A robust test for homoscedasticity in nonparametric regression

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: holger.dette@rub.de

FAX: +49 234 3214 559

Mareen Marchlewski

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: mareen.marchlewski@rub.de

9 July 2009

Abstract

We consider a nonparametric location scale model and propose a new test for homoscedasticity (constant scale function). The test is based on an estimate of a deterministic function which vanishes if and only if the hypothesis of a constant scale function is satisfied and an empirical process estimating this function is investigated. Weak convergence to a scaled Brownian bridge is established, which allows a simple calculation of critical values. The new test can detect alternatives converging to the null hypothesis at a rate $n^{-1/2}$ and is robust with respect to the presence of outliers. The finite sample properties are investigated by means of a simulation study, and the test is compared with some non-robust tests for a constant scale function, which have recently been proposed in the literature.

Keywords and Phrases: robust test for heteroscedasticity, nonparametric regression, scale check, Brownian Bridge

AMS Subject Classification: 62G10, 62F35

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote a bivariate sample of independent identically distributed observations with conditional distribution function $F(y|x) = P(Y_1 \leq y|X_1 = x)$ where F is of the

form

$$(1.1) \qquad F(y\,|\,x) = F_0 \left(\frac{y - m(x)}{\sigma(x)} \right)$$

with a fixed (but unknown) distribution function F_0 with corresponding density f_0. The function m is called location or regression function, while σ is denoted as scale function. Often $m(x)$ and $\sigma^2(x)$ are viewed as expectation and variance of the conditional distribution of Y_1 given $X_1 = x$. In this context the importance of being able to detect heteroscedasticity has been widely recognized because under the additional assumption of a constant scale function the statistical analysis can be simplified substantially. Early work on this problem considering parametric specifications for the regression and scale function can be found in Harrison and McCabe (1979), Breusch and Pagan (1979), Cook and Weisberg (1983) and Diblasi and Bowman (1997) among others. The problem of testing for heteroscedasticity in the classical nonparametric regression model with conditional expectation m and conditional variance σ^2 has been considered in Dette and Munk (1998), Dette (2002), Liero (2003), Dette and Hetzler (2009) Francisco-Fernández and Vilar-Fernández (2005), Dette et al. (2007) or Dette and Hetzler (2009). Similar testing problems in semiparametric models have been considered by You and Chen (2005). However, all these references assume the existence of the second moment of the conditional distribution. For this reason these methods may not be robust with respect to outliers in the data. Koenker and Bassett (1981) proposed a robust test for heteroscedasticity in a quantile regression model with a parametrically specified location and scale function, but the problem of testing this assumption in a nonparametric location scale model has – to the knowledge of the authors – not been considered in the literature so far.

It is the purpose of the present paper to develop a robust test for a constant scale function in the location-scale model (1.1), that is

$$(1.2) \qquad H_0 : \sigma(x) = \sigma$$

for some constant $\sigma > 0$. We will develop a test which can detect alternatives of scale functions converging to the null hypothesis with the rate $n^{-1/2}$ and is additionally robust with respect to outliers in the data. In Section 2 we introduce the necessary notation and review some results about robust nonparametric estimation of the regression function m. We also introduce a functional which allows a characterization of the hypothesis (1.2). Section 3 contains our main results. We propose a stochastic process which is the basis for our robust test of homoscedasticity. Weak convergence of this process is proved under the null hypothesis, fixed and local alternatives. Moreover, consistency is established and the finite sample properties of the new test are investigated by means of a simulation study in Section 4. Here we also compare the performance of the new test with some tests which have been recently proposed in the literature (assuming existence of a second moment of the conditional distribution) in the presence of outliers.

2 Notation and a preliminary result

We consider an equivalent formulation of model (1.1), i.e.

$$(2.1) \qquad Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \qquad i = 1, \dots, n,$$

where $(X_1, Y_1), \dots, (X_n, Y_n)$ are i.i.d, the errors $\varepsilon_1, \dots, \varepsilon_n$ are independent and have density f_0 and distribution function F_0, respectively, where f_0 is symmetric (but not necessarily has existing moments). We define the density of the conditional distribution $F(Y|X)$ by $f(y|x)$ and denote by g the marginal density of X, which is assumed to have compact support, say $\text{supp}(g) = [0, 1]$ (note that this implies that g is positive on $[0, 1]$). The corresponding distribution function is denoted by G. In order to construct a robust estimate of the regression function m we consider a root of the equation

$$(2.2) \qquad H_n(x, \vartheta) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \psi(Y_i - \vartheta)$$

with respect to ϑ, where K denotes a kernel function with compact support, say $\text{supp}(K) = [-1, 1]$, satisfying

$$(2.3) \qquad \int K(u)du = 1 \ , \quad \int uK(u)du = 0.$$

h is a bandwidth converging to 0 with increasing sample size and ψ is a strictly monotone, antisymmetric and bounded function having (a.e.) two continuous bounded derivatives with $\psi'(0) > 0$. Note that the solution of (2.2) defines a robust nonparametric regression estimate as considered by Härdle (1984), Härdle and Gasser (1984) or Stützle and Mittal (1979) among others. Let $\hat{m}_n(x)$ denote a root of the equation (2.2) where we assume that the kernel K has been appropriately modified in order to address for boundary effects. We define the stochastic process

$$(2.4) \qquad \hat{T}_t = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq t\} \psi^2\left(Y_i - \hat{m}_n(X_i)\right) - \hat{G}_n(t) \cdot \frac{1}{n} \sum_{i=1}^{n} \psi^2\left(Y_i - \hat{m}_n(X_i)\right),$$

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

$$(2.5) \qquad \hat{G}_n(t) = \frac{1}{n} \sum_{j=1}^{n} I\{X_j \leq t\}$$

denotes the empirical distribution function of the explanatory variables. A heuristic argument (which will be made rigorous in the following section) shows that for any $t \in [0, 1]$

$$(2.6) \qquad \hat{T}_t \xrightarrow{P} T_t,$$

where the deterministic process $(T_t)_{t\in[0,1]}$ is defined by

$$(2.7) \quad T_t = \int_0^t g(x) \int_{\mathbb{R}} \psi^2(\sigma(x)\omega) f_0(\omega)d\omega dx - G(t) \int_0^1 g(x) \int_{\mathbb{R}} \psi^2(\sigma(x)\omega) f_0(\omega)d\omega dx.$$

The following Lemma shows that $(T_t)_{t\in[0,1]}$ vanishes if and only if the hypothesis of homoscedasticity is satisfied.

