

# Specification tests in nonparametric regression

John H.J. Einmahl<sup>a</sup>, Ingrid Van Keilegom<sup>b,\*</sup>

<sup>a</sup>*Department of Econometrics & OR and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*

<sup>b</sup>*Institut de Statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium*

Available online 22 August 2007

---

## Abstract

Consider the location-scale regression model  $Y = m(X) + \sigma(X)\varepsilon$ , where the error  $\varepsilon$  is independent of the covariate  $X$ , and  $m$  and  $\sigma$  are smooth but unknown functions. We construct tests for the validity of this model and show that the asymptotic limits of the proposed test statistics are distribution free. We also investigate the finite sample properties of the tests through a simulation study, and we apply the tests in the analysis of data on food expenditures.

© 2007 Elsevier B.V. All rights reserved.

*JEL classification:* C12; C14; C52

*Keywords:* Bootstrap; Empirical process; Location-scale regression; Model diagnostics; Nonparametric regression; Test for independence; Weak convergence

---

## 1. Introduction

Consider the nonparametric location-scale regression model

$$Y = m(X) + \sigma(X)\varepsilon, \quad (1.1)$$

where  $Y$  is the variable of interest,  $X$  is a covariate, the error  $\varepsilon$  is independent of  $X$ , and  $m$  and  $\sigma$  are smooth but unknown location and scale curves, respectively. The location curve  $m$  is not restricted to the conditional mean  $E(Y|X = \cdot)$ , but can equally well represent the conditional trimmed mean curve, the median curve, etc. Similarly the scale curve  $\sigma$  is not restricted to the conditional standard deviation. Let  $(X_1, Y_1), \dots, (X_n, Y_n)$  be  $n$  independent replications of  $(X, Y)$ .

This model has been studied by many authors over the last years. The estimation of this model has been considered in Akritas and Van Keilegom (2001), Van Keilegom and Veraverbeke (2002), Cheng (2004), Müller et al. (2004a,b), among others, whereas Dette et al. (2007), Neumeyer et al. (2006), Van Keilegom et al. (2007) and Pardo-Fernández et al. (2007) studied various testing problems under this model.

Although the independence of the error and the covariate is a quite weak and common assumption, in several applications, especially in the recent econometrics literature, it is considered too strong as an assumption. An appropriate testing procedure for the validity of this model is therefore in demand. In

---

\*Corresponding author.

*E-mail addresses:* [j.h.j.einmahl@uvt.nl](mailto:j.h.j.einmahl@uvt.nl) (J.H.J. Einmahl), [vankeilegom@stat.ucl.ac.be](mailto:vankeilegom@stat.ucl.ac.be) (I. Van Keilegom).

Einmahl and Van Keilegom (2007) a difference-based testing approach is proposed for the homoscedastic model  $Y = m(X) + \varepsilon$ , with  $\varepsilon$  independent of  $X$ . In the present paper we consider another approach, applicable to the more general model (1.1). Although model (1.1) has been used and studied frequently, a procedure for testing the validity of this model is, to the best of our knowledge, not available. Our approach is based on the estimation of the unobserved errors, and we use Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling type test statistics based on the estimated errors and the covariate to test the independence between the error and the covariate.

Observe that the tests developed in this paper can be easily adapted for testing the validity of the homoscedastic model  $Y = m(X) + \varepsilon$ , with  $\varepsilon$  independent of  $X$ . This is also a very relevant testing problem; we will pay attention to it in Sections 3 and 4. Also note that the results in this paper will be presented for random design, but can be readily adapted to fixed design. In that case, interest lies in the fact whether or not the error terms  $\varepsilon_1, \dots, \varepsilon_n$  are identically distributed.

The paper is organized as follows. In the next section, we will construct the test statistics and present the main asymptotic results, including the asymptotic distribution of the test statistics. In Section 3 some simulation results will be shown. The analysis of data on food expenditures is carried out in Section 4. The assumptions and some technical derivations are deferred to the Appendix.

## 2. Main results

Define  $F_X(x) = P(X \leq x)$ ,  $F_\varepsilon(y) = P(\varepsilon \leq y)$ ,  $F(y|x) = P(Y \leq y|X = x)$  and  $F_{X,\varepsilon}(x, y) = P(X \leq x, \varepsilon \leq y)$ , and let  $D_X$  be the support of the covariate  $X$ . The probability density functions of these distributions will be denoted with lower case letters. Assume that  $m$  and  $\sigma$  are, respectively, a location and scale functional. This means that we can write  $m(x) = T(F(\cdot|x))$  and  $\sigma(x) = S(F(\cdot|x))$  for some functionals  $T$  and  $S$ , such that

$$T(F_{aY+b}(\cdot|x)) = aT(F_Y(\cdot|x)) + b \quad \text{and} \quad S(F_{aY+b}(\cdot|x)) = aS(F_Y(\cdot|x)),$$

for all  $a \geq 0$  and  $b \in \mathbb{R}$ , where  $F_{aY+b}(\cdot|x)$  denotes the conditional distribution of  $aY + b$  given  $X = x$  (see also Huber, 1981, pp. 59, 202). It follows (see e.g. Van Keilegom, 1998, Proposition 5.1) that if model (1.1) holds for a certain location functional  $m$  and scale functional  $\sigma$ , then it holds for all location functionals  $\tilde{m}$  and scale functionals  $\tilde{\sigma}$ , in the sense that the new error  $\tilde{\varepsilon} = (Y - \tilde{m}(X))/\tilde{\sigma}(X)$  is still independent of  $X$ . Hence, we can and will assume that  $m$  and  $\sigma^2$  are given by

$$m(x) = \int_0^1 F^{-1}(s|x)J(s) ds, \quad \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s) ds - m^2(x), \tag{2.1}$$

where  $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$  is the quantile function of  $Y$  given  $x$  and  $J$  is a given score function satisfying  $\int_0^1 J(s) ds = 1$  (e.g., the choice  $J \equiv 1$  leads to  $m(x) = E(Y|X = x)$  and  $\sigma^2(x) = \text{Var}(Y|X = x)$ ).

Our tests will be based on the difference  $\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)$  for appropriate estimators  $\hat{F}_X$ ,  $\hat{F}_{\hat{\varepsilon}}$  and  $\hat{F}_{X,\hat{\varepsilon}}$  of  $F_X$ ,  $F_\varepsilon$  and  $F_{X,\varepsilon}$ , respectively. First, let

$$\hat{F}_X(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x)$$

be the empirical distribution function of  $X$ . To estimate the distribution of  $\varepsilon$ , estimate  $m(x)$  and  $\sigma^2(x)$  by

$$\hat{m}(x) = \int_0^1 \hat{F}^{-1}(s|x)J(s) ds, \quad \hat{\sigma}^2(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s) ds - \hat{m}^2(x), \tag{2.2}$$

where

$$\hat{F}(y|x) = \sum_{i=1}^n W_i(x, a_n) I(Y_i \leq y) \tag{2.3}$$

is the Stone (1977) estimator and the  $W_i(x, a_n)$  ( $i = 1, \dots, n$ ) are the Nadaraya–Watson weights

$$W_i(x, a_n) = \frac{K((x - X_i)/a_n)}{\sum_{j=1}^n K((x - X_j)/a_n)}$$

(with  $K$  a given kernel function and  $(a_n)_{n \in \mathbb{N}}$  a bandwidth sequence). Now define  $\hat{\varepsilon}_i = \{Y_i - \hat{m}(X_i)\}/\hat{\sigma}(X_i)$  for the resulting residuals, and let

$$\hat{F}_{\hat{\varepsilon}}(y) = n^{-1} \sum_{i=1}^n I(\hat{\varepsilon}_i \leq y). \tag{2.4}$$

Finally,  $F_{X,\hat{\varepsilon}}(x, y)$  is estimated by

$$\hat{F}_{X,\hat{\varepsilon}}(x, y) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x, \hat{\varepsilon}_i \leq y).$$

To test the null hypothesis, we define the following test statistics:

$$T_{n,KS} = \sqrt{n} \sup_{x,y} |\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)|, \tag{2.5}$$

$$T_{n,CM} = n \iint (\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y))^2 d\hat{F}_X(x) d\hat{F}_{\hat{\varepsilon}}(y), \tag{2.6}$$

$$T_{n,AD} = n \iint \frac{(F_{X,\hat{\varepsilon}}(x, y) - F_X(x)\hat{F}_{\hat{\varepsilon}}(y))^2}{\hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)(1 - \hat{F}_{X-}(x))(1 - \hat{F}_{\hat{\varepsilon}-}(y))} d\hat{F}_X(x) d\hat{F}_{\hat{\varepsilon}}(y). \tag{2.7}$$

(For a distribution function  $F$ , we denote with  $F_-$  its left continuous version.)

