A note on testing symmetry of the error distribution in linear 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

September 17, 2003

Abstract

In the classical linear regression model the problem of testing for symmetry of the error distribution is considered. The test statistic is a functional of the difference between the two empirical distribution functions of the estimated residuals and their counterparts with opposite signs. The weak convergence of the difference process to a Gaussian process is established. The covariance structure of this process depends heavily on the density of the error distribution, and for this reason the performance of a symmetric wild bootstrap procedure is discussed in asymptotic theory and by means of a simulation study.

AMS Classification: 62J05, 62J10, 62G10
Keywords and Phrases: M-estimation, goodness-of-fit tests, testing for symmetry, empirical process of residuals, linear model

1 Introduction

Consider the linear model

$$Y_{ni} = x_{ni}^T\beta + \varepsilon_{ni} \quad (i = 1, \dots, n) \tag{1.1}$$

where $\beta \in \mathbb{R}^p$ denotes the unknown parameter and the errors ε_{ni} are assumed to be independent and identically distributed with $E[\varepsilon_{ni}] = 0$. Throughout this paper let $X_n \in \mathbb{R}^{n\times p}$ denote the design matrix in the linear model, where the vector $x_{ni}^T = (x_{ni1}, \dots, x_{nip})$ corresponds to the ith row of the matrix X_n and is not random. In this paper we are interested in the problem of testing the hypotheses

$$(1.2)\quad H_0: \ F(y) = 1 - F(-y) \text{ for all } y \in I\!R \quad vs \quad H_1: \ F(y) \neq 1 - F(-y) \text{ for some } y \in I\!R$$

where F is the distribution function of the random variables ε_{ni}. As pointed out by Bickel (1982) the assumption of symmetry in the linear regression model (1.1) is of importance because in this case asymptotically efficient adaptive estimates for the parameter β in the linear model (1.1) exist. The problem of testing symmetry of the unknown distribution of the residuals in regression models has been considered by numerous authors in the literature for various special cases of regression models. Most of the literature concentrates on the problem of testing the symmetry of the distribution of an i.i.d. sample about an unknown mean [see for example Bhattacheraya, Gastwirth and Wright (1982), Aki (1981), Koziol (1985), Schuster and Barker (1987) or Psaradakis (2003) among many others]. Recently, Fan and Gencay (1999) and Ahmad and Li (1997) proposed tests for this problem in the linear model (1.1), which are based on kernel density estimates from the estimated residuals. Fan and Gencay's (1999) test is based on an affinity measure between the density f of the error distribution and its reflection. Ahmad and Li (1997) transferred an approach of Rosenblatt (1975) for testing independence to the problem of testing symmetry in a linear model with homoscedastic errors. This test depends on a smoothing parameter h required for the density estimation of the error distribution and is consistent against alternatives converging to the null hypothesis at a rate $(n\sqrt{h})^{-1/2}$, where n denotes the sample size. The resulting test is attractive from a practical point of view because the corresponding test statistic can easily be calculated and is asymptotically normal distributed. However, the performance of the test is very sensitive with respect to the choice of the bandwidth h and the test cannot detect alternatives converging to the null at a rate $n^{-1/2}$. We actually construct a test for symmetry in the linear model (1.1) which does not depend on a smoothing parameter and can detect alternatives converging to the null at a rate $n^{-1/2}$. For this we consider the empirical process

$$(1.3)\qquad S_n(y) := \widehat{F}_n(y) + \widehat{F}_n(-y-) - 1 \ = \ \frac{1}{n}\sum_{i=1}^{n}\Big(I\{\widehat{\varepsilon}_{ni} \leq y\} - I\{-\widehat{\varepsilon}_{ni} \leq y\}\Big),$$

where $\widehat{\varepsilon}_{ni} = Y_{ni} - x_{ni}^T\widehat{\beta}_n$ $(i = 1, \dots, n)$ denote the residuals obtained from an M-estimator for the parameter β in the linear model (1.1). Note that a process of this type was also introduced by Koul (2002, Section 6.5) and that the statistic on the left hand side estimates the function $F(y) + F(-y) - 1$, which vanishes under the assumption of symmetry of the error distribution. Therefore, in principle, a test for the hypothesis (1.2) can be based on a Kolmogorov-Smirnov or Cramér-von Mises type functional of the process $S_n(y)$. However, as pointed out by Koul (2002) tests based on the process (1.3) are in general not asymptotically distribution-free. The purpose of the present paper is to investigate the performance of a bootstrap version of a test based on a functional of the process (1.3) from an asymptotic point of view and by means of a simulation study. In Section 2 we present some preliminary results and assumptions, while Section 3 contains our main theoretical results. We prove weak convergence of a centered version of the empirical process S_n to a Gaussian process under the null hypothesis of a symmetric error distribution and any fixed alternative. The covariance structure of the limiting process depends on the unknown distribution of the error, estimation method and design matrix in a complicated way, and as a consequence an asymptotically distribution free test statistic is not available. We propose a modification of the wild bootstrap approach to compute critical values and prove its consistency. The finite sample properties of this bootstrap procedure are discussed by means of a simulation study in Section 4 for the case of a linear regression and an ANOVA model. The numerical results indicate that the new bootstrap test is an interesting alternative to the existing procedures because it is applicable for sample sizes larger than 20 and more powerful than the tests based on kernel estimators. Finally some of the more technical details are deferred to the appendix.

2 M-estimators in linear models

We assume that the density f of the error distribution exists and is uniformly continuous, positive and denote with F the corresponding distribution function. Moreover the design matrix $X_n \in I\!R^{n\times p}$ is assumed to be of rank $p \leq n$ and to satisfy the following regularity assumptions

$$\max_{i=1,\ldots,n} x_{ni}^T (X_n^T X_n)^{-1} x_{ni} = O(\frac{1}{n}), \tag{2.1}$$

$$\lim_{n\to\infty} \frac{1}{n} X_n^T X_n = \Sigma \in I\!R^{p\times p}, \tag{2.2}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_{ni} = m \in I\!R^{p}. \tag{2.3}$$

Example 2.1 Consider the classical one-way-layout in analysis of variance, that is

$$Y_{nij} = \mu_i + \varepsilon_{nij}, \quad i = 1, \ldots, p;\ j = 1, \ldots, n_i;\ n = \sum_{i=1}^{p} n_i. \tag{2.4}$$

In this case we have $X_n^T X_n = diag(n_1, \ldots, n_p)$ and it is easy to see that assumptions (2.1) - (2.3) are satisfied if $\lim_{n\to\infty} \frac{n_i}{n} = k_i > 0$ for all $i = 1, \ldots, p$, where $\Sigma = diag(1/k_1, \ldots, 1/k_p)$ and $m = (k_1, \ldots, k_p)^T$.