Lemma 2.1. *If the function ψ is strictly increasing and antisymmetric, then $T_t = 0$ a.e. on $[0,1]$ if and only if the hypothesis (1.2) of a constant scale function is satisfied.*

Proof. If $\sigma(x) = \sigma$ for all $x \in [0,1]$, then, by Fubini's theorem $T_t = 0$ a.e. on the interval $[0,1]$. To prove the converse note that a constant process $(T_t)_{t\in[0,1]}$ implies

$$0 = T'_t = g(t) \int_0^1 g(x) \int_{\mathbb{R}} \left(\psi^2(\sigma(t)\omega) - \psi^2(\sigma(x)\omega) \right) f_0(\omega)d\omega dx.$$

Because $\mathrm{supp}(g) = [0,1]$ it follows

$$(2.8) \qquad \int_0^1 g(x) \int_{\mathbb{R}} \left(\psi^2(\sigma(t)\omega) - \psi^2(\sigma(x)\omega) \right) f_0(\omega)d\omega dx = 0$$

for a.e. $t \in [0,1]$. If there would exist $t_1 \neq t_2$ such that $0 < \sigma(t_1) < \sigma(t_2)$ then

$$
\begin{aligned}
0 &= \int_0^1 g(x) \int_{\mathbb{R}} \left(\psi^2(\sigma(t_1)\omega) - \psi^2(\sigma(x)\omega) \right) f_0(\omega)d\omega dx \\
&\quad - \int_0^1 g(x) \int_{\mathbb{R}} \left(\psi^2(\sigma(t_2)\omega) - \psi^2(\sigma(x)\omega) \right) f_0(\omega)d\omega dx \\
(2.9) \qquad &= \int_{\mathbb{R}} \left(\psi^2(\sigma(t_1)\omega) - \psi^2(\sigma(t_2)\omega) \right) f_0(\omega)d\omega.
\end{aligned}
$$

Now, if $\omega > 0$ it follows that $\psi^2(\sigma(t_1)\omega) < \psi^2(\sigma(t_2)\omega)$ by strict monotonicity of ψ. Similarly, if $\omega < 0$ we have $\psi^2(\sigma(t_1)(-\omega)) < \psi^2(\sigma(t_2)(-\omega))$ and by antisymmetry $\psi^2(\sigma(t_1)\omega) < \psi^2(\sigma(t_2)\omega)$. Consequently we obtain the inequality

$$0 = \int_{\mathbb{R}} \left(\psi^2(\sigma(t_1)\omega) - \psi^2(\sigma(t_2)\omega) \right) f_0(\omega)d\omega < 0$$

which yields a contradiction. Therefore the proof of Lemma 2.1 is completed. $\qquad\square$

From Lemma 2.1 it follows that the hypothesis (1.2) can be tested using the Kolmogorov-Smirnoff or Cramér-von Mises statistic, that is

$$(2.10) \qquad\qquad \hat{K} = \sup_{t\in[0,1]} |\hat{T}_t| \;; \quad \hat{C} = \int_0^1 \hat{T}_t^2 d\hat{G}_n(t),$$

respectively. In the following section we will investigate the asymptotic properties of the process $(\hat{T}_t)_{t\in[0,1]}$ in order to derive (asymptotic) critical values for tests based on the statistic $\hat{K}$ and $\hat{C}$.

4

3   Asymptotic properties of the process $(\hat{T}_t)_{t\in[0,1]}$

Consider the process $(\hat{T}_t)_{t\in[0,1]}$ defined by (2.4). For our asymptotic results we require the following conditions on the sequence of bandwidths used in (2.2)

$$(3.1) \qquad\qquad\qquad\qquad nh^2 \;\to\; \infty$$

$$(3.2) \qquad\qquad\qquad\qquad nh^5 \;\to\; 0$$

Moreover, we assume that the conditional density

$$(3.3) \qquad f(y|x) \quad \text{has a continuous bounded second derivative w.r.t. the variable } x,$$

for the density g of the marginal distribution of the random variable X

$$(3.4) \qquad\qquad\qquad 0 < c_0 \leq g(x) \leq c_1 \qquad \forall\, x \in [0,1]$$

$$(3.5) \qquad\qquad\qquad \frac{\partial^2}{\partial^2 x} g(x) \quad \text{exists a.e. on } [0,1]$$

and for the regression function

$$(3.6) \qquad\qquad\qquad m \text{ is twice continuously differentiable.}$$

Throughout this paper $D[0,1]$ denotes the set of cadlag functions on the interval $[0,1]$ [see Billingsley (1999)]. The following result establishes weak convergence of the process $\sqrt{n}(\hat{T}_t - T_t)$. Note that it does not require that the null hypotheses (1.2) of a constant scale function is satisfied. This case is considered in Corollary 3.2 below.

Theorem 3.1. *If the assumptions stated in Section 2 and (3.1) - (3.6) are satisfied, then the stochastic process*

$$\{\sqrt{n}(\hat{T}_t - T_t)\}_{t\in[0,1]}$$

converges weakly in $D[0,1]$ to a centered Gaussian process $\{A(t)\}_{t\in[0,1]}$ with covariance kernel

$$(3.7)\, k(t_1, t_2) = \int_0^1 g(x)\gamma_4(x)\,(I\{x \leq t_1\} - G(t_1))\,(I\{x \leq t_2\} - G(t_2))\,dx$$

$$- \int_0^1 g(x)\gamma_2(x)\,(I\{x \leq t_1\} - G(t_1))\,dx \int_0^1 g(x)\gamma_2(x)\,(I\{x \leq t_2\} - G(t_2))\,dx,$$

where for $j = 1, 2$ the constant $\gamma_j(x)$ is given by

$$(3.8) \qquad\qquad \gamma_j(x) = E[\psi^j\,(\sigma(x)\varepsilon_1) \mid X_1 = x] = \int_{\mathbb{R}} \psi^j\,(\sigma(x)\omega)\, f_0(\omega)d\omega.$$