These statistics are similar to the ones considered in [Hoeffding \(1948\)](#), [Blum et al. \(1961\)](#) and [de Wet \(1980\)](#) for testing independence between two random variables, except that here we have replaced the unknown errors  $\varepsilon_i$  by  $\hat{\varepsilon}_i$  ( $i = 1, \dots, n$ ). As we will see below, the limiting distribution of these test statistics is the same as in the case where the  $\varepsilon_i$  are observed, and hence the tests are asymptotically distribution free.

In the first theorem we obtain an i.i.d. representation for the difference  $\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)$ ,  $x \in D_X, y \in \mathbb{R}$  (weighted in an appropriate way), on which all three test statistics are based. Based on this result, the weak convergence will then be established. The assumptions mentioned below are given in the Appendix.

**Theorem 2.1.** *Assume (A), (K), (J) and (F). Then, under  $H_0$ , for  $0 \leq \beta < \frac{1}{2}$ ,*

$$\begin{aligned} & \sup_{\substack{x \in D_X \\ y \in \mathbb{R}}} \left| \frac{1}{N(x, y)^\beta} \{ \hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y) - (\hat{F}_{X,\varepsilon}(x, y) - F_{X,\varepsilon}(x, y)) - F_X(x)(\hat{F}_{\varepsilon}(y) - F_{\varepsilon}(y)) \right. \\ & \quad \left. - F_{\varepsilon}(y)(\hat{F}_X(x) - F_X(x)) \right| \\ & = o_P(n^{-1/2}), \end{aligned}$$

with

$$\hat{F}_{X,\varepsilon}(x, y) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x, \varepsilon_i \leq y),$$

$$\hat{F}_{\varepsilon}(y) = \frac{1}{n} \sum_{i=1}^n I(\varepsilon_i \leq y),$$

and

$$N(x, y) = F_X(x)F_{\varepsilon}(y)(1 - F_X(x))(1 - F_{\varepsilon}(y)).$$

**Proof.** Write

$$\begin{aligned}
 & N(x, y)^{-\beta} \{ [\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)] - [F_{X,\varepsilon}(x, y) - F_X(x)F_{\varepsilon}(y)] \} \\
 &= N(x, y)^{-\beta} \{ [\hat{F}_{X,\hat{\varepsilon}}(x, y) - F_{X,\varepsilon}(x, y)] - F_X(x)[\hat{F}_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)] - \hat{F}_{\hat{\varepsilon}}(y)[\hat{F}_X(x) - F_X(x)] \} \\
 &= N(x, y)^{-\beta} (\hat{F}_{X,\varepsilon}(x, y) - F_{X,\varepsilon}(x, y)) - F_X(x)(\hat{F}_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)) - F_{\hat{\varepsilon}}(y)(\hat{F}_X(x) - F_X(x)) \\
 &\quad + N(x, y)^{-\beta} n^{-1} \sum_{i=1}^n [I(X_i \leq x) - F_X(x)][I(\hat{\varepsilon}_i \leq y) - I(\varepsilon_i \leq y)] \\
 &\quad - N(x, y)^{-\beta} [\hat{F}_X(x) - F_X(x)][\hat{F}_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)]. \tag{2.8}
 \end{aligned}$$

From Lemma A.1 it follows that the second term on the right-hand side of (2.8) is equal to (using the notation of that lemma)

$$N(x, y)^{-\beta} (\hat{F}_X(x) - F_X(x))(F_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)) + o_P(n^{-1/2}),$$

uniformly in  $x$  and  $y$ . This term is  $o_P(n^{-1/2})$ , since by the Chibisov–O’Reilly theorem (see, e.g., [Shorack and Wellner, 1986](#), p. 462),

$$\sup_x \left| \frac{\hat{F}_X(x) - F_X(x)}{[F_X(x)(1 - F_X(x))]^\beta} \right| = O_P(n^{-1/2}),$$

and since

$$\sup_y \left| \frac{F_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)}{[F_{\varepsilon}(y)(1 - F_{\varepsilon}(y))]^\beta} \right| = o_P(1), \tag{2.9}$$

which can be shown in a similar way as in the beginning of the proof of Lemma A.1. Using again Lemma A.1, the third term of (2.8) can be written as

$$\begin{aligned}
 & N(x, y)^{-\beta} [\hat{F}_X(x) - F_X(x)][\{\hat{F}_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)\} + \{F_{\hat{\varepsilon}}(y) - F_{\varepsilon}(y)\}] + o_P(n^{-1/2}) \\
 &= o_P(n^{-1/2}),
 \end{aligned}$$

uniformly in  $x$  and  $y$ . Hence, the result follows.  $\square$

The next result follows readily from Theorem 2.1, by using standard empirical process theory.

**Theorem 2.2.** *Assume (A), (K), (J) and (F). Let  $W_0$  be a 4-sided tied-down Wiener process on  $[0, 1]^2$ , defined by  $W_0(u, v) = W(u, v) - uW(1, v) - vW(u, 1) + uvW(1, 1)$ ,  $u, v \in [0, 1]$ , where  $W$  is a standard bivariate Wiener process. Under  $H_0$ , for  $0 \leq \beta < \frac{1}{2}$ , the process*

$$\sqrt{n} \frac{\hat{F}_{X,\hat{\varepsilon}}(x, y) - \hat{F}_X(x)\hat{F}_{\hat{\varepsilon}}(y)}{N(x, y)^\beta}, \quad x \in D_X, \quad y \in \mathbb{R},$$

converges weakly to  $W_0(F_X(x), F_{\varepsilon}(y))/N(x, y)^\beta$ .

As a consequence, we find the limiting distribution of the three test statistics. Recall that these limits are distribution free and identical to the ones in the classical case, i.e. when  $m$  and  $\sigma$  are not estimated, but known.

**Theorem 2.3.** *Assume (A), (K), (J) and (F). Then, under  $H_0$ ,*

$$\begin{aligned}
 T_{n,KS} &\xrightarrow{d} \sup_{0 < u, v < 1} |W_0(u, v)|, \\
 T_{n,CM} &\xrightarrow{d} \iint W_0^2(u, v) \, du \, dv, \\
 T_{n,AD} &\xrightarrow{d} \iint \frac{W_0^2(u, v)}{uv(1-u)(1-v)} \, du \, dv.
 \end{aligned}$$

**Proof.** The result for  $T_{n,KS}$  follows readily from Theorem 2.2 and the continuous mapping theorem. The result for  $T_{n,CM}$  follows from Theorem 2.2, Lemma A.1, (2.9) and the Helly–Bray theorem.

Now we present the proof for  $T_{n,AD}$ . From the Skorohod construction and Theorem 2.2 it follows that (keeping the same notation for the new processes)

$$\sup_{x,y} \left| \frac{\sqrt{n}\hat{T}(x,y) - W_0(F_X(x), F_\varepsilon(y))}{N(x,y)^\beta} \right| \xrightarrow{P} 0, \tag{2.10}$$

where  $\hat{T}(x,y) = \hat{F}_{X,\varepsilon}(x,y) - \hat{F}_X(x)\hat{F}_\varepsilon(y)$  and  $0 \leq \beta < \frac{1}{2}$ . In what follows we will show that

$$\iint \frac{n\hat{T}^2(x,y)}{\hat{N}(x,y)} d\hat{F}_X(x) d\hat{F}_\varepsilon(y) - \iint \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)} dF_X(x) dF_\varepsilon(y) \xrightarrow{P} 0, \tag{2.11}$$

where  $\hat{N}(x,y) = \hat{F}_X(x)\hat{F}_\varepsilon(y)(1 - \hat{F}_{X-}(x))(1 - \hat{F}_{\varepsilon-}(y))$ . Define  $A_n = (\hat{F}_X^{-1}(n^{-3/4}), \hat{F}_X^{-1}(1 - n^{-3/4})) \times (\hat{F}_\varepsilon^{-1}(n^{-3/4}), \hat{F}_\varepsilon^{-1}(1 - n^{-3/4}))$ . The left-hand side of (2.11) can be written as

$$\begin{aligned} & \iint_{A_n} \frac{n\hat{T}^2(x,y) - W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)^{1/3}} \frac{N(x,y)^{1/3}}{\hat{N}(x,y)^{1/4}} \frac{d\hat{F}_X(x) d\hat{F}_\varepsilon(y)}{\hat{N}(x,y)^{3/4}} \\ & + \iint_{A_n} \frac{N(x,y) - \hat{N}(x,y)}{N(x,y)^{2/5}} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)^{4/5}} \frac{N(x,y)^{1/5}}{\hat{N}(x,y)^{1/6}} \frac{d\hat{F}_X(x) d\hat{F}_\varepsilon(y)}{\hat{N}(x,y)^{5/6}} \\ & + \iint_{A_n} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)} d\hat{F}_X(x) d\hat{F}_\varepsilon(y) - \iint_{D_X \times \mathbb{R}} \frac{W_0^2(F_X(x), F_\varepsilon(y))}{N(x,y)} dF_X(x) dF_\varepsilon(y) \\ & + \iint_{A_n^c} \frac{n\hat{T}^2(x,y)}{\hat{N}(x,y)} d\hat{F}_X(x) d\hat{F}_\varepsilon(y) \\ & = \sum_{i=1}^5 T_i. \end{aligned}$$

