehat{\vartheta}}(\varepsilon_{ni}^*) \le t\} - t) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (I\{U_i \le t\} - t)$$

$$G^*_{n,2}(t) \;=\; \Big(\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} \circ F^{-1}_{\widehat{\vartheta}}\Big)(t)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h_{\widehat{\vartheta}}(\varepsilon^*_{ni})$$

$$G^*_{n,3}(t) \;=\; (c_{\widehat{\vartheta}} \circ F^{-1}_{\widehat{\vartheta}})(t)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \widetilde{x}_{ni}\psi(\varepsilon^*_{ni}).$$

The random variables U_i are independent and have a uniform distribution on the interval $[0,1]$ and from this follows tightness of $G^*_{n,1}$. In the following we show that for almost all samples $\mathcal{Y}_n$, the process $G^*_{n,2}$ is tight in terms of stochastic equicontinuity. Tightness of $G^*_{n,3}$ follows similarly. We have for all $\epsilon > 0, \delta > 0$

$$P\Big(\sup_{\substack{s,t\in[0,1]\\ |s-t|\le\delta}} |G^*_{n,2}(s) - G^*_{n,2}(t)| > \epsilon \;\Big|\; \mathcal{Y}_n\Big)$$
$$\le \frac{1}{\epsilon^2}\int h^2_{\widehat{\vartheta}}(x) f_{\widehat{\vartheta}}(x)\,dx \sup_{\substack{s,t\in[0,1]\\ |s-t|\le\delta}} \Big|\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} \circ F^{-1}_{\widehat{\vartheta}}(s) - \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} \circ F^{-1}_{\widehat{\vartheta}}(t)\Big|.$$

The integral on the r.h.s. of the inequality is bounded almost surely. The supremum can be estimated as follows,

$$\sup_{\substack{s,t\in[0,1]\\ |s-t|\le\delta}} \Big|\frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} (F^{-1}_{\widehat{\vartheta}}(s)) - \frac{\partial F_\vartheta}{\partial \vartheta}\Big|_{\vartheta=\widehat{\vartheta}} (F^{-1}_{\widehat{\vartheta}}(t))\Big|$$
$$\le \sup_{\substack{s,t\in[0,1]\\ |s-t|\le\delta}} \Big|\frac{\partial F_\vartheta}{\partial \vartheta}|_{\vartheta=\vartheta_0} (F^{-1}_{\vartheta_0}(s)) - \frac{\partial F_\vartheta}{\partial \vartheta}|_{\vartheta=\vartheta_0} (F^{-1}_{\vartheta_0}(t))\Big|$$
$$+2\sup_{t\in[0,1]} \Big|\frac{\partial F_\vartheta}{\partial \vartheta}|_{\vartheta=\vartheta_0} (F^{-1}_{\widehat{\vartheta}}(t)) - \frac{\partial F_\vartheta}{\partial \vartheta}|_{\vartheta=\vartheta_0} (F^{-1}_{\vartheta_0}(t))\Big|.$$

The first term on the r.h.s. converges to zero for $\delta \downarrow 0$ because of the uniform continuity of $\frac{\partial F_\vartheta}{\partial \vartheta}$. The second term converges to zero almost surely for $n \to \infty$ by assumption [A 9]. This proves tightness of $G^*_{n,2}$ for almost all samples $\mathcal{Y}_n$ and completes the proof of Theorem 2.3. □

A.4 Proof of Theorem 3.1

The main difference in the proof of Theorem 3.1 compared to the proof of Theorem 2.1 is a different derivation of expansion (A.2). For this we need the following two propositions.

Proposition A.2 *Under the assumptions of Theorem 3.1 we have for all $j = 1, \dots, k$*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} h'_{\vartheta_j}(\varepsilon_i)(\widehat{m}(X_i) - m(X_i)) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varepsilon_i E_\vartheta[h'_{\vartheta_j}(\varepsilon_1)] + o_P(1).$$

Proof: The proof is similar to the proof of Proposition A.1, but uses standard calculation techniques for nonparametric regression estimation and is omitted for the sake of brevity. □

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

$$\widehat{F}_n(y) \;=\; F_n(y) + f(y)\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in I\!R$.