Similarly, in the classical linear regression model

$$Y_{nij} = \sum_{l=1}^{p} \beta_l f_l(t_i) + \varepsilon_{nij}, \quad i = 1, \ldots, k;\ j = 1, \ldots, n_i, \tag{2.5}$$

with linear independent functions $f_1, \dots, f_p$ and $k \geq p$ different points $t_1, \dots, t_k$ we have $x_{nij}^T = (f_1(t_i), \dots, f_p(t_i))$ $(i = 1, \dots, k; j = 1, \dots, n_i)$, and the assumptions (2.1) - (2.3) also follow from the condition $\lim_{n\to\infty} \frac{n_i}{n} = k_i > 0$ $(i = 1, \dots, k)$, where

$$\Sigma = \Big(\int f_i(t) f_j(t) d\xi(t) \Big)_{i,j=1}^p , \quad m = \Big(\int f_i(t) d\xi(t) \Big)_{i=1}^p ,$$

and ξ denotes the measure with masses k_i at the points t_i $(i = 1, \dots, k)$.

Throughout this paper 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

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

Example 2.2 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 $\psi(x) = \operatorname{sign}(x)$, 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, but in the examples presented below we concentrate on the least squares case.

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

$$||(X_n^T X_n)^{1/2} (\widehat{\beta}_n - \beta)|| \;=\; O_P(1), \tag{2.6}$$

and the stochastic expansion

$$(X_n^T X_n)^{1/2} (\widehat{\beta}_n - \beta) \;=\; \frac{1}{\int f \, d\psi} (X_n^T X_n)^{-1/2} \sum_{i=1}^n x_{ni} \psi(\varepsilon_{ni}) + o_P(1) \tag{2.7}$$

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

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

which is considered as a process in $D[-\infty,\infty] = D(\bar{I\!R})$, where we use the convention $\widehat{F}_n(-\infty) = 0$ and $\widehat{F}_n(\infty) = 1$. 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 can be made precise using condition (2.6) and Theorem 6.2.1 in Koul (2002, p. 232), which yields (uniformly with respect to $y \in I\!R$) the expansion

$$\begin{aligned}
\widehat{F}_n(y) \;&=\; F_n(y) + f(y)\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta) + o_P(\frac{1}{\sqrt{n}}) \\
&=\; F_n(y) + \frac{f(y)}{\int f\,d\psi}\frac{1}{n}\sum_{j=1}^n \widetilde{x}_{nj}\psi(\varepsilon_{nj}) + o_P(\frac{1}{\sqrt{n}}),
\end{aligned}$$

where the second equality follows from the expansion (2.7) and we used the notation

$$\widetilde{x}_{nj} \;=\; \sum_{i=1}^n x_{ni}^T (X_n^T X_n)^{-1} x_{nj}. \tag{2.8}$$

Note also that assumptions (2.2) and (2.3) imply

$$\frac{1}{n}\sum_{j=1}^n \widetilde{x}_{nj}^2 = \frac{1}{n}\sum_{j=1}^n \widetilde{x}_{nj} \;\longrightarrow\; m^T\Sigma^{-1}m \quad \text{for } n \to \infty. \tag{2.9}$$

For the construction of a test of symmetry we consider the process $\sqrt{n}(\widehat{F}_n(y) - F(y))$ and its reflection. This process has the stochastic expansion

$$\sqrt{n}(\widehat{F}_n(y) - F(y)) \;=\; G_n(y) + o_P(1)$$

(uniformly with respect to $y \in I\!R$), where the process G_n is defined by

$$G_n(y) \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^n \Big(I\{\varepsilon_{nj} \le y\} - F(y) + \frac{f(y)}{\int f\,d\psi}\widetilde{x}_{nj}\psi(\varepsilon_{nj}) \Big).$$

A continuous quantile transformation shows that this process in $D[-\infty,\infty]$ is equivalent to the process

$$\widetilde{G}_n(t) = G_n(F^{-1}(t)) \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Big(I\{U_{nj}\le t\} - t + \frac{f(F^{-1}(t))\widetilde{x}_{nj}}{\int f\circ F^{-1}\,d\widetilde{\psi}}\widetilde{\psi}(U_{nj})\Big), \quad t\in[0,1],$$

in $D[0,1]$, where the random variables $U_{nj} = F(\varepsilon_{nj})$ have a uniform distribution and $\widetilde{\psi} = \psi\circ F^{-1}$. Now a standard argument shows the weak convergence of G_n in $D[-\infty,\infty]$ to a centered Gaussian process G with covariance kernel

$$\begin{aligned}\mathrm{Cov}(G(x),G(y)) \;&=\; F(x\wedge y) - F(x)F(y) + m^T\Sigma^{-1}m\frac{f(x)f(y)}{(\int f\,d\psi)^2}\int \psi^2(t)f(t)\,dt\\ &\quad + m^T\Sigma^{-1}m\Big(\frac{f(x)}{\int f\,d\psi}\int_{-\infty}^{y}\psi(t)f(t)\,dt + \frac{f(y)}{\int f\,d\psi}\int_{-\infty}^{x}\psi(t)f(t)\,dt\Big).\end{aligned}$$

Example 2.3 Note that in general the covariance kernel depends in a complicated way on the design matrix (by the quantity m and Σ) and the distribution of the error. However, there are important situations, where the structure simplifies substantially. Consider for example the situation discussed in Example 2.1 for the one-way-layout (2.4), then it is easy to see that $m^T\Sigma m = 1$. Similarly, in the classical polynomial regression model (2.5) with $f_i(t) = t^{i-1}$ $(i = 1,\dots,p)$ it follows by a straightforward calculation that $m^T\Sigma m = 1$. If least squares estimation is used for the parameter β we obtain in these cases that the covariance matrix of the limiting process G is given by

$$\mathrm{Cov}(G(x),G(y)) = F(x\wedge y) - F(x)F(y) + f(x)f(y)\int t^2 f(t)dt + F(x)f(y) + F(y)f(x).$$

Nevertheless, even in this case the covariance kernel of the process G depends in a complicated manner on the distribution function F and its density and this remark applies in particular to the process S_n defined in (1.3).