Proof. Use the decomposition

$$\hat{T}_t = \hat{T}_t^{(1)} - \hat{T}_t^{(2)}, \tag{3.9}$$

where the processes $(\hat{T}_t^{(1)})_{t \in [0,1]}$ and $(\hat{T}_t^{(2)})_{t \in [0,1]}$ are defined by

$$\hat{T}_t^{(1)} = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq t\} \psi^2 \left(Y_i - \hat{m}_n(X_i)\right),$$

$$\hat{T}_t^{(2)} = \hat{G}_n(t) \cdot \frac{1}{n} \sum_{i=1}^n \psi^2 \left(Y_i - \hat{m}_n(X_i)\right).$$

For the first term we obtain by means of a Taylor expansion

$$\hat{T}_t^{(1)} = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq t\} \psi^2 \left(\sigma(X_i)\varepsilon_i\right) + D_1 + D_2, \tag{3.10}$$

where the random variables D_1 and D_2 are given by

$$D_1 = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq t\} \left(m(X_i) - \hat{m}_n(X_i)\right) \left(\psi^2\right)' \left(\sigma(X_i)\varepsilon_i\right) \tag{3.11}$$

$$D_2 = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq t\} \left(m(X_i) - \hat{m}_n(X_i)\right)^2 \frac{\left(\psi^2\right)''(\xi_i)}{2} \tag{3.12}$$

and the random variables ξ_i satisfy $|\xi_i - \sigma(X_i)\varepsilon_i| \leq |\hat{m}_n(X_i) - m(X_i)|$. Because the second derivative of ψ is bounded we obtain by similar arguments as in Härdle (1984) or Härdle and Gasser (1984) that the bias and variance of $\hat{m}_n(x)$ are of order $O(h^2)$ and $O(\frac{1}{nh})$ (uniformly with respect to $x \in [0,1]$), respectively, where it is assumed that the estimates have been modified at the boundary to address for boundary effects. This yields

$$D_2 = O_p \left(h^4 + \frac{1}{nh}\right) = o_p \left(\frac{1}{\sqrt{n}}\right). \tag{3.13}$$

From a more extensive calculation given in the Appendix it also follows that

$$D_1 = o_p \left(\frac{1}{\sqrt{n}}\right). \tag{3.14}$$

A straightforward calculation yields for the expectation of the second term in (3.9)

$$E\left[\frac{G(t)}{n} \sum_{i=1}^n \psi^2 \left(Y_i - m(X_i)\right)\right] = G(t) \int_0^1 g(x)\gamma_2(x)dx, \tag{3.15}$$

and a similar argument as used in the derivation of (3.10) shows

$$\frac{\hat{G}_n(t)}{n} \sum_{i=1}^n \psi^2 \left(Y_i - \hat{m}_n(X_i)\right) = \frac{G(t)}{n} \sum_{i=1}^n \psi^2 \left(\sigma(X_i)\varepsilon_i\right) + o_p \left(\frac{1}{\sqrt{n}}\right). \tag{3.16}$$

6

Summarizing (3.10), (3.13), (3.14) and (3.16) now yields

(3.17)
$$\hat{T}_t = \hat{T}_t^{(1)} - \hat{T}_t^{(2)} = \tilde{T}_t + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where the process $\tilde{T}_t$ is defined by

(3.18)
$$\tilde{T}_t = \frac{1}{n}\sum_{i=1}^{n} I\{X_i \le t\}\psi^2(\sigma(X_i)\varepsilon_i) - \frac{G(t)}{n}\sum_{i=1}^{n} \psi^2(\sigma(X_i)\varepsilon_i).$$

Obviously we have $E[\tilde{T}_t] = T_t$ and consequently the assertion of Theorem 3.1 follows if the weak convergence of the process

(3.19)
$$A_n(t) := \sqrt{n}\left(\tilde{T}_t - E\left[\tilde{T}_t\right]\right) = \sqrt{n}\left(\hat{T}_t - T_t\right) + o_p(1),$$

can be established. For this purpose we define the random variables

$$Z_{i,t} = L_{i,t} - E\left[L_{i,t}\right]$$
$$L_{i,t} = \psi^2\left(\sigma(X_i)\varepsilon_i\right)\left(I\{X_i \le t\} - G(t)\right),$$

and obtain the representation

(3.20)
$$A_n(t) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n} Z_{i,t}.$$

Note that the random variables $Z_{i,t}$ are i.i.d. with $E[Z_{i,t}] = 0$. In order to determine the covariance $\mathrm{Cov}(A_n(t_1), A_n(t_2))$ we note

(3.21)
$$E\left[L_{i,t}\right] = \int_0^1 g(x)\gamma_2(x)\left(I\{x \le t\} - G(t)\right)dx,$$
$$E\left[L_{i,t}L_{i,s}\right] = \int_0^1 g(x)\gamma_4(x)\left(I\{x \le t\} - G(t)\right)\left(I\{x \le s\} - G(s)\right)dx,$$

which yields by a straightforward calculation

(3.22)
$$\mathrm{Cov}\left(A_n(t_1), A_n(t_2)\right) = k(t_1, t_2),$$

where the kernel k is defined by (3.7). In order to prove asymptotic normality we use the Cramér-Wold device and the central limit theorem (note that the random variables $Z_{i,t}$ are bounded) and obtain

(3.23)
$$(A_n(t_1), \ldots, A_n(t_k))^T \xrightarrow{D} (A(t_1), \ldots, A(t_k))^T,$$

where $(A(t))_{t \in [0,1]}$ denotes a Gaussian process with covariance kernel k defined in (3.7).

7

According to Theorem 13.5 in Billingsley (1999) the assertion of the theorem now follows if the condition

$$(3.24) \qquad E\left[(A_n(t) - A_n(s))^2 (A_n(r) - A_n(t))^2\right] \leq C(r - s) \quad \forall \; 0 \leq s < t < r \leq 1$$

for some constant $C > 0$ can be established, which implies tightness of the process (3.20). In order to prove (3.24) we calculate

$$h_n(s, t, r) = E\left[(A_n(t) - A_n(s))^2 (A_n(r) - A_n(t))^2\right].$$

A straightforward calculation yields

$$
\begin{aligned}
h_n(s, t, r) \;\leq\; & E\left[(L_{i,s,t} - E[L_{i,s,t}])^2 (L_{i,t,r} - E[L_{i,t,r}])^2\right] + E\left[(L_{i,s,t} - E[L_{i,s,t}])^2\right] E\left[(L_{i,t,r} - E[L_{i,t,r}])^2\right] \\
& + 2E^2\left[(L_{i,s,t} - E[L_{i,s,t}])(L_{i,t,r} - E[L_{i,t,r}])\right]
\end{aligned}
$$

$$
\begin{aligned}
\;=\; & E\left[L_{1,s,t}^2 L_{1,t,r}^2\right] - 2E\left[L_{1,s,t}^2 L_{1,t,r}\right] E[L_{1,t,r}] - 2E\left[L_{1,s,t} L_{1,t,r}^2\right] E[L_{1,s,t}] \\
& + E\left[L_{1,s,t}^2\right] E\left[L_{1,t,r}^2\right] + 2E^2\left[L_{1,s,t} L_{1,t,r}\right]
\end{aligned}
$$

where the random variables $L_{i,j,l}$ are defined by

$$L_{i,j,l} = \psi^2(\sigma(X_i)\varepsilon_i)\left(I\{j < X_i \leq l\} - (G(l) - G(j))\right).$$