Proof: Analogous to Lemma B.1 of Akritas and van Keilegom (2001) one can derive the following expansion of the empirical process based on residuals,

$$\begin{aligned}\widehat{F}_n(y) \;&=\; \frac{1}{n}\sum_{i=1}^{n} I\{\widehat{\varepsilon}_i \le y\} \;=\; \frac{1}{n}\sum_{i=1}^{n} I\{\varepsilon_i \le y + \widehat{m}(X_i) - m(X_i)\} \\ &=\; \frac{1}{n}\sum_{i=1}^{n} I\{\varepsilon_i \le y\} + P(\varepsilon \le y + \widehat{m}(X) - m(X) \mid \mathcal{Y}_n) - F(y) + o_P(\frac{1}{\sqrt{n}})\end{aligned}$$

uniformly with respect to $y \in I\!R$, where (ε, X) has the same distribution as (ε_1, X_1) but is independent from the sample $\mathcal{Y}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. Now by a Taylor expansion it follows

$$\begin{aligned}\widehat{F}_n(y) \;&=\; F_n(y) + \int (F(y + \widehat{m}(x) - m(x)) - F(y)) f_X(x)\, dx + o_P(\frac{1}{\sqrt{n}}) \\ &=\; F_n(y) + f(y)\int (\widehat{m}(x) - m(x)) f_X(x)\, dx + o_P(\frac{1}{\sqrt{n}}).\end{aligned}$$

Inserting the definition (3.2) of the regression estimator $\widehat{m}$ and some straightforward calculations of expectations and variances yield the assertion of the Proposition. □

For the proof of Theorem 3.1 we only consider the case $k = 1$ for the sake of transparent notation. From the two propositions above and (A.1) analogous to the expansion (A.2) in the linear case we obtain the representation

$$\sqrt{n}(\widehat{F}_n(F_\vartheta^{-1}(t)) - F_{\widehat{\vartheta}}(F_\vartheta^{-1}(t))) \;=\; G_n(t) + o_P(1)$$

uniformly with respect to $t \in [0,1]$. Here the process G_n is defined by

$$G_n(t) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{ I\{\varepsilon_i \le F_\vartheta^{-1}(t)\} - t - \frac{\partial F_\vartheta}{\partial \vartheta}(F_\vartheta^{-1}(t)) h_\vartheta(\varepsilon_i) + \varepsilon_i c_\vartheta(F_\vartheta^{-1}(t))\Big\},$$

where

$$c_\vartheta(y) \;=\; f_\vartheta(y) + \frac{\partial F_\vartheta(y)}{\partial \vartheta}\int h'_\vartheta(z) f_\vartheta(z)\, dz, \quad y \in I\!R.$$

Now the proof of Theorem 3.1 follows exactly the lines of the proof of Theorem 2.1. □

A.5 Proof of Theorem 3.2

For the proof of Theorem 3.2 we need the following two propositions.

Proposition A.4 *Under the assumptions of Theorem 3.2 we have the expansion*

$$\widehat{F}_n^*(y) \;=\; F_n^*(y) + \frac{1}{n}\sum_{i=1}^{n}\Big(F_{\widehat{\vartheta}}(y + \widehat{m}^*(X_i) - \widehat{m}(X_i)) - F_{\widehat{\vartheta}}(y)\Big) + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $y \in I\!R$.

Proof: We consider the empirical process

$$Z_n(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i, U_i),$$

indexed by a class $\mathcal{G}$ of functions defined by

$$\mathcal{G} = \Big\{ g_{y,\vartheta,\ell} \;\Big|\; y \in I\!R, \vartheta \in \Theta, ||\vartheta - \vartheta_0|| \le M, \ell \in \mathcal{C}_1^{1+\delta/2}[0,1] \Big\}$$

where M is a fixed positive constant, $\mathcal{C}_1^{1+\delta/2}[0,1]$ denotes a class of bounded smooth functions $\ell : [0,1] \to I\!R$ [see van der Vaart and Wellner (1996, p. 154) for the exact definition], and

$$\begin{aligned} g_{y,\vartheta,\ell}(x,u) \;=\;& I\{u \le F_\vartheta(y + \ell(x))\} - I\{u \le F_\vartheta(y)\} \\ & - E[I\{U \le F_\vartheta(y + \ell(X))\}] + E[I\{U \le F_\vartheta(y)\}]. \end{aligned}$$