3 Testing for a symmetric error distribution

Recall the definition of the hypothesis of symmetry in (1.2) and the definition of the empirical process in (1.3). Note that the statistic $S_n(y)$ on the left hand side estimates the function $F(y) + F(-y) - 1$, which vanishes under the assumption of symmetry of the error distribution and consequently a test for symmetry can be based on a Kolmogorov-Smirnov or Cramér-von-Mises type functional of the process $\widehat{F}_n(y) + \widehat{F}_n(-y-) - 1$. As indicated in Example 2.3 and pointed out by Koul (2002) in more detail, tests based on the process (1.3) are in general not asymptotically distribution-free. For this reason we investigate the performance of a bootstrap procedure to obtain critical values for these tests. Our first result states the weak convergence of the process (1.3) in the space $D[-\infty,\infty]$ and is needed to establish the consistency of a modification of the wild bootstrap (adapted to the problem of testing for symmetry), which will be given in the second part of this section (see Theorem 3.4).

Theorem 3.1 *The estimated empirical symmetry process*

$$\widehat{S}_n(y) \;=\; \sqrt{n}\left(\widehat{F}_n(y)+\widehat{F}_n(-y-)-F(y)-F(-y)\right)$$

converges weakly in $D[-\infty,\infty]$ to a Gaussian process G with covariance

$$\begin{aligned}\mathrm{Cov}(G(x),G(y)) \;=\;& F(x\wedge y)-F(x)F(y)+F((-x)\wedge y)-F(-x)F(y)+F(x\wedge(-y))\\ &-F(x)F(-y)+F((-x)\wedge(-y))-F(-x)F(-y)\\ &+(f(x)+f(-x))(U(y)+U(-y))m^T\Sigma^{-1}m/(\textstyle\int f\,d\psi)\\ &+(f(y)+f(-y))(U(x)+U(-x))m^T\Sigma^{-1}m/(\textstyle\int f\,d\psi)\\ &+(f(x)+f(-x))(f(y)+f(-y))m^T\Sigma^{-1}m\sigma^2/(\textstyle\int f\,d\psi)^2\end{aligned}$$

where

$$\sigma^2 = E[\psi^2(\varepsilon_{n1})] = \int_{-\infty}^{\infty}\psi^2(x)f(x)\,dx,\quad U(x)=E[I\{\varepsilon_{n1}\le x\}\psi(\varepsilon_{n1})]=\int_{-\infty}^{x}\psi(t)f(t)\,dt$$

and the matrix Σ and the vector m are defined in (2.2) and (2.3), respectively.
If the function ψ is symmetric and the null hypothesis of symmetry is valid this covariance kernel reduces to:

$$\begin{aligned}\mathrm{Cov}(G(x),G(y)) \;=\;& 2F(-(|x|\vee|y|))+\frac{4m^T\Sigma^{-1}m}{(\int f\,d\psi)}(f(x)U(y)+f(y)U(x))\\ &+\;\frac{4f(x)f(y)m^T\Sigma^{-1}m\sigma^2}{(\int f\,d\psi)^2}.\end{aligned}$$

Suitable statistics for testing symmetry of the error distribution F are, for example, of Kolmogorov–Smirnov or Cramer–von–Mises type,

$$\sup_{t\in I\!R}|\widehat{S}_n(t)| \quad\text{and}\quad \int \widehat{S}_n^2(t)\,d\widehat{H}_n(t), \tag{3.1}$$

where $\widehat{H}_n$ is the empirical distribution function of the absolute values of the estimated residuals $|\widehat{\varepsilon}_{1n}|,\ldots,|\widehat{\varepsilon}_{nn}|$, and the null hypothesis of symmetry is rejected for large values of these statistics. The asymptotic distribution of the test statistics can be obtained from Theorem 3.1, an application of the Continuous Mapping Theorem and (in the latter case) the uniform convergence of $\widehat{H}_n$,

$$\sup_{t\in I\!R}|\widehat{H}_n(t)-H(t)| \;=\; o_p(1),$$

where H denotes the distribution function of $|\varepsilon_{n1}|$. A standard argument on contiguity [see e.g. Witting, Müller–Funk (1995), Theorem 6.113, 6.124 and 6.138 or van der Vaart (1998), Section 6] now shows that the resulting tests are consistent with respect to local alternatives converging to the null at a rate $n^{-1/2}$. However, the null distribution of the process $\widehat{S}_n(t)$ is not asymptotically distribution free and the critical values cannot be computed without estimating the unknown features of the error distribution of the data generating process.
To avoid the problem of estimating the distribution function F and its density function f we propose a modification of the wild bootstrap approach, which is adapted to the specific problem of testing symmetry. For this let $v_{n1}, \ldots, v_{nn}$ be independent and identically distributed random variables, which are independent of the sample $\mathcal{Y}_n = \{\mathcal{Y}_{n1}, \ldots, \mathcal{Y}_{nn}\}$ such that $P(v_{ni} = 1) = P(v_{ni} = -1) = 1/2, i = 1, \ldots, n$. Note that whether the underlying error distribution F is symmetric or not the distribution of the random variable $v_{ni}\varepsilon_{ni}$ is symmetric with density $\widetilde{f}$ and distribution function $\widetilde{F}$ given by

$$(3.2) \qquad \widetilde{f}(t) = \frac{1}{2}(f(t) + f(-t)), \quad \widetilde{F}(t) = \frac{1}{2}(F(t) + 1 - F(-t)),$$

respectively. We now define bootstrap residuals by

$$\varepsilon^*_{ni} = v_{ni}\widehat{\varepsilon}_{ni}$$

and build new bootstrap observations by

$$(3.3) \qquad Y^*_{ni} = x^T_{ni}\widehat{\beta}_n + \varepsilon^*_{ni} \quad (i = 1, \ldots, n).$$

Finally, we calculate estimated residuals from the bootstrap sample

$$\widehat{\varepsilon}^*_{ni} = Y^*_{ni} - x^T_{ni}\widehat{\beta}^*_n = \varepsilon^*_{ni} - x^T_{ni}(\widehat{\beta}^*_n - \widehat{\beta}_n)$$

where $\widehat{\beta}^*_n$ is defined as the M-estimator for the "parameter" $\widehat{\beta}_n$ in the linear model (3.3). Throughout the remaining part of this section we assume that a symmetric function $\psi = \psi(-x) = -\psi(x)$ is used for the calculation of the M-estimates. As a consequence we obtain bootstrap observations under the null hypothesis.
The bootstrap version of the empirical distribution function of the residuals is now given by