Using this definition and the notation (3.8) it follows

$$
\begin{aligned}
h_n(s, t, r) \;\leq\; & (G(r) - G(t))^2 \int_s^t g(x)\gamma_8(x)dx + (G(t) - G(s))^2 \int_t^r g(x)\gamma_8(x)dx \\
& + (G(t) - G(s))^2 (G(r) - G(t))^2 \int g(x)\gamma_8(x)dx \\
& + 4(G(r) - G(t))\left(\int_s^t g(x)\gamma_6(x)dx + (G(t) - G(s))^2 \int g(x)\gamma_6(x)dx\right)\left(\int_t^r g(x)\gamma_2(x)dx\right) \\
& + 4\left(2(G(t) - G(s))(G(r) - G(t))\int_s^t g(x)\gamma_6(x)dx + (G(t) - G(s))^2 \int_t^r g(x)\gamma_6(x)dx\right) \\
& \times \left((G(r) - G(t))\int g(x)\gamma_2(x)dx\right) \\
& + \left(\int_s^t g(x)\gamma_4(x)dx + (G(t) - G(s))^2 \int g(x)\gamma_4(x)dx\right) \\
& \times \left(\int_t^r g(x)\gamma_4(x)dx + (G(r) - G(t))^2 \int g(x)\gamma_4(x)dx\right) \\
& + 4\left((G(t) - G(s))\int_s^t g(x)\gamma_4(x)dx\right)\left((G(r) - G(t))\int_t^r g(x)\gamma_4(x)dx\right) \\
& + 2\left((G(t) - G(s))(G(r) - G(t))\int g(x)\gamma_4(x)dx\right. \\
& \left. - (G(r) - G(t))\int_s^t g(x)\gamma_4(x)dx - (G(t) - G(s))\int_t^r g(x)\gamma_4(x)dx\right)^2.
\end{aligned}
$$

which yields (3.24) by the mean value theorem. Consequently, tightness and convergence of the finite dimensional distributions imply the assertion of Theorem 3.1. □

Recall that by Lemma 2.1 the null hypothesis is satisfied if and only if $T_t = 0$ a.e. on $[0, 1]$. In this case Theorem 3.1 simplifies substantially, and the process $\sqrt{n}\{\hat{T}_t\}_{t \in [0,1]}$ converges to a scaled Brownian Bridge.

Corollary 3.2. *If the assumption of Theorem 3.1 and the null hypothesis (1.2) are satisfied, then*

$$(3.25) \qquad \{\sqrt{n}\hat{T}_t\}_{t \in [0,1]} \Rightarrow \sqrt{m_4}\{B \circ G\}_{t \in [0,1]},$$

on $D[0, 1]$ where B denotes a Brownian bridge and $m_4 = \int_{\mathbb{R}} \psi^4 (\sigma \omega) f_0(\omega)d\omega$.

Proof. If $\sigma(x) = \sigma$ it follows by Fubini's theorem that

$$
\begin{aligned}
k(t_1, t_2) &= \int_0^1 g(x) \int_{\mathbb{R}} \psi^4 (\sigma \omega) f_0(\omega)d\omega \, (I\{x \le t_i\} - G(t_i))\,(I\{x \le t_j\} - G(t_j))\, dx \\
&\quad - \left(\int_0^1 g(x) \int_{\mathbb{R}} \psi^2 (\sigma \omega) f_0(\omega)d\omega \, (I\{x \le t_i\} - G(t_i))\, dx \right) \\
&\quad \times \left(\int_0^1 g(x) \int_{\mathbb{R}} \psi^2 (\sigma \omega) f_0(\omega)d\omega \, (I\{x \le t_j\} - G(t_j))\, dx \right) \\
&= \int_{\mathbb{R}} \psi^4 (\sigma \omega) f_0(\omega)d\omega \, (G(t_i \wedge t_j) - G(t_i)G(t_j)),
\end{aligned}
$$

which proves the assertion of Corollary 3.2. □

Similar calculations as used in the proof of Theorem 3.1 show that a consistent estimate of m_4 is given by

$$(3.26) \qquad \hat{m}_4 = \frac{1}{n} \sum_{i=1}^n \psi^4(Y_i - \hat{m}_n(X_i)),$$

and consequently it follows from the continuous mapping and Slutzky's theorem that under the null hypothesis (1.2)

$$\frac{\sqrt{n}\hat{K}}{\sqrt{\hat{m}_4}} \xrightarrow{\mathcal{D}} \sup_{t \in [0,1]} |B(t)| \; ; \quad \frac{\sqrt{n}\hat{C}}{\hat{m}_4} \xrightarrow{\mathcal{D}} \int_0^1 B^2(t)dt.$$

If $k_{1-\alpha}$ and $c_{1-\alpha}$ denote the corresponding $(1 - \alpha)$ quantiles of the distributions on the right hand side, respectively, it follows that rejecting the null hypothesis of a constant scale function if

$$(3.27) \qquad \sqrt{n}\hat{K} > k_{1-\alpha}\sqrt{\hat{m}_4}$$

9

or if

(3.28)
$$\sqrt{n}\hat{C} > c_{1-\alpha}\hat{m}_4$$

yields an asymptotic level α test. The consistency of this test follows from Theorem 3.1 and Lemma 2.1 which show that under the alternative $\sqrt{n}\hat{K} \xrightarrow{P} \infty$; $\sqrt{n}\hat{C} \xrightarrow{P} \infty$.