With these definitions we have

$$\sqrt{n}\Big(\widehat{F}_n^*(y) - F_n^*(y) - \frac{1}{n}\sum_{i=1}^{n}(F_{\widehat{\vartheta}}(y + \widehat{m}^*(X_i) - \widehat{m}(X_i)) - F_{\widehat{\vartheta}}(y))\Big) = Z_n(g_{y,\widehat{\vartheta},\widehat{m}^*-\widehat{m}}),$$

where $\lim_{n\to\infty} P(g_{y,\widehat{\vartheta},\widehat{m}^*-\widehat{m}} \in \mathcal{G}) = 1$. The proof of the assertion

$$\sup_{y\in I\!R} |Z_n(g_{y,\widehat{\vartheta},\widehat{m}^*-\widehat{m}})| = o_P(1)$$

follows along the lines of the proof of Lemma B.1 of Akritas and van Keilegom (2001), but is more complicated due to the fact that our function class is additionally indexed by the parameter ϑ. For the sake of self-consistency we will explain the main differences here. First we show that the bracketing number satisfies the integral condition

$$\int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{G}, L^2)}\, d\epsilon < \infty. \tag{A.7}$$

To keep the proof of this property readable we confine ourselves to a proof of condition (A.7) for the class

$$\mathcal{G}_1 = \Big\{ (x,u) \mapsto I\{u \le F_\vartheta(y + \ell(x))\} \;\Big|\; y \in I\!R, \vartheta \in \Theta, ||\vartheta - \vartheta_0|| \le M, \ell \in \mathcal{C}_1^{1+\delta/2}[0,1] \Big\}.$$

To this end we fix for $\epsilon > 0$ a grid $\vartheta_1, \ldots, \vartheta_K$ in Θ to cover $\{\vartheta \in \Theta \mid ||\vartheta - \vartheta_0|| \leq M\}$ with $K = O(\epsilon^{-2k})$ balls of radius ϵ^2. Using Example 19.7. in van der Vaart (1998, p. 271) and assumption [A 5] we obtain K brackets of the form

$$[F_{\vartheta_j} - B\epsilon^2, F_{\vartheta_j} + B\epsilon^2], \quad j = 1, \ldots, K,$$

for the parametric class of functions $\{F_\vartheta \mid \vartheta \in \Theta, ||\vartheta - \vartheta_0|| \leq M\}$, where the constant B is defined as

$$B = \sup_{y \in I\!R} \sup_{||\vartheta - \vartheta_0|| \leq M} \Big|\Big| \frac{\partial F_\vartheta(y)}{\partial \vartheta} \Big|\Big|.$$

For some constant C, the class $\mathcal{C}_1^{1+\delta/2}[0,1]$ can be covered by $m = O(\exp(C\epsilon^{-2/(1+\delta/2)}))$ brackets $[\ell_i^L, \ell_i^U]$, $i = 1, \ldots, m$ [see Corollary 2.7.2 in van der Vaart and Wellner (1996)]. For fixed $j \in \{1, \ldots, K\}$ and $i \in \{1, \ldots, m\}$ we choose $M = O(\epsilon^{-4})$ brackets for $y \in I\!R$ analogous to the argument in Lemma B.1 of Akritas and van Keilegom (2001). Now $\mathcal{G}_1$ can be covered by KmM brackets of the form

$$\Big[I\{u \leq F_{\vartheta_j}(y_{jih}^L + \ell_i^L(x)) - B\epsilon^2\}, I\{u \leq F_{\vartheta_j}(y_{jih}^U + \ell_i^U(x)) + B\epsilon^2\}\Big]$$

($j = 1, \ldots, K$, $i = 1, \ldots, m$, $h = 1, \ldots, M$). The L^2–distance between upper and lower bracket is then

$$\begin{aligned}
&\Big[P\Big(F_{\vartheta_j}(y_{jih}^L + \ell_{ji}^L(X_{n1})) - B\epsilon^2 < U_1 \leq F_{\vartheta_j}(y_{jih}^U + \ell_{ji}^U(X_{n1})) + B\epsilon^2\Big)\Big]^{1/2} \\
\leq &\Big[\int \sup_{y \in I\!R} \Big(F_{\vartheta_j}(y + \ell_{ji}^U(x)) - F_{\vartheta_j}(y + \ell_{ji}^L(x))\Big) f_X(x)\, dx + \widetilde{B}\epsilon^2\Big]^{1/2} \\
\leq &\; O(\epsilon),
\end{aligned}$$