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

and the empirical distribution function of the symmetrized (unobservable) errors is denoted by

$$F^*_n(y) = \frac{1}{n}\sum_{i=1}^n I\{v_{ni}\varepsilon_{ni} \le y\}.$$

Finally, we define the analogue of the empirical symmetry process in the bootstrap setting, that is

$$\widehat{S}^*_n(y) = \sqrt{n}(\widehat{F}^*_n(y) - 1 + \widehat{F}^*_n(-y-)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\widehat{\varepsilon}^*_{ni} \le y\} - I\{-\widehat{\varepsilon}^*_{ni} \le y\}\Big).$$

Note that no centering is necessary here because of the identity

$$E[F_n^*(y) - 1 + F_n^*(-y-) \mid \mathcal{Y}_n] = 0.$$

The proof of the following auxiliary result, which gives an asymptotic expansion for the process $\widehat{S}_n^*(y)$, is deferred to the appendix.

Proposition 3.2 *Uniformly with respect to $y \in I\!R$ the expansion*

$$\widehat{S}_n^*(y) = \sqrt{n}(\widehat{F}_n^*(y) - 1 + \widehat{F}_n^*(-y-)) \;=\; G_n^*(y) + o_P(1)$$

is valid, where the process G_n^* is defined by

$$G_n^*(y) = \sqrt{n}\Big(F_n^*(y) - 1 + F_n^*(-y-) + (f(y) + f(-y))\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)\Big).$$

In order to establish the consistency of the proposed bootstrap procedure we require the following additional assumptions for the estimator $\widehat{\beta}_n^*$ from the bootstrap sample

$$||(X_n^T X_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n)|| \;=\; O_P(1) \tag{3.4}$$

$$\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) \;=\; \frac{1}{\int f\, d\psi}\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}\psi(v_{ni}\varepsilon_{ni}) + o_P(\frac{1}{\sqrt{n}}). \tag{3.5}$$

Remark 3.3 *(a)* Note that the conditions (3.4) and (3.5) are fulfilled if $\widehat{\beta}_n$ and $\widehat{\beta}_n^*$ are defined as least squares estimators, that is $\psi(x) = x$. In this case we have

$$\begin{aligned}(X_n^T X_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n) \;&=\; (X_n^T X_n)^{-1/2}\sum_{i=1}^n x_{ni}\varepsilon_{ni}^* \\ &=\; (X_n^T X_n)^{-1/2}\sum_{i=1}^n x_{ni}(v_{ni}\varepsilon_{ni} - v_{ni}x_{ni}^T(\widehat{\beta}_n - \beta)),\end{aligned}$$

and therefore (with Example 2.2)

$$\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) \;=\; \frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}v_{ni}\varepsilon_{ni} - Z_n,$$

where the random variable Z_n is defined by

$$Z_n \;=\; \frac{1}{n}\sum_{j=1}^n\sum_{k=1}^n v_{nk}\varepsilon_{nj}\widetilde{x}_{nk}x_{nk}^T(X_n^T X_n)^{-1}x_{nj} \;=\; o_P(\frac{1}{\sqrt{n}}).$$

The last equality follows from $E[Z_n] = 0$ and a calculation of the variance, which gives $\mathrm{Var}(Z_n) = o(\frac{1}{n})$.

(b) If ψ is continuously differentiable then the condition (3.5) follows from the assumption (compare with (2.7)]

$$(X_n^T X_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n) = \frac{1}{\int f\, d\psi}(X_n^T X_n)^{-1/2} \sum_{i=1}^n x_{ni}\psi(\varepsilon_{ni}^*) + o_P(1) \tag{3.6}$$

by means of a Taylor expansion

$$\psi(\varepsilon_{ni}^*) = \psi(v_{ni}\widehat{\varepsilon}_{ni}) = \psi(v_{ni}\varepsilon_{ni} - v_{ni}x_{ni}^T(\widehat{\beta}_n - \beta)) = \psi(v_{ni}\varepsilon_{ni}) - v_{ni}x_{ni}^T(\widehat{\beta}_n - \beta)\psi'(\xi_{ni})$$

and an application of the expansion (2.7).

(c) It is notable that the choice of a symmetric function ψ in the estimating equations for the M-estimator is essential for the appropriate performance of the bootstrap procedure. As a direct consequence we obtain $E[\psi(\varepsilon_{ni}^*) \,|\, \mathcal{Y}_n] = E[\psi(v_{ni}\widehat{\varepsilon}_{ni}) \,|\, \widehat{\varepsilon}_{ni}] = (\psi(\widehat{\varepsilon}_{ni}) + \psi(-\widehat{\varepsilon}_{ni}))/2 = 0$ and therefore we do not have to center the definition of M-estimator in the bootstrap setting, in the contrast to the work of Koul and Lahiri (1994) in this context. The dominating parts of the expansion (3.5) and (3.6) have expectation zero, conditionally on the sample $\mathcal{Y}_n$.

Theorem 3.4 *Assume that the density f of the error distribution is Lipschitz continuous of order $\gamma \geq 1/2$, then the process*

$$\widehat{S}_n^*(y) = \sqrt{n}\left(\widehat{F}_n^*(y) - 1 + \widehat{F}_n^*(-y-)\right),$$

conditioned on $\mathcal{Y}_n$, converges weakly in $D[-\infty, \infty]$ to a centered Gaussian process $\widetilde{G}$ with covariance kernel

$$\begin{aligned}\mathrm{Cov}(\widetilde{G}(x), \widetilde{G}(y)) = {} & 2\widetilde{F}(-(|x| \vee |y|)) + 2\widetilde{f}(x)(U(y) + U(-y))m^T\Sigma^{-1}m/(\textstyle\int f(t)\, d\psi(t)) \\ & + 2\widetilde{f}(y)(U(x) + U(-x))m^T\Sigma^{-1}m/(\textstyle\int f(t)\, d\psi(t)) \\ & + 4\widetilde{f}(x)\widetilde{f}(y)m^T\Sigma^{-1}m\sigma^2/(\textstyle\int f(t)\, d\psi(t))^2\end{aligned}$$

in probability, where the function $\widetilde{f}(y) = (f(y) + f(-y))/2$ denotes the density of the symmetrized error and $\widetilde{F}(y) = (F(y) + 1 - F(-y))/2$ the corresponding distribution function. Under the null hypothesis of symmetry $H_0 : f = \widetilde{f}$ we have $\mathrm{Cov}(\widetilde{G}(x), \widetilde{G}(y)) = \mathrm{Cov}(G(x), G(y))$ where the process G and its covariance kernel are defined in Theorem 3.1.