Remark 3.3. From similar arguments as given in the proof of Theorem 3.1 it follows that the test for the hypothesis of a constant scale function based on the process $\{\sqrt{n}\hat{T}_t\}_{t\in[0,1]}$, can detect local alternatives of the form

(3.29)
$$\sigma_n(x) = \sigma + n^{-1/2}h(x),$$

where $h : [0,1] \to \mathbb{R}$ denotes a fixed function, such that the variance function $\sigma_n(x)$ is nonnegative for all $x \in [0,1]$. To be precise, recall that the stochastic approximation in (3.17) is also valid under local alternatives. Therefore a Taylor expansion of $\psi^2((\sigma + n^{-1/2}h(x))\varepsilon_i)$ yields

$$
\begin{aligned}
\sqrt{n}\hat{T}_t &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I\{X_i \le t\} - G(t))\,\psi^2(\sigma\varepsilon_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}(I\{X_i \le t\} - G(t))\,h(X_i)\varepsilon_i\,(\psi^2)'(\sigma\varepsilon_i) + o_p(1).
\end{aligned}
$$

If the first moment of $h(X_i)\varepsilon_i\,(\psi^2)'(\sigma\varepsilon_i)$ exists the last term converges almost surely. Consequently, the process $\{\sqrt{n}\hat{T}_t\}_{t\in[0,1]}$ converges weakly to

$$
\left\{\sqrt{m_4}(B \circ G) + \int_0^1 (I\{x \le t\} - G(t))\,h(x)g(x)dx\int_{\mathbb{R}}\omega\,(\psi^2)'(\sigma\omega)f_0(\omega)d\omega\right\}_{t\in[0,1]}.
$$

In the remaining part of this section we will briefly discuss a corresponding result under the assumption of a fixed design. More precisely, we consider the model

(3.30)
$$Y_{i,n} = m(x_{i,n}) + \sigma(x_{i,n})\varepsilon_{i,n}$$

where $\{\varepsilon_{i,n}, \dots, \varepsilon_{n,n}\}$ denotes a triangular array of independent random variables with symmetric density f_0 and the design points $x_{i,n}, \dots, x_{n,n}$ satisfy the condition

(3.31)
$$\frac{i - 0.5}{n} = \int_0^{x_{i,n}} g(t)dt,$$

where g denotes a positive density on the interval $[0,1]$ [see Sacks and Ylvisaker (1966) or Sacks and Ylvisaker (1968)]. The process $\hat{T}_t$ is defined as in (2.4), where the random variables X_j are

10

replaced by the deterministic quantities $x_{j,n}$ $(j = 1, \ldots, n)$. The following results show that weak convergence also holds in the case of a fixed design, but with a different limiting process.

Theorem 3.4. *Consider the nonparametric regression model (3.30). If the assumptions stated in Section 2 and (3.1) - (3.6) are satisfied, then the process*

$$\{\sqrt{n}\left(\hat{T}_t - T_t\right)\}_{t\in[0,1]}$$

(with the obvious modifications for the fixed design assumption) converges weakly in $D[0,1]$ to a centered Gaussian process $\{\bar{A}(t)\}_{t\in[0,1]}$ with covariance kernel

$$
\begin{aligned}
\bar{k}(t_1, t_2) \;=\; & \int_0^1 g(x)\gamma_4(x)\left(I\{x \le t_1\} - G(t_1)\right)\left(I\{x \le t_2\} - G(t_2)\right) dx \\
& - \int_0^1 g(x)\gamma_2^2(x)\left(I\{x \le t_1\} - G(t_1)\right)\left(I\{x \le t_2\} - G(t_2)\right) dx
\end{aligned}
$$

(3.32)

Corollary 3.5. *If the assumptions of Theorem 3.4 and the null hypothesis (1.2) are satisfied, then*

(3.33)
$$\{\sqrt{n}\hat{T}_t\}_{t\in[0,1]} \Rightarrow \sqrt{m_4 - m_2^2}\{B \circ G\}_{t\in[0,1]},$$

on $D[0,1]$ where B denotes a Brownian bridge and $m_j = \int_{\mathbb{R}} \psi^j\left(\sigma\omega\right) f_0(\omega)d\omega$ for $j = 2, 4$.

A comparison of Corollary 3.2 and 3.5 shows that in the case of a fixed design the limiting process under the null hypothesis of a constant scale function has a smaller variance. The larger variance in Corollary 3.2 is caused by the additional randomness of the explanatory variables.

4 Finite sample properties

In this section we will study the finite sample properties of the Cramér-von Mises test (3.28) and will also compare the new test with two alternative methods, which are similar in spirit with the present method.

Following Dette and Munk (1998) we have considered the three models

$$
\begin{array}{lll}
\text{(I)} & m(x) = 1 + \sin(x) & \sigma(x) = 0.5 \ \exp{(cx)}, \\
\text{(II)} & m(x) = 1 + x & \sigma(x) = 0.5 \ (1 + c \ \sin(x))^2, \\
\text{(III)} & m(x) = 1 + x & \sigma(x) = 0.5 \ (1 + cx)^2,
\end{array}
$$

where the case $c = 0$ corresponds to a constant scale function, the choices $c = 0.5$ and 1 to two alternatives. We consider a fixed design of the form (3.31) with $g(x) = 1$. The results for the

random design are similar and not depicted for the sake of brevity. The null hypothesis of a constant scale function is rejected if

$$(4.1) \qquad \sqrt{n}\, \frac{\hat{C}}{\hat{m}_4 - \hat{m}_2^2} > c_{1-\alpha} \;,$$

where $\hat{C}$ denotes the Cramer-von Mises statistic defined in (2.10), the estimate $\hat{m}_4$ is defined in (3.26) (with an obvious modification for the fixed design), $c_{1-\alpha}$ denotes the $(1 - \alpha)$ quantile of the random variable $\int_0^1 B^2(t)dt$ and

$$\hat{m}_2 = \frac{1}{n} \sum_{i=1}^{n} \psi^2(Y_i - \hat{m}(x_i)) \;.$$

For the errors in model (2.1) we used a standard normal distribution with a $0\%, 10\%$ and 20% contamination by a t-distribution with 4 degrees of freedom. The robust test proposed in this paper requires the specification of the function ψ and we used the function

$$(4.2) \qquad \psi(x) = \max\{-\kappa, \min\{u, \kappa\}\}$$

with $\kappa = 1$. Further, a bandwidth is needed which we calculated according to a robust cross-validation method proposed by Leung (2005). For each scenario 1000 simulation runs have been performed to calculate the rejection probabilities. In Table 1 and 2 we show the simulated level of the new test (4.1) for sample sizes $n = 50$ and 100. We observe a rather precise estimation of the nominal level. The table also contains the corresponding results for the test of Dette and Munk (1998) and Dette and Hetzler (2009). While the level of the L^2-test of Dette and Munk (1998) is slightly affected by the presence of outliers, the test of Dette and Hetzler (2009) is more sensitive in such situations, and the nominal level is underestimated between $10\% - 50\%$ or $25\% - 90\%$ if the contamination rate is 10% or 20%, respectively. On the other hand the test proposed in this paper yields a good approximation of the nominal level in all cases under consideration.