The term  $T_1$  is  $o_P(1)$  by (2.10) and Lemma A.2. For showing that  $T_2 = o_P(1)$  use is made of Lemmas A.1 and A.2 and the Chibisov–O’Reilly theorem. The convergence in probability to 0 of  $T_3 + T_4$  follows from the Helly–Bray theorem. Remains to consider  $T_5$ . We will only show that

$$\iint_{B_n} \frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_\varepsilon(y)} d\hat{F}_X(x) d\hat{F}_\varepsilon(y) \xrightarrow{P} 0,$$

where  $B_n$  is the intersection of  $A_n^c$  and  $(-\infty, m_1) \times (-\infty, m_\varepsilon)$ , with  $m_1$  and  $m_\varepsilon$  the medians of  $F_X$  and  $F_\varepsilon$ , respectively; the other parts can be dealt with similarly. First consider (with  $c_n = \hat{F}_X^{-1}(n^{-3/4})$ ,  $d_n = \hat{F}_\varepsilon^{-1}(n^{-3/4})$ )

$$\begin{aligned} & \int_{-\infty}^{c_n} \int_{-\infty}^{d_n} \frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_\varepsilon(y)} d\hat{F}_\varepsilon(y) d\hat{F}_X(x) \\ & \leq 2n \int_{-\infty}^{c_n} \int_{-\infty}^{d_n} \frac{\hat{F}_{X,\varepsilon}^2(x,y) + \hat{F}_X^2(x)\hat{F}_\varepsilon^2(y)}{\hat{F}_X(x)\hat{F}_\varepsilon(y)} d\hat{F}_\varepsilon(y) d\hat{F}_X(x) \\ & \leq 4n\hat{F}_X(c_n)\hat{F}_\varepsilon(d_n) \\ & \leq 4n(n^{-3/4} + n^{-1})^2 \rightarrow 0. \end{aligned}$$

Next, consider

$$\begin{aligned} & \int_{c_n}^{m_1} \int_{-\infty}^{d_n} \frac{n\hat{T}^2(x,y)}{\hat{F}_X(x)\hat{F}_\varepsilon(y)} d\hat{F}_\varepsilon(y) d\hat{F}_X(x) \\ & \leq 2n \int_{c_n}^{m_1} \int_{-\infty}^{d_n} \left( \frac{\hat{F}_\varepsilon(y)}{\hat{F}_X(x)} + \hat{F}_X(x)\hat{F}_\varepsilon(y) \right) d\hat{F}_\varepsilon(y) d\hat{F}_X(x) \\ & \leq n(\log n^{3/4} + 1)(n^{-3/4} + n^{-1})^2 \rightarrow 0. \end{aligned}$$

Finally, the integral over  $(-\infty, c_n) \times (d_n, m_\varepsilon)$  can be dealt with in a similar way.  $\square$

### 3. Simulations

The test statistics considered in the previous section are asymptotically distribution free, and hence the asymptotic critical values of the tests can be obtained by simulation or from tables. However, for finite sample sizes simulations show that these asymptotic critical values do not respect well the size of the test (for  $\alpha = 0.05$  and  $n$  at most 200, the size is about 0.005 or even smaller). Hence, this approximation is not of much use in practice when the sample size is not very large. This is maybe not so surprising, given that the estimators of  $m$  and  $\sigma$  converge at a slow ‘nonparametric’ rate, and hence, even though the effect of estimating these functions disappears asymptotically, the estimators do affect the rejection probabilities substantially for finite  $n$ .

Therefore, a bootstrap procedure is a useful alternative and can be performed in the following way. Fix  $B$  and let  $b = 1, \dots, B$ .

*Step 1:* Let  $\varepsilon_1^{*b}, \dots, \varepsilon_n^{*b}$  be an i.i.d. sample from the distribution of the residuals  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$ .

*Step 2:* Define  $Y_i^{*b} = \hat{m}(X_i) + \hat{\sigma}(X_i)\varepsilon_i^{*b}$  ( $i = 1, \dots, n$ ).

*Step 3:* Let  $T_{n,KS}^{*b}$ ,  $T_{n,CM}^{*b}$  and  $T_{n,AD}^{*b}$  be the test statistics obtained from the bootstrap sample  $\{(X_i, Y_i^{*b}), i = 1, \dots, n\}$ .

If we denote  $T_{n,KS}^{*(b)}$  for the order statistics of the values  $T_{n,KS}^{*1}, \dots, T_{n,KS}^{*B}$  obtained in Step 3, and analogously for  $T_{n,CM}^{*(b)}$  and  $T_{n,AD}^{*(b)}$ , then  $T_{n,KS}^{*[(1-\alpha)B]}$ ,  $T_{n,CM}^{*[(1-\alpha)B]}$  and  $T_{n,AD}^{*[(1-\alpha)B]}$  approximate the  $(1 - \alpha)$ th quantiles of the distributions of  $T_{n,KS}$ ,  $T_{n,CM}$  and  $T_{n,AD}$ , respectively.

We carry out two different simulation studies. In the first study, we compare the rejection probabilities of the proposed tests with those of the tests studied in Einmahl and Van Keilegom (2007). Since in the latter paper it is assumed that  $\sigma \equiv 1$ , we replace  $\hat{\sigma}$  everywhere by 1 in our test statistics. In Einmahl and Van Keilegom (2007) the same type of test statistics is used as in the present paper, but the bivariate empirical distribution function on which these statistics are based is very different. Instead of estimating the location curve  $m$ , in that paper the smooth, unknown  $m$  is almost eliminated by taking appropriate differences of  $Y$ -values that correspond to three neighboring  $X$ -values. The thus obtained limiting distributions of the test statistics are not distribution free and more complicated than the ones in this paper.

Consider the following simulation set up. Suppose that  $X$  has a uniform-(0, 1) distribution,  $m(x) = E(Y|X = x) = x - 0.5x^2$ ,  $\sigma^2 = \text{Var}(Y|X = x) = 0.1^2$  and under the null hypothesis  $\varepsilon$  follows a standard normal distribution. The simulations are carried out for samples of sizes  $n = 100$  and 200 and the significance level  $\alpha = 0.05$ . The results are based on 250 samples and for each of them 250 bootstrap replications are created (except under the null hypothesis, where we use 500 samples and 500 bootstrap replications). The bandwidth  $a_n$ , for estimating  $m$ , is selected by means of a least-squares cross-validation procedure; for computing the bootstrap test statistics the same bandwidth has been used. The kernel  $K$  is equal to the Epanechnikov kernel  $K(u) = 3/(4\sqrt{5})(1 - u^2/5)I(u^2 < 5)$ .

The following alternative hypotheses are studied. First consider

$$H_{1,A} : \varepsilon | X = x \sim N(0, 1 + ax),$$

with  $a > 0$ . Next, let

$$H_{1,B} : \varepsilon | X = x \stackrel{d}{=} \frac{W_x - r_x}{\sqrt{2r_x}},$$

where  $W_x \sim \chi_{r_x}^2$ ,  $r_x = 1/(bx)$  and  $b > 0$  controls the skewness of the distribution. Note that the first and second moment of the variable  $\varepsilon$  created in the latter way do not depend on  $x$  and coincide with the respective moments under  $H_0$ . When  $b$  tends to 0, the distribution of  $\varepsilon | X = x$  converges to its null distribution. Finally, let

$$H_{1,C} : \varepsilon | X = x \sim \sqrt{1 - (cx)^{1/4}} t_{2/(cx)^{1/4}},$$

where  $0 < c \leq 1$  is a parameter controlling the kurtosis of the distribution. By construction, the conditional moments up to order three of  $\varepsilon$  given  $X$  are constant and coincide with the respective moments under the null hypothesis, while the fourth conditional moment does depend on  $X$  (note that the third and fourth moment do

not need to exist). The conditional distribution of  $\varepsilon$  under  $H_{1,C}$  converges to the conditional null distribution of  $\varepsilon$  when  $c$  tends to 0.

Tables 1–3 summarize the results for these three alternative hypotheses. Table 1 shows that under the alternative hypothesis  $H_{1,A}$ , the new method clearly outperforms the difference approach of Einmahl and Van Keilegom (2007), except for the Kolmogorov–Smirnov test. Under the alternative  $H_{1,B}$  (Table 2), the new approach performs better than the difference approach for small  $b$ ; for larger  $b$  the difference approach is somewhat better. Finally, the results under the alternative  $H_{1,C}$ , given in Table 3, show that the difference approach gives higher power than the present approach in most cases, but for the Anderson–Darling statistic (which is the best one for detecting this alternative) it is the other way around. In summary, we see good behavior of the present method and we can observe that both methods perform quite different and therefore both have their merits for detecting certain alternatives.