The proof of the theorem is deferred to the appendix. From the theorem the consistency of a test for symmetry based on the wild bootstrap procedure can be deduced as follows. Let T_n denote the test statistic based on a continuous functional of the process $\widehat{S}_n$ and let T_n^* denote the corresponding bootstrap statistic based on $\widehat{S}_n^*$. If t_n is the realization of the test statistic T_n based on the sample $\mathcal{Y}_n$ then a level α–test is obtained by rejecting symmetry 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

$$P(T_n^* > c^*_{1-\alpha} \mid \mathcal{Y}_n) \;=\; \alpha. \tag{3.7}$$

¿From Theorem 3.4 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 simulation in Section 4.

4 Finite sample properties

In this section we investigate the finite sample properties of the bootstrap test proposed in Section 3. We consider the classical one-way-layout in analysis of variance and a quadratic regression model on the interval $[0, 1]$. If T_n is a functional of the empirical process (1.3), then the null hypothesis is rejected if

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

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

4.1 Testing for a symmetric error distribution in a regression model

In this example we illustrate the performance of the bootstrap procedure in the problem of testing for a symmetric error distribution in a quadratic regression model, that is

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

where the explanatory variables $x_{ni} = i/n \quad (i = 1, \ldots, n)$ correspond to the uniform design. The finite sample performance of the Cramér-von-Mises type test statistic

$$T_n = \int_{\mathbb{R}} \Big(\widehat{F}_n(x) - 1 + \widehat{F}_n(-x-)\Big)^2 d\widehat{H}_n(x) \tag{4.3}$$

is investigated, where $\widehat{F}_n$ denotes the empirical distribution function of the residuals $\widehat{\varepsilon}_{n1}, \ldots, \widehat{\varepsilon}_{nn}$ and $\widehat{H}_n(x) = \widehat{F}_n(x) - \widehat{F}_n(-x-)$ denotes the empirical distribution function of the absolute values of the residuals $|\widehat{\varepsilon}_{n1}|, \ldots, |\widehat{\varepsilon}_{nn}|$ [see Shorack and Wellner (1986), p. 747]. The null hypothesis is rejected by the bootstrap test if the inequality (4.1) is satisfied, where the resampling scheme is described in Section 3. The parameters in model (4.2) are given by $\beta_1 = \beta_2 = 1$, while the error distribution is given by

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

The results are depicted in Table 4.1 and show a reasonable approximation of the nominal level in all cases. The alternative of non-symmetry is detected with high probability, even in the case $k = 4$, which is rather close to symmetry.
It might be of interest to compare these results with the test based on kernel estimators, which was presented by Ahmad and Li (1997). In their simulation study these authors considered the problem of testing the symmetry of an i.i.d. sample of observations and a comparison is not directly possible. For this reason we also included the symmetry test of Ahmad and Li (1997) in our study. Our numerical results show that the approximation of the nominal level of the test of Ahmad and Li (1997) is not too accurate. This phenomenon was also observed by other authors in the context of testing model assumptions [see e.g. Härdle and Mammen (1993), Hjellvik and Tjøstheim (1996) or Fan and Linton (2003)]. Therefore we implemented a bootstrap version of Ahmad and Li's (1997) test. We investigated the performance of the test in the situation discussed in the previous paragraph. The test requires a specification of a bandwidth, say a, which was chosen as $a = \widehat{\sigma}^2/n^{1/2}$, which was recommended by Ahmad and Li (1997). Here $\widehat{\sigma}^2$ is the least squares estimator of the variance in the linear regression model (4.2). The rejection probabilities of Ahmad and Li's (1997) test for the error distributions given in (4.4) are presented in Table 4.2. We observe that in all cases under consideration the test based on the empirical process approach yields substantially higher rejection probabilities than the test of Ahmad and Li (1997). Moreover, this test does not rely on a smoothing parameter.

n	25			50			75			100		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	.031	.055	.109	.034	.058	.106	.029	.586	.105	.029	.056	.102
df_1	.728	.806	.888	.987	.993	.995	1.000	1.000	1.000	1.000	1.000	1.000
df_2	.495	.618	.738	.921	.957	.980	.993	.997	.999	.999	1.000	1.000
df_3	.367	.462	.561	.795	.859	.920	.958	.977	.989	.993	.996	.997
df_4	.275	.373	.493	.674	.773	.853	.880	.925	.961	.968	.983	.994

Table 4.1: *Simulated rejection probabilities of the bootstrap test for a symmetric error distribution. 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) normalized such that $E[\varepsilon_{ni}] = 0,\ E[\varepsilon_{ni}^2] = 1$.*

n	25			50			75			100		
α	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	.034	.058	.102	.032	.057	.105	.031	.053	.109	.028	.054	.104
df_1	.681	.717	.782	.929	.947	.967	.992	.994	.998	.996	.996	1.000
df_2	.479	.556	.658	.748	.809	.882	.906	.930	.965	.970	.979	.996
df_3	.346	.412	.510	.579	.649	.753	.755	.806	.870	.878	.913	.954
df_4	.251	.304	.418	.459	.522	.627	.645	.697	.808	.745	.811	.868

Table 4.2: *Simulated rejection probabilities of Ahmad and Li's (1997) test for symmetry in the model (4.2). The error distribution is normal* (df_0) *and chi-square with k degrees of freedom* (df_k) *normalized such that* $E[\varepsilon_{ni}] = 0,\ E[\varepsilon_{ni}^2] = 1$.

4.2 Testing for symmetry of the error distribution 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

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \qquad i = 1, \dots, p; j = 1, \dots, n_i; n = \sum_{i=1}^{p} n_i. \tag{4.5}$$

We consider the case of $p = 4$ factors and $n = 20$ and $n = 40$ total observations, where the sample sizes for the different groups are

$$\begin{aligned} A &: n_1 = n_2 = n_3 = n_4 = \frac{n}{4}; \\ B &: n_1 = \frac{3n}{20}; n_2 = \frac{n}{4}; n_3 = \frac{n}{4}; n_4 = \frac{7n}{20}. \end{aligned} \tag{4.6}$$

The finite sample properties of the test for symmetry of the error distribution in the model (4.5) are of interest, where the statistic T_n is defined by the Kolmogorov-Smirnov functional, that is

$$T_n = \sup_{x \in \mathbb{R}} |\widehat{F}_n(x) - 1 + \widehat{F}_n(-x-)|.$$

The variance of the error in the model (4.5) is given by $\sigma^2 = 1$, while the error distributions are defined in (4.4). The simulated rejection probabilities are shown in Table 4.3.