with the same reasoning as in the proof of Lemma B.1 of Akritas and van Keilegom (2001). We have $N_{[\,]}(\epsilon, \mathcal{G}_1, L^2) = O(\epsilon^{-4-2k}\exp(C\epsilon^{-2/(1+\delta/2)}))$ brackets in total. This yields the assertion (A.7) for the class $\mathcal{G}_1$ and therefore by a similar argument for the remaining cases the weak convergence of the process $\{Z_n(g)\}_{g \in \mathcal{G}}$ to a Gaussian process $\{Z(g)\}_{g \in \mathcal{G}}$. For the next step of the reasoning one can show in a similar way as in the aforementioned proof the estimate

$$\sup_{y \in I\!R} \sup_{||\vartheta - \vartheta_0|| \leq M} \mathrm{Var}(g_{y,\vartheta,\ell}(X_1, U_1)) \leq c \int \ell(x) f_X(x)\, dx$$

for some constant c. From this it follows directly

$$\sup_{y \in I\!R} E\Big[g^2_{y,\widehat{\vartheta},\widehat{m}^* - \widehat{m}}(X, U) \;\Big|\; \mathcal{Y}_n, U_1, \ldots, U_n\Big] = o_p(1),$$

where (X, U) has the same distribution as (X_1, U_1) but is independent from the sample. Now the assertion of the proposition can be deduced by applying Corollary 2.3.12. in van der Vaart and Wellner (1996) [confer the end of the proof of Lemma B.1, Akritas and van Keilegom (2001)] or by making use of a uniform version of Lemma 19.24 [compare with the proof of Theorem 19.26.] in van der Vaart (1998). □

Proposition A.5 *Under the assumptions of Theorem 3.2 we have the expansion*

$$\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) \;=\; F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) + f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ni}^* + o_{P_n^*}(\frac{1}{\sqrt{n}})$$

uniformly with respect to $t \in [0,1]$.

Proof: By Proposition A.4, Markov's inequality, and a Taylor expansion we directly obtain

$$\begin{aligned}
&\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) = \frac{1}{n}\sum_{i=1}^{n}\Big(F_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t) + \widehat{m}^*(X_i) - \widehat{m}(X_i)) - t\Big) + o_{P_n^*}(\frac{1}{\sqrt{n}})\\
&= f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{n}\sum_{i=1}^{n}(\widehat{m}^*(X_i) - \widehat{m}(X_i)) + \frac{1}{2}f_{\widehat{\vartheta}}'(\xi_{t,i})\frac{1}{n}\sum_{i=1}^{n}(\widehat{m}^*(X_i) - \widehat{m}(X_i))^2 + o_{P_n^*}(\frac{1}{\sqrt{n}})\\
&= f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t)) + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ni}^* + o_{P_n^*}(\frac{1}{\sqrt{n}}),
\end{aligned}$$

uniformly with respect to $t \in [0,1]$, where the last equality follows by inserting the definition of the estimators $\widehat{m}^*$ and $\widehat{m}$ and some straightforward but cumbersome calculations of expectations and variances. □

Proposition A.5 yields together with assumptions [A 5] and [A 17] the following stochastic expansion

$$\begin{aligned}
&\sqrt{n}\left(\widehat{F}_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t))\right)\\
&= \sqrt{n}\Big(F_n^*(F_{\widehat{\vartheta}}^{-1}(t)) - t + f_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{ni}^* - F_{\widehat{\vartheta}^*}(F_{\widehat{\vartheta}}^{-1}(t)) + t\Big) + o_{P_n^*}(1)\\
&= G_n^*(F_{\widehat{\vartheta}}^{-1}(t)) + o_{P_n^*}(1),
\end{aligned}$$

where

$$\begin{aligned}
G_n^*(F_{\widehat{\vartheta}}^{-1}(t)) \;&=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{I\{F_{\widehat{\vartheta}}(\varepsilon_{ni}^*) \le t\} - t - \frac{\partial F_\vartheta}{\partial\vartheta}\Big|_{\vartheta=\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))h_{\widehat{\vartheta}}(\varepsilon_{ni}^*) + \varepsilon_{ni}^* c_{\widehat{\vartheta}}(F_{\widehat{\vartheta}}^{-1}(t))\Big\},\\
c_{\widehat{\vartheta}}(y) \;&=\; f_{\widehat{\vartheta}}(y) + \frac{\partial F_\vartheta(y)}{\partial\vartheta}\Big|_{\vartheta=\widehat{\vartheta}}\int h_{\widehat{\vartheta}}'(z)f_{\widehat{\vartheta}}(z)\,dz, \quad y \in I\!R.
\end{aligned}$$

The rest of the proof follows exactly in the same way as the proof of Theorem 2.3. □

Acknowledgements. The authors are grateful to Prof. H. Koul for some useful comments on related literature and for sending us the paper of Khmaladze and Koul (2004) before publication and to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise. The work of the authors was supported by the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen).