Next, in the second simulation study we consider the general heteroscedastic model, in which the function  $\sigma$  is supposed to be unknown. The same simulation setup is chosen as for the first study, except that we take now  $n = 50$  and  $100$ ,  $\sigma^2(x) = (2 + x)^2/100$ , and the results are based on 500 samples and 500 bootstrap replications.

Table 1  
Power under  $H_{1,A}$  with known variance

| $a$ | Meth. | $n = 100$ |           |           | $n = 200$ |           |           |
|-----|-------|-----------|-----------|-----------|-----------|-----------|-----------|
|     |       | <i>KS</i> | <i>CM</i> | <i>AD</i> | <i>KS</i> | <i>CM</i> | <i>AD</i> |
| 0   | Est   | 0.068     | 0.072     | 0.072     | 0.070     | 0.060     | 0.066     |
|     | Diff  | 0.044     | 0.072     | 0.034     | 0.062     | 0.050     | 0.040     |
| 1   | Est   | 0.080     | 0.096     | 0.132     | 0.136     | 0.208     | 0.376     |
|     | Diff  | 0.088     | 0.124     | 0.092     | 0.148     | 0.184     | 0.168     |
| 2.5 | Est   | 0.152     | 0.268     | 0.316     | 0.312     | 0.624     | 0.788     |
|     | Diff  | 0.176     | 0.236     | 0.216     | 0.304     | 0.412     | 0.432     |
| 5   | Est   | 0.224     | 0.444     | 0.524     | 0.540     | 0.872     | 0.960     |
|     | Diff  | 0.240     | 0.352     | 0.308     | 0.492     | 0.672     | 0.716     |
| 10  | Est   | 0.328     | 0.568     | 0.668     | 0.708     | 0.964     | 1.00      |
|     | Diff  | 0.344     | 0.488     | 0.428     | 0.656     | 0.856     | 0.872     |

The new method is indicated by ‘Est’, the difference approach by ‘Diff’.

Table 2  
Power under  $H_{1,B}$  with known variance

| $b$ | Meth. | $n = 100$ |           |           | $n = 200$ |           |           |
|-----|-------|-----------|-----------|-----------|-----------|-----------|-----------|
|     |       | <i>KS</i> | <i>CM</i> | <i>AD</i> | <i>KS</i> | <i>CM</i> | <i>AD</i> |
| 0   | Est   | 0.068     | 0.072     | 0.072     | 0.070     | 0.060     | 0.066     |
|     | Diff  | 0.044     | 0.072     | 0.034     | 0.062     | 0.050     | 0.040     |
| 1   | Est   | 0.212     | 0.160     | 0.224     | 0.324     | 0.256     | 0.396     |
|     | Diff  | 0.060     | 0.100     | 0.068     | 0.092     | 0.176     | 0.116     |
| 2.5 | Est   | 0.392     | 0.236     | 0.344     | 0.568     | 0.328     | 0.468     |
|     | Diff  | 0.120     | 0.216     | 0.116     | 0.224     | 0.348     | 0.232     |
| 5   | Est   | 0.524     | 0.324     | 0.388     | 0.600     | 0.408     | 0.468     |
|     | Diff  | 0.148     | 0.300     | 0.200     | 0.460     | 0.672     | 0.572     |
| 10  | Est   | 0.616     | 0.396     | 0.412     | 0.728     | 0.484     | 0.496     |
|     | Diff  | 0.256     | 0.512     | 0.380     | 0.712     | 0.880     | 0.816     |

The new method is indicated by ‘Est’, the difference approach by ‘Diff’.

We consider only  $H_{1,B}$  and  $H_{1,C}$ , since  $H_{1,A}$  is now contained in the null hypothesis. The bandwidths used to estimate  $m$  and  $\sigma$  are different. They are both selected by means of a cross-validation procedure; again, these bandwidths are also used for computing the bootstrap test statistics. No competing procedures exist for testing this general model.

The results are given in Tables 4 and 5 and show that the significance level is quite close to the nominal value of 0.05, both for  $n = 50$  and 100. For the alternative hypothesis  $H_{1,B}$ , the Kolmogorov–Smirnov test usually

Table 3  
Power under  $H_{1,C}$  with known variance

| $c$ | Meth. | $n = 100$ |           |           | $n = 200$ |           |           |
|-----|-------|-----------|-----------|-----------|-----------|-----------|-----------|
|     |       | <i>KS</i> | <i>CM</i> | <i>AD</i> | <i>KS</i> | <i>CM</i> | <i>AD</i> |
| 0   | Est   | 0.068     | 0.072     | 0.072     | 0.070     | 0.060     | 0.066     |
|     | Diff  | 0.044     | 0.072     | 0.034     | 0.062     | 0.050     | 0.040     |
| 0.2 | Est   | 0.080     | 0.096     | 0.116     | 0.084     | 0.100     | 0.124     |
|     | Diff  | 0.044     | 0.080     | 0.040     | 0.088     | 0.108     | 0.096     |
| 0.4 | Est   | 0.108     | 0.100     | 0.160     | 0.112     | 0.128     | 0.216     |
|     | Diff  | 0.052     | 0.088     | 0.048     | 0.132     | 0.180     | 0.156     |
| 0.6 | Est   | 0.132     | 0.156     | 0.224     | 0.148     | 0.248     | 0.344     |
|     | Diff  | 0.072     | 0.160     | 0.100     | 0.220     | 0.312     | 0.252     |
| 0.8 | Est   | 0.192     | 0.240     | 0.360     | 0.236     | 0.388     | 0.584     |
|     | Diff  | 0.172     | 0.280     | 0.200     | 0.444     | 0.604     | 0.532     |
| 1   | Est   | 0.308     | 0.432     | 0.572     | 0.512     | 0.752     | 0.876     |
|     | Diff  | 0.376     | 0.612     | 0.520     | 0.836     | 0.944     | 0.940     |

The new method is indicated by ‘Est’, the difference approach by ‘Diff’.

Table 4  
Power under  $H_{1,B}$  with unknown variance

| $b$ | $n = 50$  |           |           | $n = 100$ |           |           |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|
|     | <i>KS</i> | <i>CM</i> | <i>AD</i> | <i>KS</i> | <i>CM</i> | <i>AD</i> |
| 0   | 0.046     | 0.030     | 0.048     | 0.068     | 0.050     | 0.044     |
| 1   | 0.098     | 0.074     | 0.116     | 0.156     | 0.136     | 0.298     |
| 2.5 | 0.314     | 0.204     | 0.270     | 0.430     | 0.302     | 0.422     |
| 5   | 0.530     | 0.348     | 0.410     | 0.592     | 0.354     | 0.474     |
| 10  | 0.556     | 0.342     | 0.360     | 0.594     | 0.356     | 0.416     |

Table 5  
Power under  $H_{1,C}$  with unknown variance

| $c$ | $n = 50$  |           |           | $n = 100$ |           |           |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|
|     | <i>KS</i> | <i>CM</i> | <i>AD</i> | <i>KS</i> | <i>CM</i> | <i>AD</i> |
| 0   | 0.046     | 0.030     | 0.048     | 0.068     | 0.050     | 0.044     |
| 0.2 | 0.098     | 0.088     | 0.086     | 0.078     | 0.082     | 0.096     |
| 0.4 | 0.098     | 0.096     | 0.094     | 0.080     | 0.096     | 0.100     |
| 0.6 | 0.108     | 0.116     | 0.100     | 0.098     | 0.114     | 0.130     |
| 0.8 | 0.146     | 0.142     | 0.144     | 0.140     | 0.154     | 0.214     |
| 1   | 0.258     | 0.242     | 0.262     | 0.272     | 0.366     | 0.478     |



outperforms the two other tests, whereas for the alternative  $H_{1,C}$  there is not so much difference between the behavior of the three test statistics for  $n = 50$ , whereas the Anderson–Darling test comes out as winner for  $n = 100$ . Note that, under both alternatives, the power of the three test statistics increases with  $b$  and  $c$ , except when  $b$  increases from 5 to 10. This seems to be due to the fact that the conditional error distribution is very skewed.