The results are similar as in Example 4.1. We observe a reasonable approximation of the nominal level, even for the sample size $n = 20$, and deviations from the null hypothesis are detected with high probability. It is interesting to note that 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 level is usually not so accurate.

	$n=20$						$n=40$					
	A			B			A			B		
	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
df_0	0.020	0.042	0.092	0.022	0.048	0.088	0.025	0.045	0.091	0.026	0.050	0.089
df_1	0.327	0.422	0.572	0.303	0.394	0.528	0.878	0.935	0.965	0.844	0.903	0.946
df_2	0.190	0.259	0.399	0.165	0.242	0.379	0.580	0.690	0.798	0.608	0.716	0.816
df_3	0.126	0.192	0.308	0.135	0.198	0.319	0.431	0.541	0.663	0.422	0.532	0.656
df_4	0.111	0.174	0.280	0.098	0.153	0.259	0.313	0.407	0.540	0.338	0.430	0.564

Table 4.3: *Simulated rejection probabilities of the bootstrap test (4.1) for a symmetric error distribution in the one-way-layout (4.5) with four factors for various distributions defined in (4.4) and designs given in (4.6).*

Acknowledgements. The authors are grateful to Prof. H. Koul for some useful comments on related literature 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).

5 Appendix: Proofs

5.1 Proof of Theorem 3.1

The proof follows by standard arguments and for this reason we will only indicate the main steps. From Theorem 6.2.1, Koul (2002, p. 232), we have the following decomposition of the process, uniformly with respect to $y \in I\!R$,

$$\begin{aligned}\widehat{S}_n(y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{\widehat{\varepsilon}_{ni} \le y\} - I\{-\widehat{\varepsilon}_{ni} \le y\} - F(y) + 1 - F(-y)\Big) \\ &= \sqrt{n}\Big(F_n(y) - 1 + F_n(-y-) - F(y) + 1 - F(-y) \\ &\quad + (f(y) + f(-y))\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n - \beta)\Big) + o_P(1).\end{aligned}$$

With the expansion (2.7) of the M-estimator we obtain

$$\widehat{S}_n(y) = G_n(y) + o_P(1),$$

uniformly with respect to $y \in \mathbb{R}$, where the process G_n is defined by

$$G_n(y) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{\varepsilon_{ni} \leq y\} - I\{-\varepsilon_{ni} \leq y\} - F(y) + 1 - F(-y) \\ + (f(y) + f(-y)) \frac{1}{\int f \, d\psi} \widetilde{x}_{ni} \psi(\varepsilon_{ni}) \Big).$$

A straightforward calculation of the covariances gives

$$\mathrm{Cov}(G_n(x), G_n(y)) = \mathrm{Cov}(G(x), G(y)) + o(1),$$

where $\mathrm{Cov}(G(x), G(y))$ is defined in Theorem 3.1.
The weak convergence of the finite dimensional distributions of the process G_n is now obtained by an application of the Cramér–Wold device and the Lindeberg condition. Finally, tightness of the process G_n can be shown in terms of asymptotic equicontinuity, that is for all $\epsilon > 0$

$$\lim_{\delta \searrow 0} \limsup_{n \to \infty} P\Big(\sup_{\substack{x,y \in \bar{\mathbb{R}} \\ |x-y| \leq \delta}} \Big| G_n(x) - G_n(y) \Big| > \epsilon \Big) = 0.$$

This follows from tightness of the standard empirical process and from uniform continuity of the density f. □

5.2 Proof of Proposition 3.2.

For vectors $a, b \in \mathbb{R}^p$ and real values $x \in \mathbb{R}$, we introduce the notation

$$Y(x, a, b) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(I\{v_{ni}\varepsilon_{ni} \leq x + v_{ni} c_{ni}^T a + c_{ni}^T b\} - P(v_{ni}\varepsilon_{ni} \leq x + v_{ni} c_{ni}^T a + c_{ni}^T b) \Big),$$

where we assume for the (nonrandom) weights $c_{ni} \in \mathbb{R}^p$

$$\text{(4.7)} \qquad \max_{i=1,\ldots,n} ||c_{ni}|| = o(1) \quad \text{and} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ||c_{ni}|| = O(1).$$

The proof of the proposition (3.2) requires the following auxiliary result.

Lemma 5.1 *For each $L > 0$:*

$$\sup_{x \in \mathbb{R}, ||a|| \leq L, ||b|| \leq L} |Y(x, a, b) - Y(x, 0, 0)| = o_P(1).$$

Proof. The proof can be done in a similar manner as the proofs of Lemmata 2.3.2. – 2.3.4. (Koul, 2002, p. 53) and is therefore omitted. □

Now we apply Lemma 5.1 with the definitions $c_{ni}^T = x_{ni}^T(X_n^TX_n)^{-1/2}$, $a = (X_n^TX_n)^{1/2}(\widehat{\beta}_n - \beta)$ and $b = (X_n^TX_n)^{1/2}(\widehat{\beta}_n^* - \widehat{\beta}_n)$ and obtain uniformly with respect to $y \in I\!R$,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\widehat{S}_n^*(y) - (F_n^*(y) - 1 + F_n^*(-y-)) \\
= \;& \frac{1}{n}\sum_{i=1}^n \Big(I\{v_{ni}\varepsilon_{ni} \le y + v_{ni}x_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)\} - I\{v_{ni}\varepsilon_{ni} \le y\} \\
&\qquad - I\{-v_{ni}\varepsilon_{ni} \le y - v_{ni}x_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)\} + I\{-v_{ni}\varepsilon_{ni} \le y\}\Big) \\
= \;& \frac{1}{n}\sum_{i=1}^n \Big(P(v\varepsilon \le y + vx_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) \mid \mathcal{Y}_n) - P(v\varepsilon \le y) \\
&\qquad - P(-v\varepsilon \le y - vx_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) \mid \mathcal{Y}_n) + P(-v\varepsilon \le y)\Big) \\
&+ o_P(\frac{1}{\sqrt{n}}) \\
= \;& \frac{1}{2n}\sum_{i=1}^n \Big(F(y + x_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) - F(y) \\
&- F(-y + x_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) + F(-y) \\
&+ F(-y + x_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) - F(-y) \\
&- F(y + x_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) + F(y)\Big) \\
&+ o_P(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Here ε and v denote independent random variables, independent of $\mathcal{Y}_n$, such that $P(v = -1) = P(v = 1) = 1/2$ and ε has density f. By an application of Lemma 2.3.1. of Koul (2002) we further have, uniformly with respect to $y \in I\!R$,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\widehat{S}_n^*(y) - (F_n^*(y) - 1 + F_n^*(-y-)) \\
= \;& \frac{1}{2n}\sum_{i=1}^n \Big(f(y)(x_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) - f(-y)(x_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) \\
&+ f(-y)(x_{ni}^T(\widehat{\beta}_n - \beta) + x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n)) - f(y)(x_{ni}^T(\widehat{\beta}_n - \beta) - x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n))\Big) \\
&+ o_P(\frac{1}{\sqrt{n}}) \\
= \;& (f(y) + f(-y))\frac{1}{n}\sum_{i=1}^n x_{ni}^T(\widehat{\beta}_n^* - \widehat{\beta}_n) + o_P(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Finally we note that the conditions for an application of Lemma 2.3.1 of Koul (2002, p. 52) are satisfied because of the uniform continuity of the density f and assumptions (2.1), (2.6) and (3.4). $\square$