In Table 3 and 4 we present the rejection probabilities of the test (4.1) in model (I), (II) and (III) for the alternatives $c = 0.5$ and $c = 1$. We observe reasonable rejection probabilities but a loss in power of up to 40% in the presence of 10% outliers and up to 60% for a contamination rate of 20%. For the sake of comparison we show in Table 5 and 6 the corresponding results of the test of Dette and Hetzler (2009) and Dette and Munk (1998) in the case $c = 0.5$. These results are directly comparable with Table 3. We observe that the test of Dette and Munk (1998) yields a small loss of power in the presence of outliers [see Table 6], while the power of the test of Dette and Hetzler (2009) is strongly affected by outliers [see Table 5]. Here a loss of power of more than 50% can be observed in many cases. Moreover, some of the clear advantages of the test of Dette and Hetzler (2009) against the test of Dette and Munk (1998) cannot be observed in the case of contaminated errors. On the other hand the new test (4.1) yields larger rejection

12

probabilities in nearly all cases under investigation. Only in model II for sample size $n = 50$ and $n = 100$ the test of Dette and Munk (1998) yields the best results. On the other hand for the sample size $n = 200$ the asymptotic advantages of the new test (4.1) are again observable [note that the test of Dette and Munk (1998) can only detect alternatives converging to the null hypothesis with rate $n^{-1/4}$ while the test of Dette and Hetzler (2009) and the test proposed in this paper can detect alternatives converging to the null with rate $n^{-1/2}$]. The results for the alternatives with $c = 1$ show similar advantages of the test (4.1) and are not displayed for the sake of brevity.

5 Appendix: Proof of the estimate (3.14)

Note that it follows from a Taylor expansion of the function $H_n(x, \hat{m}_n(x))$ defined in (2.2) at the point $m(x)$

$$\hat{m}_n(x) - m(x) = \frac{H_n(x)}{D_n(x)},$$

where

$$H_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \psi(Y_i - m(x))$$

and

$$D_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \psi'(Y_i - m(x) + \omega_i(\hat{m}_n(x) - m(x)))$$

with $\omega_i \in (0, 1)$ [see Härdle (1984)]. We define

$$c(x) = E\left[\frac{1}{h} K\left(\frac{x - X_i}{h}\right) \psi'(Y_i - m(x))\right],$$

then

$$
\begin{aligned}
D_1 &= \frac{1}{n} \sum_{k=1}^{n} I\{X_k \le t\} \frac{H_n(X_k)}{c(X_k)} (\psi^2)'(\sigma(X_k)\varepsilon_k) \\
&\quad + \frac{1}{n} \sum_{k=1}^{n} I\{X_k \le t\} H_n(X_k) \left(\frac{1}{D_n(X_k)} - \frac{1}{c(X_k)}\right) (\psi^2)'(\sigma(X_k)\varepsilon_k) \\
&=: D_{11} + D_{12},
\end{aligned}
$$

where the last identity defines D_{11} and D_{12} in an obvious manner. For the variable D_{11} we note that

$$E[D_{11}] = \frac{K(0)}{nh} \int_0^t \frac{g(x)}{c(x)} \int_{\mathbb{R}} \psi(\sigma(x)\omega)(\psi^2)'(\sigma(x)\omega) f_0(\omega) d\omega dx = O\left(\frac{1}{nh}\right).$$

Because of the antisymmetry of the function $\psi(\sigma(x)\omega)((\psi^2)'(\sigma(x)\omega))^2 f_0(\omega)$ (with respecht to ω) we obtain by a tedious calculation

$$
\begin{aligned}
E[D_{11}^2] &= \frac{1}{n} E\left[I\{X_1 \le t\} \frac{H_n^2(X_1)}{c^2(X_1)} ((\psi^2)'(\sigma(X_1)\varepsilon_1))^2 \right] \\
&\quad + \frac{n-1}{n} E\left[I\{X_1 \le t, X_2 \le t\} \frac{H_n(X_1)H_n(X_2)}{c(X_1)c(X_2)} (\psi^2)'(\sigma(X_1)\varepsilon_1)(\psi^2)'(\sigma(X_2)\varepsilon_2) \right] \\
&= \frac{n-1}{n^3 h^2} \int_0^t \frac{g(x)}{c^2(x)} E\left[K^2\left(\frac{x-X_i}{h}\right) \psi^2(Y_i - m(x)) \right] E\left[((\psi^2)'(\sigma(x)\varepsilon_1))^2 \right] dx \\
&\quad + \frac{1}{n^3 h^2} K^2(0) \int_0^t \frac{g(x)}{c^2(x)} E\left[\psi^2(\sigma(x)\varepsilon_1)((\psi^2)'(\sigma(x)\varepsilon_1))^2 \right] dx \\
&\quad + \frac{n-1}{n^3 h^2} \int_0^t \int_0^t \frac{g(x_1)g(x_2)}{c^2(x_1)c^2(x_2)} K^2\left(\frac{x_1-x_2}{h}\right) E\left[\psi(m(x_2) - m(x_1) + \sigma(x_2)\varepsilon_2)(\psi^2)'(\sigma(x_2)\varepsilon_2) \right] \\
&\quad \times E\left[\psi(m(x_1) - m(x_2) + \sigma(x_1)\varepsilon_1)(\psi^2)'(\sigma(x_1)\varepsilon_1) \right] dx_1 dx_2 \\
&\quad + \frac{n-1}{n^3 h^2} K^2(0) \left(\int_0^t \frac{g(x)}{c^2(x)} E\left[\psi(\sigma(x)\varepsilon_1)(\psi^2)'(\sigma(x)\varepsilon_1) \right] dx \right)^2 \\
&= o\left(\frac{1}{n}\right),
\end{aligned}
$$