**4. Data analysis**

We consider monthly expenditures in Dutch Guilders ( $\approx 0.45$  Euro) of Dutch households on several commodity categories and a number of background variables. These data can be found in the Data Archive of the Journal of Applied Econometrics, see [Adang and Melenberg \(1995\)](#). We use accumulated expenditures on food and total expenditures over the year October 1986 through September 1987 for households consisting of two persons ( $n = 159$ ) and want to regress two responses, namely

$$Y_1 = \text{share of food expenditure in household budget,}$$

$$Y_2 = \log(\text{expenditure on food per household})$$

to the regressor  $X = \log(\text{total expenditures})$ . Scatterplots of these responses versus the regressor are given in [Fig. 1](#). We want to use our tests to see if model (1.1) is appropriate. The bandwidths for estimating  $m$  and  $\sigma$  are, as in the simulation section, determined by means of a cross-validation procedure. The  $P$ -values of the tests are presented in [Table 6](#).

The table shows that model (1.1) is violated by  $Y_1$  (except for the Kolmogorov–Smirnov test, whose  $P$ -value is borderline), but not by  $Y_2$ . Next, we like to test whether  $Y_2$  satisfies the more restrictive homoscedastic model  $Y_2 = m(X) + \varepsilon$ , with  $\varepsilon$  independent of  $X$ . The  $P$ -values given in the last column of [Table 6](#) indicate that the homoscedastic model is valid too and can be used for an analysis of the log food expenditure data. This is in agreement with the findings in [Einmahl and Van Keilegom \(2007\)](#).

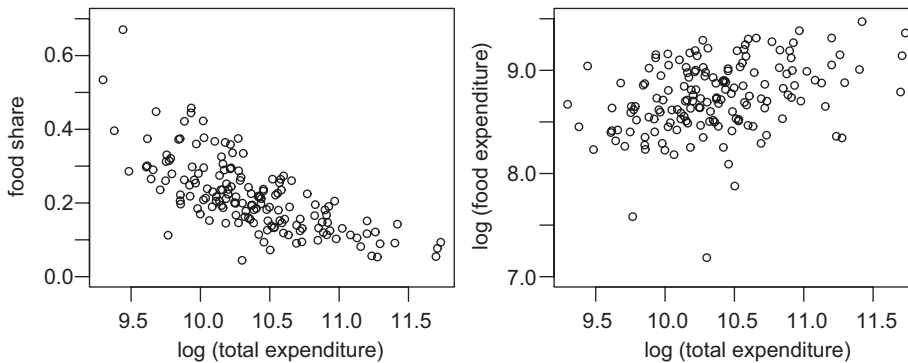


Fig. 1. Scatterplot of  $Y_1$  versus  $X$  (left) and of  $Y_2$  versus  $X$  (right).

Table 6  
 $P$ -values for the household data for the heteroscedastic and homoscedastic model

| Test      | $Y_1$  |  | $Y_2$  |       |
|-----------|--------|--|--------|-------|
|           | Hetero |  | Hetero | Homo  |
| <i>KS</i> | 0.071  |  | 0.369  | 0.378 |
| <i>CM</i> | 0.011  |  | 0.315  | 0.371 |
| <i>AD</i> | 0.011  |  | 0.477  | 0.638 |

**Acknowledgment**

The second author acknowledges financial support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy) and from CentER, Tilburg University (as extramural fellow).

**Appendix A**

The asymptotic results given in Section 2 require the following assumptions.

(A) The sequence  $(a_n)_{n \in \mathbb{N}}$  satisfies  $na_n^4 \rightarrow 0$  and  $na_n^{3+2\delta}(\log a_n^{-1})^{-1} \rightarrow \infty$  for some  $\delta > 0$ .

(K) The probability density function  $K$  has compact support,  $\int uK(u)du = 0$  and  $K$  is twice continuously differentiable.

(J) (i)  $J(s) = I(0 \leq s \leq 1)$  or (ii) there exist  $0 \leq s_0 \leq s_1 \leq 1$  such that  $s_0 \leq \inf\{s \in [0, 1] : J(s) \neq 0\}$ ,  $s_1 \geq \sup\{s \in [0, 1] : J(s) \neq 0\}$  and  $\inf_{x \in D_X} \inf_{s_0 \leq s \leq s_1} f(F^{-1}(s|x)|x) > 0$  and  $J$  is twice continuously differentiable on the interior of its support,  $\int_0^1 J(s) ds = 1$  and  $J(s) \geq 0$  for all  $0 \leq s \leq 1$ .

(F) (i) The support  $D_X$  of  $X$  is a bounded interval,  $F_X$  is twice continuously differentiable and  $\inf_{x \in D_X} f_X(x) > 0$ .

(ii)  $F(y|x)$  is differentiable in  $y$  and twice differentiable in  $x$  and the derivatives are continuous in  $(x, y)$ . Moreover,  $\sup_{x,y} |y|f(y|x) < \infty$  and  $\sup_{x,y} |y \frac{\partial^k}{\partial x^k} F(y|x)| < \infty$  for  $k = 1, 2$ .

(iii) For every  $\gamma \in (0, 1)$  there exists an  $\alpha \in (0, 1)$ , such that

$$\sup_{y, |z_1 - 1| \leq \alpha, |z_2| \leq \alpha} \frac{\max(|y|, 1)f_\varepsilon(z_1y + z_2)}{\min(F_\varepsilon(y), 1 - F_\varepsilon(y))^\gamma} < \infty,$$

and

$$\sup_{y, |z_1 - 1| \leq \alpha, |z_2| \leq \alpha} \frac{f_\varepsilon(z_1y + z_2)}{f_\varepsilon(y)} \min(F_\varepsilon(y), 1 - F_\varepsilon(y))^\gamma < \infty.$$

(iv)  $\inf_{x \in D_X} \sigma(x) > 0$ .

Note that condition (F) (iii) is only needed for the Anderson–Darling statistic and controls the denominator of that statistic. This condition is satisfied for error distributions encountered in practice, in particular for the normal distribution (used as null distribution in the simulation section) and for the Student  $t$ -distribution.

In addition to  $F_\varepsilon, \hat{F}_\varepsilon$  and  $\hat{F}_{\hat{\varepsilon}}$ , we will need  $F_{\hat{\varepsilon}}(y) = P(\{Y - \hat{m}(X)\}/\hat{\sigma}(X) \leq y | \hat{m}, \hat{\sigma})$ , where  $(X, Y)$  is independent of  $(X_1, Y_1), \dots, (X_n, Y_n)$ . The proofs of Section 2 are based on the two following crucial results.

**Lemma A.1.** Assume (A), (K), (J) and (F). Then, for  $0 \leq \beta < \frac{1}{2}$ ,

$$\sup_{y \in \mathbb{R}} |[F_\varepsilon(y)(1 - F_\varepsilon(y))]^{-\beta} [\hat{F}_{\hat{\varepsilon}}(y) - \hat{F}_\varepsilon(y) - F_{\hat{\varepsilon}}(y) + F_\varepsilon(y)]| = o_P(n^{-1/2})$$

and

$$\sup_{\substack{x \in D_X \\ y \in \mathbb{R}}} \left| N(x, y)^{-\beta} n^{-1} \sum_{i=1}^n [I(X_i \leq x) - F_X(x)][I(\hat{\varepsilon}_i \leq y) - I(\varepsilon_i \leq y) - F_{\hat{\varepsilon}}(y) + F_\varepsilon(y)] \right| = o_P(n^{-1/2}).$$

**Proof.** We will show the first statement. The second one can be proved in a similar way. For reasons of symmetry we restrict attention to the case where  $y < F_\varepsilon^{-1}(1/2)$ . Since  $1 - F_\varepsilon(y)$  is bounded away from 0 in this case, we only need to consider  $F_\varepsilon(y)$  in the denominator. In order to simplify the presentation, we will present the proof for the case  $\sigma \equiv 1$  and known. If this is not the case, the estimator  $\hat{\sigma}$  can be handled in much the same way as the estimator  $\hat{m}$ .