References

M. Akritas and I. van Keilegom (2001). Nonparametric estimation of the residual distribution. Scand. J. Statist. 28, 549–567.

R. Beran (1982). Estimated sampling distributions: The bootstrap and competitors. Ann. Statist. 10, 212-225.

L. Birgé and P. Massart (2001). Gaussian model selection. J. Eur. Math. Soc. 3, 203–268.

L. D. Brown and M. G. Low (1996). Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist. 24, 2384–2398.

L. D. Brown, T. T. Cai, M. G. Low and C.-H. Zhang (2002). Asymptotic equivalence theory for nonparametric regression with random design. Ann. Statist. 30, 688–707.

E. Brunner and N. Neumann (1987). Nonparametric methods for the 2-period-crossover design under weak model assumptions. Biom. J. 29, 907-920.

D. A. Darling (1955). The Cramér-Smirnov test in the parametric case. Ann. Math. Statist. 26, 1-20.

M. Denker (1985). Asymptotic Distribution Theory in Nonparametric Statistics. Vieweg & Sons, Braunschweig.

J. Durbin (1973). Weak convergence of the sample distribution function when the parameters are estimated. Ann. Statist. 2, 279-290.

R. L. Eubank and J. D. Hart (1992). Testing goodness-of-fit in regression via order selection criteria. Ann. Statist. 20, 1412–1425.

J. Fan and I. Gijbels (1996). Local polynomial modelling and its applications. Chapman and Hall, London.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W.A. Stahel (1986). Robust Statistics. Wiley, New York.

E. V. Khmaladze and H. L. Koul (2004). Martingale transforms goodness-of-fit tests in regression models. Ann. Statist., to appear.

H. L. Koul and S. N. Lahiri (1994). On Bootstrapping M-Estimated Residual Processes in Multiple Linear Regression Models. Journal Multivariate Analysis 49, 255–265.

H. L. Koul (2002). Weighted Empirical Processes in Dynamic Nonlinear Models (Second Edition). Springer, New York.

O. V. Lepski and A. B. Tsybakov (2000). Asymptotically and exact nonparametric hypothesis testing in sup-norm and at a fixed point. Probab. Theory Related Fields 117, 17–48.

R. M. Loynes (1980). The empirical distribution function of residuals from generalized regression. Ann. Statist. 8, 285-298.

J. I. Marden and M. E. T. Muyot (1995). Rank tests for main and interaction effects in analysis of variance. J. Amer. Stat. Assoc. 90, 1388-1398.

H. Milhem (2003). Testing the Goodness-of-Fit in a Gaussian Regression. Proceedings of the EYSM 2003, 101–111.

D. A. Pierce and K. J. Kopecky (1979). Testing goodness of fit for the distribution of errors in regression models. Biometrika 66, 1-6.

F. Pukelsheim (1993). Optimum Experimental Design. Wiley, N.Y.

R. H. Randles and D. A. Wolfe (1979). Introduction to the Theory of Nonparametric Statistics. Krieger Publishing Company.

J. Rice (1984). Bandwidth choice for nonparametric regression. Ann. Statist. 12, 1215-1230.

R. J. Serfling (1980). Approximation Theorems of Mathematical Statistics. Wiley, N.Y.

W. Stute, W. González Manteiga and M. Presedo Quindimil (1998). Bootstrap Approximation in Model Checks for Regression. J. Amer. Statist. Assoc. 93, 141–149.

S. Sukhatme (1972). Fredholm determinant of a positive kernel of a special type and its applications. Ann. Math. Statist. 43, 1914-1926.

G. L. Thompson (1991). A unified approach to rank tests for multivariate and repeated measures designs. J. Amer. Statist. Assoc. 86, 410–419.

A. W. van der Vaart (1998). Asymptotic Statistics. Cambridge University Press.

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

S. G. Young and A. W. Bowman (1995). Non-parametric analysis of covariance. Biometrics 51, 920-931.