which yields $D_{11} = o_p(n^{-1/2})$. For the derivation of a corresponding estimate we note that

$$
\begin{aligned}
D_{12} &\le \frac{1}{n} \sum_{i=k}^n I\{X_k \le t\} \mid \hat{m}_n(X_k) - m(X_k) \mid \cdot \left| \frac{c(X_k) - D_n(X_k)}{c(X_k)} \right| \cdot \mid (\psi^2)'(\sigma(X_k)\varepsilon_k) \mid \\
&\le \max_{1 \le k \le n} \mid \hat{m}_n(X_k) - m(X_k) \mid \max_{1 \le k \le n} \left| \frac{c(X_k) - D_n(X_k)}{c(X_k)} \right| \cdot \frac{1}{n} \sum_{k=1}^n I\{X_k \le t\} \mid (\psi^2)'(\sigma(X_k)\varepsilon_k) \mid.
\end{aligned}
$$

A Taylor expansion of the function in $D_n(X_k)$ yields

$$| D_n(X_k) - c(X_k) | \leq \left| \frac{1}{nh} \sum_{i \neq k} K \left(\frac{X_k - X_i}{h} \right) \psi'(Y_i - m(X_k)) - c(X_k) \right|$$

$$+ \frac{1}{nh} \sum_{i=1}^{n} \left| K \left(\frac{X_k - X_i}{h} \right) \psi''(Y_i - m(X_k)) \right| \omega_i \mid \hat{m}_n(X_k) - m(X_k) \mid$$

$$+ \frac{1}{nh} \sum_{i=1}^{n} \left| K \left(\frac{X_k - X_i}{h} \right) \psi''' \left(Y_i - m(X_k) + \tilde{\omega}_i \left(\hat{m}_n(X_k) - m(X_k) \right) \right) \right|$$

$$\times \frac{\omega_i^2}{2} (\hat{m}_n(X_k) - m(X_k))^2 + O_p \left(\frac{1}{nh} \right)$$

where $\tilde{\omega}_i \in (0, \omega_i)$. Obviously

$$E \left[\frac{1}{nh} \sum_{i \neq k} K \left(\frac{X_k - X_i}{h} \right) \psi'(Y_i - m(X_k)) - c(X_k) \right] = O \left(\frac{1}{n} \right)$$

and

$$\text{Var} \left(\frac{1}{nh} \sum_{i \neq k} K \left(\frac{X_k - X_i}{h} \right) \psi'(Y_i - m(X_k)) - c(X_k) \right)$$

$$= \frac{n-1}{n^2} E \left[\frac{1}{h^2} K^2 \left(\frac{X_k - X_i}{h} \right) (\psi')^2 (Y_i - m(X_k)) \right] - 2 \frac{n-1}{n^2} E^2 \left[\frac{1}{h} K \left(\frac{X_k - X_i}{h} \right) \psi'(Y_i - m(X_k)) \right]$$

$$= O \left(\frac{1}{n} \right).$$

Therefore $\mid D_n(X_k) - c(X_k) \mid = o_p \left(n^{-1/4} \right) + O_p \left(\frac{1}{nh} \right) + O_p(h^2)$ uniformly with respect to $k = 1, \ldots, n$ and $D_{12} = O_p \left(h^4 \right) = o_p \left(n^{-1/2} \right)$, which completes the proof. $\square$

References

Billingsley, P. (1999). *Convergence of Probability Measures. 2nd Ed.* Wiley Series in Probability and Statistics, Chichester.

Breusch, T. and Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrika*, 47:1287–1294.

Cook, R. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, 70:1–10.

Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *J. Statist. Planning Infer.*, 103:311–329.

Dette, H. and Hetzler, B. (2009). A simple test for the parametric form of the variance function in nonparametric regression. *Annals of the Institute of Statistical Mathematics, to appear.*

Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *J. R. Statist. Soc., Ser. B*, 60:693–708.

Dette, H., Neumeyer, N., and 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*, 69:903–917.

Diblasi, A. and Bowman, A. (1997). Testing for constant variance in a linear model. *Statist. Probab. Lett.*, 33:95–103.

Francisco-Fernández, M. and Vilar-Fernández, J. (2005). Two tests for heteroscedasticity in nonparametric regression. Technical report, Universidad de Coruña.

Härdle, W. (1984). Robust regression function estimation. *Journal of Multivariate Analysis*, 14:169–180.

Härdle, W. and Gasser, T. (1984). Robust non-parametric function fitting. *Journal of the Royal Statistical Society - Series B*, 1:42–51.

Harrison, M. and McCabe, B. (1979). A test for heteroscedasticity based on least squares residuals. *J. Amer. Statist. Assoc.*, 74:494–500.

Koenker, R. and Bassett, G. (1981). Robust tests for heteroscedasticity based on regression quantiles. *Econometrika*, 50:43–61.

Leung, D. H.-Y. (2005). Cross-validation in nonparametric regression with outliers. *Ann. Stat.*, 33:2291–2310.

Liero, H. (2003). Testing homoscedasticity in nonparametric regression. *J. Nonpar. Statist.*, 15:31–51.

Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.*, 37:66–89.

Sacks, J. and Ylvisaker, N. D. (1968). Designs for regression problems with correlated errors; many parameters. *Ann. Math. Statist.*, 39:49–69.

Stützle, W. and Mittal, Y. (1979). Some comments on the asymptotic behaviour of robust smoothers. In Gasser, T. and Rosenblatt, M., editors, *Smoothing techniques for Curve Estimation, Lecture Notes in Mathematics*. Springer-Verlag, Heidelberg.