Choose  $0 < \delta_1 < (\frac{1}{2} - \beta) / (\frac{1}{2} + \beta)$ . Write

$$\begin{aligned} \frac{F_\varepsilon(y)}{F_\varepsilon(y)^{1-\delta_1}} - F_\varepsilon(y)^{\delta_1} &= \int \frac{P(\varepsilon \leq y + \hat{m}(x) - m(x) | \hat{m}) - P(\varepsilon \leq y)}{F_\varepsilon(y)^{1-\delta_1}} dF_X(x) \\ &= \int \frac{f_\varepsilon(\xi_y(x))}{F_\varepsilon(y)^{1-\delta_1}} (\hat{m}(x) - m(x)) dF_X(x), \end{aligned}$$

for some  $\xi_y(x)$  between  $y$  and  $y + \hat{m}(x) - m(x)$ . Since  $\sup_{y, |z| \leq \alpha} f_\varepsilon(y+z)/F_\varepsilon(y)^{1-\delta_1} < \infty$  (for some  $\alpha > 0$ ) and  $\sup_x |\hat{m}(x) - m(x)| = o_P(1)$  (see Proposition 4.3 in Akritas and Van Keilegom, 2001), the above is  $o_P(1)$ , uniformly in  $y$ . Hence, it follows that

$$\sup_y \frac{F_{\hat{\varepsilon}}(y)^{\beta+\delta_2}}{F_\varepsilon(y)^\beta} = O_P(1)$$

with  $\delta_2 = \beta\delta_1/(1 - \delta_1)$  and so it suffices to consider  $F_{\hat{\varepsilon}}(y)^{-\beta-\delta_2} n^{-1} \sum_{i=1}^n [I(\hat{\varepsilon}_i \leq y) - I(\varepsilon_i \leq y) - F_{\hat{\varepsilon}}(y) + F_\varepsilon(y)]$ . Next, note that in a similar way (but with replacing  $\delta_1$  by  $\delta_1/(1 + \delta_1)$ ), we can show that

$$\sup_y \left| \frac{F_\varepsilon(y)^{\beta+2\delta_2}}{F_{\hat{\varepsilon}}(y)^{\beta+\delta_2}} - F_\varepsilon(y)^{\delta_2} \right| \xrightarrow{P} 0.$$

So since  $\beta + 2\delta_2 < \frac{1}{2}$ , it follows from the Chibisov–O’Reilly theorem that

$$\sup_y \left| F_\varepsilon(y)^{-\beta-2\delta_2} n^{-1} \sum_{i=1}^n [I(\varepsilon_i \leq y) - F_\varepsilon(y)] \right| = O_P(n^{-1/2}),$$

and hence it suffices to show that

$$\sup_y \left| n^{-1} \sum_{i=1}^n \left[ \frac{I(\hat{\varepsilon}_i \leq y)}{F_{\hat{\varepsilon}}(y)^a} - \frac{I(\varepsilon_i \leq y)}{F_\varepsilon(y)^a} - F_{\hat{\varepsilon}}(y)^{1-a} + F_\varepsilon(y)^{1-a} \right] \right| = o_P(n^{-1/2}), \tag{A.1}$$

where  $a = \beta + \delta_2$  throughout the proof. Note that  $0 \leq a < 1/2$ . Let  $d_n(x) = \hat{m}(x) - m(x)$ , and consider the class

$$\mathcal{F} = \left\{ (x, e) \mapsto \frac{I(e \leq y + d(x))}{P(e \leq y + d(X))^a} - \frac{I(e \leq y)}{P(e \leq y)^a} - P(e \leq y + d(X))^{1-a} + P(e \leq y)^{1-a}; \right. \\ \left. y < F_\varepsilon^{-1}(1/2), d \in C_1^{1+\delta}(D_X) \right\},$$

where  $C_1^{1+\delta}(D_X)$  (with  $\delta > 0$  as in Assumption (A)) is the class of all differentiable functions  $d$  defined on  $D_X$  such that  $\|d\|_{1+\delta} \leq \alpha/2$  (with  $\alpha > 0$  as in Assumption (F) (iii)), where

$$\|d\|_{1+\delta} = \max \left\{ \sup_x |d(x)|, \sup_x |d'(x)| \right\} + \sup_{x, x'} \frac{|d'(x) - d'(x')|}{|x - x'|^\delta}.$$

Note that by Propositions 4.3–4.5 in Akritas and Van Keilegom (2001), we have that  $P(d_n \in C_1^{1+\delta}(D_X)) \rightarrow 1$  as  $n \rightarrow \infty$ . In the next part of this proof we will show that the class  $\mathcal{F}$  is Donsker, i.e. we will establish the weak convergence of  $n^{-1/2} \sum_{i=1}^n f(X_i, \varepsilon_i)$ ,  $f \in \mathcal{F}$ . This is done by verifying the conditions of Theorem 2.11.9 in van der Vaart and Wellner (1996):

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\bar{\varepsilon}, \mathcal{F}, L_2^n)} d\bar{\varepsilon} \rightarrow 0 \quad \text{for every } \delta_n \downarrow 0, \tag{A.2}$$

$$n^{1/2} E \left\{ \sup_{f \in \mathcal{F}} |f(X, \varepsilon)| I \left( \sup_{f \in \mathcal{F}} |f(X, \varepsilon)| > n^{1/2} \eta \right) \right\} \rightarrow 0 \quad \text{for every } \eta > 0, \tag{A.3}$$

where  $N_{[]}(\bar{\varepsilon}, \mathcal{F}, L_2^n)$  is the bracketing number, defined as the minimal number of sets  $N_{\bar{\varepsilon}}$  in a partition  $\mathcal{F} = \bigcup_{j=1}^{N_{\bar{\varepsilon}}} \mathcal{F}_{\bar{\varepsilon}j}$ , such that for every  $j = 1, \dots, N_{\bar{\varepsilon}}$ :

$$E \left\{ \sup_{f, g \in \mathcal{F}_{\bar{\varepsilon}j}} |f(X, \varepsilon) - g(X, \varepsilon)|^2 \right\} \leq \bar{\varepsilon}^2. \tag{A.4}$$

According to Theorem 2.10.6 in van der Vaart and Wellner (1996), we can deal with the four terms in the definition of  $\mathcal{F}$  separately. We will restrict ourselves to showing (A.2) and (A.3) for

$$\mathcal{F}_1 = \left\{ (x, e) \mapsto \frac{I(e \leq y + d(x))}{P(e \leq y + d(X))^a}; y < F_\varepsilon^{-1}(1/2), d \in C_1^{1+\delta}(D_X) \right\},$$

since the other terms are similar, but much easier. We will assume  $0 < \bar{\varepsilon} \leq 1$ . In Corollary 2.7.2 of the aforementioned book it is stated that  $m = N_{\lfloor (K_1 \bar{\varepsilon})^2, C_1^{1+\delta}(D_X), L_2(P) \rfloor}$  is bounded by  $\exp(K(K_1 \bar{\varepsilon})^{-2/(1+\delta)})$ , with  $K_1 > 0$  to be determined later. Let  $d_1^L \leq d_1^U, \dots, d_m^L \leq d_m^U$  be the functions defining the  $m$  brackets for  $C_1^{1+\delta}(D_X)$ . Thus, for each  $d$  and each fixed  $y$ :

$$I(\varepsilon \leq y + d_i^L(X)) \leq I(\varepsilon \leq y + d(X)) \leq I(\varepsilon \leq y + d_i^U(X)).$$

Let  $b = \min(2a, 1 - 2a)$ . Define  $F_i^L(y) = P(\varepsilon \leq y + d_i^L(X))$  and let  $-\infty = y_{i1}^L < y_{i2}^L < \dots < y_{i,m_L}^L = +\infty$  ( $m_L = O(\bar{\varepsilon}^{-2/b})$ ) partition the line in segments having  $F_i^L$ -probability less than or equal to  $K_2 \bar{\varepsilon}^{2/b}$  where  $K_2 > 0$  will be chosen later. Similarly, define  $F_i^U(y) = P(\varepsilon \leq y + d_i^U(X))$  and let  $-\infty = y_{i1}^U < y_{i2}^U < \dots < y_{i,m_U}^U = +\infty$  ( $m_U = O(\bar{\varepsilon}^{-2/b})$ ) partition the line in segments having  $F_i^U$ -probability less than or equal to  $K_2 \bar{\varepsilon}^{2/b}$ .

Let  $\mathcal{F}_{\bar{\varepsilon}ik}$  ( $i = 1, \dots, m, k = 1, \dots, m_L - 1$ ) be the subset of  $\mathcal{F}_1$  defined by the functions  $d_i^L \leq d \leq d_i^U$  and  $\tilde{y}_{ik}^L \leq y \leq \tilde{y}_{ik}^U$ , where  $\tilde{y}_{ik}^L = y_{ik}^L$  and  $\tilde{y}_{ik}^U$  is the smallest of the  $y_{ik}^U$  which is larger than (or equal to)  $y_{i,k+1}^L$ . Fix  $i, k$  and fix  $X$  and  $\varepsilon$ . We consider three cases:

Case 1: For all  $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X, \varepsilon) = 0$ . The supremum in (A.4) equals zero in that case.