5.3 Proof of Theorem 3.4

Using the expansion in Proposition 3.2 and assumption (3.5) we have uniformly with respect to $y \in I\!R$

$$\widehat{S}_n^*(y) = \widetilde{G}_n^*(y) + o_P(1)$$

where the process $\widetilde{G}_n^*$ is defined by

$$\begin{aligned}\widetilde{G}_n^*(y) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big(I\{v_{ni}\varepsilon_{ni} \le y\} - I\{-v_{ni}\varepsilon_{ni} \le y\} + 2\widetilde{f}(y)\frac{1}{\int f\,d\psi}\widetilde{x}_{ni}\psi(v_{ni}\varepsilon_{ni})\Big)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n v_{ni}\Big(I\{\varepsilon_{ni} \le y\} - I\{-\varepsilon_{ni} \le y\} + 2\widetilde{f}(y)\frac{1}{\int f\,d\psi}\widetilde{x}_{ni}\psi(\varepsilon_{ni})\Big).\end{aligned}$$

The last equality follows with the definition of the random variables v_{ni} and the symmetry of the function ψ. Now we obtain for the conditional covariance of the asymptotically equivalent process $\widetilde{G}_n^*$

$$\begin{aligned}E[\widetilde{G}_n^*(x)\widetilde{G}_n^*(y)\,|\,\mathcal{Y}_n] = \frac{1}{n}\sum_{i=1}^n \Big(&\frac{1}{2}(I\{\varepsilon_{ni} \le x\wedge y\} + I\{-\varepsilon_{ni} \le x\wedge y\})\\ &- \frac{1}{2}(I\{\varepsilon_{ni} \le x\} + I\{-\varepsilon_{ni} \le x\})\frac{1}{2}(I\{\varepsilon_{ni} \le y\} + I\{-\varepsilon_{ni} \le y\})\\ &+ \frac{1}{2}(I\{\varepsilon_{ni} \le (-x)\wedge y\} + I\{-\varepsilon_{ni} \le (-x)\wedge y\})\\ &- \frac{1}{2}(I\{\varepsilon_{ni} \le -x\} + I\{-\varepsilon_{ni} \le -x\})\frac{1}{2}(I\{\varepsilon_{ni} \le y\} + I\{-\varepsilon_{ni} \le y\})\\ &+ \frac{1}{2}(I\{\varepsilon_{ni} \le x\wedge(-y))\} + I\{-\varepsilon_{ni} \le x\wedge(-y))\})\\ &- \frac{1}{2}(I\{\varepsilon_{ni} \le x\} + I\{-\varepsilon_{ni} \le x\})\frac{1}{2}(I\{\varepsilon_{ni} \le -y\} + I\{-\varepsilon_{ni} \le -y\})\\ &+ \frac{1}{2}(I\{\varepsilon_{ni} \le (-x)\wedge(-y))\} + I\{-\varepsilon_{ni} \le (-x)\wedge(-y))\})\\ &- \frac{1}{2}(I\{\varepsilon_{ni} \le -x\} + I\{-\varepsilon_{ni} \le -x\})\frac{1}{2}(I\{\varepsilon_{ni} \le -y\} + I\{-\varepsilon_{ni} \le -y\})\Big)\\ &+ 4\widetilde{f}(y)\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}\psi(\varepsilon_{ni})\frac{1}{2}(I\{\varepsilon_{ni} \le x\} - I\{-\varepsilon_{ni} \le x\})/(\textstyle\int f\,d\psi)\\ &+ 4\widetilde{f}(x)\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}\psi(\varepsilon_{ni})\frac{1}{2}(I\{\varepsilon_{ni} \le y\} - I\{-\varepsilon_{ni} \le y\})/(\textstyle\int f\,d\psi)\\ &+ 4\widetilde{f}(x)\widetilde{f}(y)\frac{1}{n}\sum_{i=1}^n \widetilde{x}_{ni}^2\psi^2(\varepsilon_{ni})/(\textstyle\int f\,d\psi)^2 \qquad \text{(a.s.)}\end{aligned}$$

and by the weak law of large numbers this random variable converges in probability to the covariance kernel $\mathrm{Cov}(\widetilde{G}(x), \widetilde{G}(y))$ defined in Theorem 3.4.