You, J. and Chen, G. (2005). Testing heteroscedasticity in partially linear regression models. *Statist. Probab. Lett.*, 73(1):61–70.

p	model	Dette/Munk				Dette/Hetzler				(4.1)			
		2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
10%	I	.028	.050	.098	.167	.027	.038	.068	.148	.025	.050	.106	.196
	II	.035	.054	.100	.179	.020	.031	.064	.133	.023	.048	.092	.181
	III	.031	.051	.088	.169	.026	.044	.073	.147	.025	.052	.107	.203
20%	I	.029	.048	.084	.158	.013	.026	.047	.119	.023	.044	.089	.198
	II	.028	.049	.087	.159	.019	.030	.053	.119	.024	.055	.108	.214
	III	.030	.050	.088	.165	.014	.024	.049	.122	.020	.047	.102	.194
0%	I	.032	.058	.106	.192	.052	.080	.138	.235	.024	.053	.111	.229
	II	.038	.054	.102	.183	.067	.092	.147	.236	.027	.051	.100	.223
	III	.031	.051	.092	.185	.057	.086	.136	.241	.033	.054	.106	.225

Table 1: *Rejection probabilities of the test (4.1) for a constant scale factor. The sample size is $n = 50$ and the constant c in model (I)-(III) is given by $c = 0$, which corresponds to the null hypothesis of a constant scale factor.*

p	model	Dette/Munk				Dette/Hetzler				(4.1)			
		2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
10%	I	.031	.053	.095	.174	.015	.023	.051	.123	.023	.043	.095	.200
	II	.029	.050	.087	.171	.014	.025	.056	.149	.023	.044	.106	.191
	III	.032	.052	.092	.169	.014	.025	.061	.120	.026	.043	.087	.208
20%	I	.020	.042	.075	.155	.012	.022	.053	.141	.014	.033	.077	.182
	II	.026	.046	.084	.154	.003	.015	.052	.144	.010	.030	.083	.157
	III	.027	.041	.078	.147	.006	.017	.051	.131	.018	.031	.069	.165
0%	I	.031	.051	.091	.192	.041	.063	.123	.206	.028	.056	.112	.225
	II	.029	.052	.094	.189	.037	.066	.114	.213	.030	.054	.118	.223
	III	.031	.054	.096	.187	.028	.059	.107	.210	.031	.055	.108	.216

Table 2: *Rejection probabilities of the test (4.1) for a constant scale factor. The sample size is $n = 100$ and the constant c in model (I)-(III) is given by $c = 0$, which corresponds to the null hypothesis of a constant scale factor.*

		n = 50				n = 100				n = 200			
p	model	2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
	I	.100	.151	.228	.357	.166	.256	.371	.517	.420	.540	.659	.783
10%	II	.089	.168	.327	.582	.195	.424	.716	.933	.825	.963	.999	1
	III	.218	.298	.426	.563	.433	.556	.680	.807	.799	.860	.916	.967
	I	.064	.112	.192	.326	.119	.198	.310	.444	.316	.409	.538	.668
20%	II	.062	.113	.234	.465	.112	.245	.472	.805	.496	.765	.951	.996
	III	.144	.216	.322	.466	.381	.481	.615	.748	.699	.777	.870	.930
	I	.133	.226	.343	.480	.280	.385	.518	.651	.552	.669	.773	.868
0%	II	.082	.172	.402	.762	.318	.616	.899	.992	.983	1	1	1
	III	.293	.409	.537	.643	.574	.686	.800	.890	.907	.951	.974	.992

Table 3: *Rejection probabilities of the test (4.1) for a constant scale factor. The constant c in model (I)-(III) is given by c = 0.5, which corresponds to an alternative.*

		n = 50				n = 100				n = 200			
p	model	2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
	I	.317	.411	.546	.688	.617	.726	.815	.896	.938	.969	.989	.994
10%	II	.112	.272	.512	.844	.627	.881	.985	.999	1	1	1	1
	III	.532	.625	.735	.823	.806	.886	.938	.967	.993	1	1	1
	I	.205	.310	.446	.593	.479	.595	.717	.825	.868	.923	.961	.987
20%	II	.080	.176	.397	.727	.336	.635	.893	.988	.970	.994	1	1
	III	.382	.501	.622	.746	.704	.803	.885	.943	.974	.985	.993	.997
	I	.419	.529	.655	.770	.766	.833	.904	.946	.979	.995	.997	1
0%	II	.136	.341	.723	.966	.827	.983	1	1	1	1	1	1
	III	.572	.663	.771	.869	.891	.938	.974	.989	.999	1	1	1

Table 4: *Rejection probabilities of the test (4.1) for a constant scale factor. The constant c in model (I)-(III) is given by c = 1, which corresponds to an alternative.*

p	model	$n=50$				$n=100$				$n=200$			
		2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
10%	I	.071	.108	.167	.291	.103	.153	.234	.359	.189	.257	.369	.514
	II	.096	.135	.203	.319	.103	.153	.241	.423	.113	.217	.359	.592
	III	.198	.274	.376	.514	.312	.419	.528	.645	.527	.630	.730	.810
20%	I	.055	.089	.152	.246	.065	.104	.166	.273	.080	.126	.218	.354
	II	.062	.091	.150	.251	.059	.106	.178	.307	.064	.125	.219	.411
	III	.138	.188	.217	.391	.167	.255	.380	.496	.303	.403	.498	.628
0%	I	.169	.251	.352	.499	.252	.342	.481	.625	.510	.626	.750	.836
	II	.146	.189	.275	.420	.174	.259	.406	.617	.357	.529	.769	.948
	III	.301	.432	.566	.725	.527	.667	.799	.892	.902	.941	.976	.989

Table 5: *Rejection probabilities of the test of Dette and Hetzler (2009) for a constant scale factor. The constant c in model (I)-(III) is given by c = 0.5, which corresponds to an alternative.*

p	model	$n=50$				$n=100$				$n=200$			
		2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
10%	I	.050	.085	.128	.210	.053	.082	.134	.241	.069	.105	.170	.392
	II	.170	.244	.348	.491	.295	.380	.489	.639	.551	.461	.668	.794
	III	.068	.104	.173	.272	.104	.149	.233	.331	.151	.213	.316	.450
20%	I	.044	.071	.127	.214	.046	.066	.114	.222	.061	.083	.139	.240
	II	.179	.245	.339	.486	.294	.390	.498	.618	.397	.485	.641	.754
	III	.066	.115	.166	.265	.090	.126	.188	.285	.115	.173	.244	.376
0%	I	.041	.066	.134	.234	.061	.101	.153	.246	.078	.121	.185	.302
	II	.186	.264	.381	.520	.333	.432	.551	.705	.553	.654	.770	.882
	III	.068	.106	.174	.286	.092	.148	.232	.367	.169	.234	.359	.491

Table 6: *Rejection probabilities of the test of Dette and Munk (1998) for a constant scale factor. The constant c in model (I)-(III) is given by c = 0.5, which corresponds to an alternative.*