Case 2: For certain  $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X, \varepsilon) = 0$  and for certain  $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X, \varepsilon) \neq 0$ . This happens only if  $\tilde{y}_{ik}^L + d_i^L(X) \leq \varepsilon \leq \tilde{y}_{ik}^U + d_i^U(X)$ . Also, the supremum in (A.4) is bounded by  $F_\varepsilon(\varepsilon)^{-2a}$  in that case. Hence, the expected value in (A.4), restricted to those  $(X, \varepsilon)$  that belong to case 2, is bounded by

$$\begin{aligned} & \iint_{\tilde{y}_{ik}^L + d_i^L(x)}^{\tilde{y}_{ik}^U + d_i^U(x)} F_\varepsilon(y)^{-2a} dF_\varepsilon(y) dF_X(x) \\ &= \frac{1}{1 - 2a} \int \left[ F_\varepsilon(\tilde{y}_{ik}^U + d_i^U(x))^{1-2a} - F_\varepsilon(\tilde{y}_{ik}^L + d_i^L(x))^{1-2a} \right] dF_X(x) \\ &= \frac{1}{1 - 2a} \int \left[ F_\varepsilon(\tilde{y}_{ik}^U + d_i^U(x))^{1-2a} - F_\varepsilon(\tilde{y}_{i,k+1}^L + d_i^U(x))^{1-2a} \right] dF_X(x) \\ & \quad + \frac{1}{1 - 2a} \int \left[ F_\varepsilon(\tilde{y}_{i,k+1}^L + d_i^U(x))^{1-2a} - F_\varepsilon(\tilde{y}_{i,k+1}^L + d_i^L(x))^{1-2a} \right] dF_X(x) \\ & \quad + \frac{1}{1 - 2a} \int \left[ F_\varepsilon(\tilde{y}_{i,k+1}^L + d_i^L(x))^{1-2a} - F_\varepsilon(\tilde{y}_{ik}^L + d_i^L(x))^{1-2a} \right] dF_X(x) \\ & \leq \frac{1}{1 - 2a} \left[ F_i^U(\tilde{y}_{ik}^U) - F_i^U(\tilde{y}_{i,k+1}^L) \right]^{1-2a} + \int F_\varepsilon(\xi_{ik}(x))^{-2a} f_\varepsilon(\xi_{ik}(x)) (d_i^U(x) - d_i^L(x)) dF_X(x) \\ & \quad + \frac{1}{1 - 2a} \left[ F_i^L(\tilde{y}_{i,k+1}^L) - F_i^L(\tilde{y}_{ik}^L) \right]^{1-2a} \\ & \leq \frac{2K_2^{1-2a}}{1 - 2a} \bar{\varepsilon}^{2(1-2a)/b} + K' \|d_i^U - d_i^L\|_{L_1(P)}, \end{aligned}$$

and this is bounded by  $\bar{\varepsilon}^2$  for proper choice of  $K_1$  and  $K_2$ , where  $K' > 0$  and where  $\xi_{ik}(x)$  is between  $\tilde{y}_{i,k+1}^L$  and  $d_i^L(x)$  and  $\tilde{y}_{i,k+1}^L + d_i^U(x)$ .

Case 3: For all  $f \in \mathcal{F}_{\bar{\varepsilon}ik}, f(X, \varepsilon) \neq 0$ . This implies that  $k > 1$  and hence  $F_i^L(\tilde{y}_{ik}^L) \geq K \bar{\varepsilon}^{2/b}$ . Hence, the expected value at the left-hand side of (A.4), restricted to those  $(X, \varepsilon)$  that satisfy the condition of case 3, is bounded by

$$\begin{aligned} & [F_i^L(\tilde{y}_{ik}^L)^{-a} - F_i^U(\tilde{y}_{ik}^U)^{-a}]^2 F_i^L(\tilde{y}_{ik}^L) \\ &= \left\{ \left[ F_i^L(\tilde{y}_{ik}^L)^{-a} - F_i^L(\tilde{y}_{i,k+1}^L)^{-a} \right] + \left[ F_i^L(\tilde{y}_{i,k+1}^L)^{-a} - F_i^U(\tilde{y}_{i,k+1}^L)^{-a} \right] + \left[ F_i^U(\tilde{y}_{i,k+1}^L)^{-a} - F_i^U(\tilde{y}_{ik}^U)^{-a} \right] \right\}^2 F_i^L(\tilde{y}_{ik}^L) \\ &=: \{T_1 + T_2 + T_3\}^2 F_i^L(\tilde{y}_{ik}^L) \leq 3\{T_1^2 + T_2^2 + T_3^2\} F_i^L(\tilde{y}_{ik}^L). \end{aligned}$$

It is easy to see that

$$T_1^2 F_i^L(\tilde{y}_{ik}^L) \leq \left[ F_i^L(\tilde{y}_{i,k+1}^L) - F_i^L(\tilde{y}_{ik}^L) \right]^{2a} F_i^L(\tilde{y}_{ik}^L)^{1-4a}$$

and this is bounded by  $\bar{\varepsilon}^2$  for proper choice of  $K_2 > 0$  (consider separately  $a \leq 1/4$  and  $a > 1/4$ ). It can be shown in a similar way that  $T_\ell^2 F_i^L(\tilde{y}_{ik}^L) \leq \bar{\varepsilon}^2$  for  $\ell = 2, 3$  and for  $K_1, K_2 > 0$  small enough. This shows that (A.4) is

satisfied and hence

$$N_{[\cdot]}(\bar{\varepsilon}, \mathcal{F}_1, L_2^n) = O\left(\exp\left(2K(K_1\bar{\varepsilon})^{-2/(1+\delta)}\right)\right).$$

It now follows that (A.2) holds, since

$$\int_0^{\delta n} \sqrt{\log N_{[\cdot]}(\bar{\varepsilon}, \mathcal{F}_1, L_2^n)} d\bar{\varepsilon} \leq 2K \int_0^{\delta n} (K_1\bar{\varepsilon})^{-1/(1+\delta)} d\bar{\varepsilon} = 2K \frac{1+\delta}{\delta} (K_1\delta n)^{\delta/(1+\delta)} \rightarrow 0.$$

Next, by writing

$$\begin{aligned} \sup_{f \in \mathcal{F}_1} |f(X, \varepsilon)| &\leq \sup_{d \in C_1^{1+\delta}(D_X), -\infty < y < \infty} \frac{I(\varepsilon \leq y + d(X))}{\mathbb{P}(\varepsilon \leq y + d(X))^a} \\ &= \sup_{d \in C_1^{1+\delta}(D_X)} \left[ \int F_\varepsilon(\varepsilon - d(X) + d(x))^a dF_X(x) \right]^{-1} \\ &\leq F_\varepsilon(\varepsilon - \alpha)^{-a}, \end{aligned}$$

it follows that the left-hand side of (A.3) is bounded by

$$n^{1/2} \mathbb{E}\{F_\varepsilon(\varepsilon - \alpha)^{-a} I(F_\varepsilon(\varepsilon - \alpha)^{-a} > n^{1/2}\eta)\} = n^{1/2} \int_{-\infty}^{\kappa n} F_\varepsilon(y - \alpha)^{-a} f_\varepsilon(y) dy, \tag{A.5}$$

where  $\kappa n = F_\varepsilon^{-1}[(n^{1/2}\eta)^{-1/a}] + \alpha$ . It now follows from condition (F) (iii) that (A.5) is bounded, for  $n$  large enough, by (where  $\gamma > 0$  is chosen such that  $a + \gamma < 1/2$  and  $K$  is some positive constant)

$$\begin{aligned} &Kn^{1/2} \int_{-\infty}^{\kappa n} F_\varepsilon(y - \alpha)^{-(a+\gamma)} dF_\varepsilon(y - \alpha) \\ &= Kn^{1/2} \int_0^{(n^{1/2}\eta)^{-1/a}} u^{-(a+\gamma)} du \\ &= \frac{Kn^{1/2}}{1 - a - \gamma} \left(n^{1/2}\eta\right)^{-(1-a-\gamma)/a} = O\left(n^{(2a+\gamma-1)/(2a)}\right) = o(1). \end{aligned}$$

This shows that the class  $\mathcal{F}_1$  (and hence  $\mathcal{F}$ ) is Donsker.