In a second step we show the conditional weak convergence in probability of the finite dimensional distributions applying the Cramér–Wold device and Lindeberg's condition. For this let $a_1, \dots, a_k, y_1, \dots, y_k \in \mathbb{R}$ denote arbitrary constants and consider the linear combination

$$\sum_{j=1}^{k} a_j \widetilde{G}_n^*(y_j) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{ni} A_{ni}$$

with

$$A_{ni} \;=\; \sum_{j=1}^{k} a_j \Big(I\{\varepsilon_{ni} \le y_j\} - I\{-\varepsilon_{ni} \le y_j\} + 2 \frac{\widetilde{f}(y_j)}{\int f \, d\psi} \widetilde{x}_{ni} \psi(\varepsilon_{ni}) \Big).$$

It is easy to see that

$$A_{ni} \;\le\; c_1 + c_2 \psi(\varepsilon_{ni})$$

for some suitable constants c_1 and c_2. For $\delta > 0$ we consider the quantity

$$L_n(\delta) \;=\; \frac{1}{n} \sum_{i=1}^{n} E\Big[(A_{ni} v_{ni})^2 I\{|A_{ni} v_{ni}| \ge \sqrt{n}\delta\} \Big| \mathcal{Y}_n \Big].$$

Observing that $|v_{ni}| = 1$, it follows for each $K > 0$ and some constants d_1, d_2, d_3, d_4

$$\limsup_{n\to\infty} L_n(\delta) \;\le\; \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (d_1 + d_2 \psi^2(\varepsilon_{ni})) I\{d_3 + d_4 |\psi(\varepsilon_{ni})| \ge K\}.$$

Applying the strong law of large numbers for $n \to \infty$ and considering $K \to \infty$ we obtain the assertion,

$$\limsup_{n\to\infty} L_n(\delta) \;=\; 0 \quad \text{with probability 1.}$$

Finally we show conditional tightness of the process $\widetilde{G}_n^*$ in probability. To this end we apply the continuous quantile transformation $t = F(y)$ and consider the process

$$H_n^*(t) \;=\; \widetilde{G}_n^*(F^{-1}(t)) \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_{ni} h(\varepsilon_{ni}, t),$$

in $D[0,1]$, where $t \in [0,1]$ and

$$h(\varepsilon_{ni}, t) \;=\; I\{F(\varepsilon_{ni}) \le t\} - I\{F(-\varepsilon_{ni}) \le t\} + 2\widetilde{f}(F^{-1}(t)) \frac{1}{\int f \, d\psi} \widetilde{x}_{ni} \psi(\varepsilon_{ni}).$$

We obtain for $0 \le s \le t \le 1$ with suitable constants $c, c' > 0$

$$E\Big[(H_n^*(t) - H_n^*(s))^4 \Big| \mathcal{Y}_n \Big] \;=\; \frac{3}{n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \ne i}}^{n} (h(\varepsilon_{ni}, t) - h(\varepsilon_{ni}, s))^2 (h(\varepsilon_{nj}, t) - h(\varepsilon_{nj}, s))^2$$

$$
\begin{aligned}
&+ \frac{1}{n^2}\sum_{i=1}^n (h(\varepsilon_{ni},t) - h(\varepsilon_{ni},s))^4 \\
\le\ & 3\Big(\frac{1}{n}\sum_{i=1}^n (h(\varepsilon_{ni},t) - h(\varepsilon_{ni},s))^2\Big)^2 \\
\le\ & c\Big(\frac{1}{n}\sum_{i=1}^n (I\{s < F(\varepsilon_{ni}) \le t\} + I\{s < F(-\varepsilon_{ni}) \le t\} \\
&\qquad + (\widetilde{f}(F^{-1}(t)) - \widetilde{f}(F^{-1}(s)))^2 \widetilde{x}_{ni}^2 \psi^2(\varepsilon_{ni}))\Big)^2 \\
\le\ & c'\Big(M_n(t) - M_n(s)\Big)^2
\end{aligned}
$$

where the last inequality follows from the Lipschitz continuity of f and we introduced the definition

$$
M_n(t) \ = \ \frac{1}{n}\sum_{i=1}^n (I\{F(\varepsilon_{ni}) \le t\} + I\{F(-\varepsilon_{ni}) \le t\} + F^{-1}(t)\widetilde{x}_{ni}^2 \psi^2(\varepsilon_{ni})).
$$

$(M_n(t))_{n\in\mathbb{N}}$ is a sequence of random functions, increasing in $t \in [0,1]$, which converges uniformly in $t \in [0,1]$ almost surely to the increasing function $M(t)$ defined by

$$
M(t) \ = \ P(F(\varepsilon_{n1}) \le t) + P(F(-\varepsilon_{ni}) \le t) + F^{-1}(t) m^T \Sigma^{-1} m E[\psi^2(\varepsilon_{n1})].
$$

This proves tightness of the process H_n^*, conditionally on the sample $\mathcal{Y}_n$, with probability 1 and completes the proof of Theorem 3.4. □

References

I.A. Ahmad, Q. Li (1997). *Testing symmetry of an unknown density function by kernel method.* Nonparam. Statistics 7, 279-293.

S. Aki (1981). *Asymptotic distribution of a Cramér-von-Mises type statistic for testing symmetry when the center ist estimated.* Ann. Inst. Statist. Math. 33, 1–14.

P. K. Bhattacharya, J. L. Gastwirth, A. L. Wright (1982). *Two modified Wilcoxon tests for symmetry about an unknown location parameter.* Biometrika 69, 377–382.

P. J. Bickel (1982). *On adaptive estimation.* Ann. Stat. 10, 647-671.

Y. Fan, R. Gencay (1999). *A consistent nonparametric test of symmetry in linear regression models.* J. Amer. Statist. Assoc. 90, 551-557.

Y. Fan, O. Linton (2003). *Some higher-order theory for a consistent non-parametric specification test.* J. Statist. Plann. Inf. 109, 125-251.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986). *Robust statistics. The approach based on influence functionals.* Wiley, New York.

W. Härdle, E. Mammen (1993).*Comparing nonparametric versus parametric regression fits.* Ann. Statist. 21, 1926-1947.

V. Hjellvik, D. Tjøstheim (1996). *Nonparametric statistics for testing linearity and serial dependence.* J. Nonparam. Stat. 6, 223-251.

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

H. L. Koul, S. N. Lahiri (1994).*On bootstrapping M-estimated residual processes in multiple linear regression models.* J. Multiv. Anal. 49, 255-265.

J. A. Koziol (1985). *A note on testing symmetry with estimated parameters.* Statist. Prob. Letters 3, 227–230.

Z. Psaradakis (2003). *A bootstrap test for symmetry of dependent data based on a Komogorov-Smirnov type statistic.* Comm. Statist., Ser. B, Simul. 8 Comput. 32, 113–126.

M. Rosenblatt (1975). *A quadratic measure of deviation and a test of independence.* Ann. Statist. 3, 1–14.

E. F. Schuster, R. C. Barker (1987). *Using the bootstrap in testing symmetry versus asymmetry.* Comm. Statist., Ser. B, Simul. & Comput. 16, 69–84.

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

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

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

H. Witting, U. Müller–Funk (1995). *Mathematische Statistik II.* B. G. Teubner, Stuttgart.