Next, let us calculate

$$\begin{aligned} &\text{Var} \left[ \frac{I(\varepsilon \leq y + d_n(X))}{\mathbb{P}(\varepsilon \leq y + d_n(X))^a} - \frac{I(\varepsilon \leq y)}{\mathbb{P}(\varepsilon \leq y)^a} - \mathbb{P}(\varepsilon \leq y + d_n(X))^{1-a} + \mathbb{P}(\varepsilon \leq y)^{1-a} \middle| d_n \right] \\ &\leq \mathbb{E} \left[ \mathbb{E} \left\{ \left( \frac{I(\varepsilon \leq y + d_n(X))}{\mathbb{P}(\varepsilon \leq y + d_n(X))^a} - \frac{I(\varepsilon \leq y)}{\mathbb{P}(\varepsilon \leq y)^a} \right)^2 \middle| X, d_n \right\} \middle| d_n \right]. \end{aligned} \tag{A.6}$$

The conditional expectation is equal to (suppose that  $d_n(X) \geq 0$  for simplicity)

$$\begin{aligned} &\mathbb{E} \left[ \frac{I(\varepsilon \leq y + d_n(X))}{F_\varepsilon(y + d_n(X))^{2a}} + \frac{I(\varepsilon \leq y)}{F_\varepsilon(y)^{2a}} - 2 \frac{I(\varepsilon \leq y)}{F_\varepsilon(y + d_n(X))^a F_\varepsilon(y)^a} \middle| X, d_n \right] \\ &= F_\varepsilon(y + d_n(X))^{1-2a} - F_\varepsilon(y)^{1-2a} + 2 \frac{F_\varepsilon(y)^{1-2a}}{F_\varepsilon(y + d_n(X))^a} [F_\varepsilon(y + d_n(X))^a - F_\varepsilon(y)^a] \\ &= (1 - 2a)F_\varepsilon(\xi_y(X))^{-2a} f_\varepsilon(\xi_y(X)) d_n(X) + 2a \frac{F_\varepsilon(y)^{1-2a}}{F_\varepsilon(y + d_n(X))^a} F_\varepsilon(\tilde{\xi}_y(X))^{a-1} f_\varepsilon(\tilde{\xi}_y(X)) d_n(X) \\ &\leq F_\varepsilon(\xi_y(X))^{-2a} f_\varepsilon(\xi_y(X)) d_n(X) \end{aligned}$$

and, by condition (F) (iii), this is bounded by  $Kd_n(X)$  for some  $K > 0$ , where  $\xi_y(X)$  and  $\tilde{\xi}_y(X)$  are between  $y$  and  $y + d_n(X)$ . A similar derivation can be given when  $d_n(X) \leq 0$ . It follows that the right-hand side of (A.6) is bounded by  $K \sup_x |d_n(x)| = o_p(1)$ , by Proposition 4.3 in Akritas and Van Keilegom (2001).

Since the class  $\mathcal{F}$  is Donsker, it follows from Corollary 2.3.12 in van der Vaart and Wellner (1996) that

$$\lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{f \in \mathcal{F}, \text{Var}(f) < \eta} n^{-1/2} \left| \sum_{i=1}^n f(X_i, \varepsilon_i) \right| > \bar{\varepsilon} \right) = 0$$

for each  $\bar{\varepsilon} > 0$ . By restricting the supremum inside this probability to the elements in  $\mathcal{F}$  corresponding to  $d(X) = d_n(X)$  as defined above, (A.1) follows.  $\square$

**Lemma A.2.** Assume (A), (K), (J) and (F), let  $\eta > 0, 0 < \zeta < 1$  and  $b_n = n^{-(1-\zeta)}$ . Then,

$$\sup_{\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{ F_{\hat{\varepsilon}}(y)^\eta \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \right\} = o_{\mathbb{P}}(1)$$

and

$$\sup_{1 - \hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{ (1 - F_{\hat{\varepsilon}}(y))^\eta \left| \frac{1 - F_{\hat{\varepsilon}}(y)}{1 - \hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \right\} = o_{\mathbb{P}}(1).$$

**Proof.** We only prove the first statement. The second one follows in a similar way. Choose  $\nu > 0$  such that  $\nu < 1/2$  and  $(1 - \zeta)(1 - \nu) < 1/2$ . Then,

$$\begin{aligned} \sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \left| \frac{\hat{F}_{\hat{\varepsilon}}(y)}{F_{\hat{\varepsilon}}(y)} - 1 \right| &\leq \sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \frac{|\hat{F}_{\hat{\varepsilon}}(y) - F_{\hat{\varepsilon}}(y)|}{F_{\hat{\varepsilon}}(y)^\nu b_n^{1-\nu}} \\ &= o(1) \sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \frac{|\sqrt{n}(\hat{F}_{\hat{\varepsilon}}(y) - F_{\hat{\varepsilon}}(y))|}{F_{\hat{\varepsilon}}(y)^\nu} \\ &= o(1) \left( \sup \frac{|\sqrt{n}(\hat{F}_{\hat{\varepsilon}}(y) - F_{\hat{\varepsilon}}(y))|}{F_{\hat{\varepsilon}}(y)^\nu} + o_{\mathbb{P}}(1) \right) = o_{\mathbb{P}}(1), \end{aligned}$$

where the last equality follows from the Chibisov–O’Reilly theorem and the one but last equality from the proof of Lemma A.1. Hence, it follows that

$$\sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| = o_{\mathbb{P}}(1). \tag{A.7}$$

We next show that the supremum in (A.7) can be replaced by the supremum over  $\{y : \hat{F}_{\hat{\varepsilon}}(y) \geq b_n\}$ . Indeed, it follows from (A.7) that there exists a sequence  $\delta_n \downarrow 0$  such that

$$\mathbb{P} \left( \sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \geq \delta_n \right) \rightarrow 0$$

as  $n \rightarrow \infty$ . Hence,

$$\mathbb{P} \left( \sup_{\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \geq \varepsilon \right) \leq \mathbb{P} \left( \sup_{F_{\hat{\varepsilon}}(y) \geq b_n(1-\delta_n)} \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \geq \varepsilon \right) + \mathbb{P} \left( \sup_{F_{\hat{\varepsilon}}(y) \geq b_n} \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \geq \delta_n \right) \rightarrow 0.$$

Finally, consider

$$\sup_{\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{ F_{\hat{\varepsilon}}(y)^\eta \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \right\} = \sup_{\hat{F}_{\hat{\varepsilon}}(y) \geq b_n} \left\{ F_{\hat{\varepsilon}}(y)^\eta \left| \frac{F_{\hat{\varepsilon}}(y)}{\hat{F}_{\hat{\varepsilon}}(y)} - 1 \right| \right\} + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1),$$

since  $\sup_y |F_{\hat{\varepsilon}}(y)^\eta - F_{\hat{\varepsilon}}(y)^\eta| = o_{\mathbb{P}}(1)$  and since it follows from the proof of Lemma A.1 that

$$\sup_y \left| \frac{F_{\hat{\varepsilon}}(y)^{1+\eta}}{\hat{F}_{\hat{\varepsilon}}(y)} - F_{\hat{\varepsilon}}(y)^\eta \right| = o_{\mathbb{P}}(1). \quad \square$$

## References

- Adang, P.J.M., Melenberg, B., 1995. Nonnegativity constraints and intratemporal uncertainty in multi-good life-cycle models. *Journal of Applied Econometrics* 10, 1–15.
- Akritas, M.G., Van Keilegom, I., 2001. Nonparametric estimation of the residual distribution. *Scandinavian Journal of Statistics* 28, 549–568.
- Blum, J.R., Kiefer, J., Rosenblatt, M., 1961. Distribution free tests of independence based on the sample distribution function. *Annals of Mathematical Statistics* 32, 485–498.
- Cheng, F., 2004. Weak and strong uniform consistency of a kernel error density estimator in nonparametric regression. *Journal of Statistical Planning and Inference* 119, 95–107.
- Dette, H., Neumeyer, N., Van Keilegom, I., 2007. A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society Series B*, to appear.
- Einmahl, J.H.J., Van Keilegom, I., 2007. Tests for independence in nonparametric regression. *Statistica Sinica*, to appear.
- Hoeffding, W., 1948. A non-parametric test of independence. *Annals of Mathematical Statistics* 19, 546–547.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, New York.
- Müller, U.U., Schick, A., Wefelmeyer, W., 2004a. Estimating functionals of the error distribution in parametric and nonparametric regression. *Journal of Nonparametric Statistics* 16, 525–548.
- Müller, U.U., Schick, A., Wefelmeyer, W., 2004b. Estimating linear functionals of the error distribution in nonparametric regression. *Journal of Statistical Planning and Inference* 119, 75–93.
- Neumeyer, N., Dette, H., Nagel, E.-R., 2006. Bootstrap tests for the error distribution in linear and nonparametric regression models. *Australian & New Zealand Journal of Statistics* 48, 129–156.
- Pardo-Fernández, J.C., Van Keilegom, I., González-Manteiga, W., 2007. Testing for the equality of  $k$  regression curves. *Statistica Sinica* 17, 1115–1137.
- Shorack, G.R., Wellner, J.A., 1986. *Empirical Processes with Applications to Statistics*. Wiley, New York.
- Stone, C.J., 1977. Consistent nonparametric regression. *Annals of Statistics* 5, 595–645.
- van der Vaart, A.W., Wellner, J.A., 1996. *Weak Convergence and Empirical Processes*. Springer, New York.
- Van Keilegom, I., 1998. Nonparametric estimation of the conditional distribution in regression with censored data. Ph.D. Thesis, Hasselt University, Hasselt, Belgium.
- Van Keilegom, I., Veraverbeke, N., 2002. Density and hazard estimation in censored regression models. *Bernoulli* 8, 607–625.
- Van Keilegom, I., González Manteiga, W., Sánchez Selloero, C., 2007. Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, to appear.
- de Wet, T., 1980. Cramér–von Mises tests for independence. *Journal of Multivariate Analysis* 10, 38–50.