ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Dette, Holger; van Keilegom, Ingrid

Working Paper

new test for the parametric form of the variance function in nonparametric regression

Technical Report, No. 2005,32

Provided in Cooperation with:
Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475), University of Dortmund

Suggested Citation: Dette, Holger; van Keilegom, Ingrid (2005) : new test for the parametric form of the variance function in nonparametric regression, Technical Report, No. 2005,32, Universität Dortmund, Sonderforschungsbereich 475 - Komplexitätsreduktion in Multivariaten Datenstrukturen, Dortmund

This Version is available at:
https://hdl.handle.net/10419/22622

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

A new test for the parametric form of the variance function in nonparametric regression

Holger Dette[1] Ingrid Van Keilegom[2]

Ruhr-Universität Bochum *Université catholique de Louvain*

September 9, 2005

Abstract

In the common nonparametric regression model the problem of testing for the parametric form of the conditional variance is considered. A stochastic process based on the difference between the empirical processes obtained from the standardized nonparametric residuals under the null hypothesis (of a specific parametric form of the variance function) and the alternative is introduced and its weak convergence established. This result is used for the construction of a Cramér von Mises type statistic for testing the parametric form of the conditional variance. The finite sample properties of a bootstrap version of this test are investigated by means of a simulation study. In particular the new procedure is compared with some of the currently available methods for this problem and its performance is illustrated by means of a data example.

Key Words: Bootstrap; Kernel estimation; Nonparametric regression; Residual distribution; Testing heteroscedasticity; Testing homoscedasticity.

[1] Financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt B1) is gratefully acknowledged. [2] Financial support from the IAP research network nr. P5/24 of the Belgian government (Belgian Science Policy) is gratefully acknowledged.

1 Introduction

Consider the common nonparametric heteroscedastic regression model

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

where (X, Y) denotes a random vector, Y is a possible transformation of the variable of interest, X is a covariate, and the centered error variable ε is independent of X with mean 0 and variance 1. The function $m(X) = E(Y|X)$ is the unknown regression function and $\sigma^2(X) = \mathrm{Var}(Y|X)$ is the unknown conditional variance function. The importance of being able to test data corresponding to model (1.1) for heteroscedasticity is widely recognized because under the additional assumption of homoscedasticity the statistical analysis can be simplified substantially in most cases. On the other hand, if the assumption of homoscedasticity is not met, efficient inference for the regression function requires that the heteroscedasticity is taken into account. This may result in transformations of the data, weighted least squares (or modified likelihood) procedures or the choice of a variable bandwidth in nonparametric kernel smoothing [see Müller and Stadtmüller (1987)].

Early work on detecting heteroscedasticity is based on diagnostic plots based on residuals after fitting a specific regression model to the data. Harrison and McCabe (1979), Breusch and Pagan (1979), Koenker and Bassett (1981), Cook and Weisberg (1983), Carroll and Ruppert (1988), Sec. 3.4, and Diblasi and Bowman (1997) developed formal tests for the form of the variance function under the additional assumption that a parametric model for the regression function can be specified. A diagnostic test under a smoothness assumption on the regression function and the assumption of a normally distributed error distribution has been proposed by Eubank and Thomas (1993). More recent work on testing homoscedasticity considers a completely nonparametric specification of the regression model (1.1). Dette and Munk (1998) used an estimate of the L^2-distance between the variance function and its integral as test statistic, while Zhu, Fujikoshi and Naito (2001) proposed a test based on a marked empirical process of the squared residuals for testing homoscedasticity. Recently Dette (2002), Liero (2003) and Francisco-Fernández and Vilar-Fernández (2005) used estimates for the L^2-distance between the variance estimators in both models – the underlying heteroscedastic model and the hypothetical homoscedastic model – as test statistic.

It is the purpose of the present paper to present an alternative approach for the problem of testing for a parametric form of the variance function in the nonparametric regression model (1.1). Our investigations are mainly motivated by the observation that all papers

on testing homoscedasticity in a completely nonparametric regression model consider test statistics based on squared residuals from a nonparametric fit. In contrast to this approach we are interested in procedures, which use the residuals directly. Our interest in such methods is motivated twofold. On one hand we expect tests based on nonparametric residuals to be more sensitive for detecting deviations from homoscedasticity in the data. On the other hand the consideration of the residuals instead of their squares is more naturally related to the commonly used graphical procedures based on visual examination of the residuals [see Atkinson (1985)].

Moreover, the literature published so far is restricted to tests for homoscedasticity but there are numerous applications where a test for given parametric form of the variance function is required. Thus a further difference to the work cited in the previous paragraph is that we consider the more general problem of testing for a specific parametric form of the variance function (see our formulation of the null hypothesis in Section 2). Finally, in the case of testing for homoscedaticity the test statistic we will develop has the nice property that it can detect local alternatives converging to the null hypothesis of constant variance with a rate faster than root-n, whereas all other existing tests only attain the rate root-n or even slower. We propose to compare the empirical processes of the standardized residuals from a nonparametric fit under the null hypothesis (of a specific parametric form, e.g. homoscedasticity) and under the alternative (e.g. heteroscedasticity) and reject the null hypothesis for large differences between these processes. In Section 2 we present some preliminary notation and a motivation of our approach. Section 3 contains the main results. We derive a stochastic expansion for the difference between the empirical processes of the standardized residuals under the null hypothesis and under the alternative and use this result to prove the weak convergence of the corresponding difference process. As a consequence a Kolmogorov-Smirnov and a Cramér-von Mises test based on the difference process are proposed. In Section 4 we study the finite sample properties of a bootstrap version of the new test and demonstrate that this procedure yields tests with a reliable approximation of the nominal level and reasonable power. We also compare the test with the currently available methods proposed in Dette and Munk (1998), Zhu, Fujikoshi and Naito (2001), Dette (2002) and Francisco-Fernández and Vilar-Fernández (2005), and demonstrate that in many cases the new procedure yields substantially larger power in the problem of testing for homoscedasticity. Finally, we analyze data on the concentration of sulfate in rain at two towns in North Carolina. Some technical details are deferred to the Appendix.

2 Notation and motivation of the test statistic

Consider the random vector (X, Y) satisfying model (1.1) and define $F_\varepsilon(y) = P(\varepsilon \leq y)$, $F(y|x) = P(Y \leq y|X = x)$ and $F_X(x) = P(X \leq x)$ as the distribution function of the error, the conditional distribution function of Y given $X = x$ and the distribution function of the predictor, respectively. The probability density functions of $F_\varepsilon(y)$ and $F_X(x)$ are assumed to exist and will be denoted by $f_\varepsilon(y)$ and $f_X(x)$, respectively. We are interested in a test for the hypothesis

$$H_0 : \sigma^2 \in \mathcal{M} \text{ versus } H_1 : \sigma^2 \notin \mathcal{M}, \tag{2.1}$$

where

$$\mathcal{M} = \{\sigma_\theta^2 : \theta \in \Theta\} \tag{2.2}$$

is some parametric class of variance functions and $\Theta \subset I\!R^p$ is a set of parameters satisfying some regularity conditions, which will be specified in the Appendix. Note that the hypothesis of homoscedasticity is obtained for $p = 1$, $\sigma_\theta^2(x) = \theta$, but (2.1) contains many other hypotheses of practical interest.

The easiest way to motivate our approach is to consider the problem of testing for the parametric form of the variance function in the nonparametric regression model (1.1) in terms of distribution functions. For this we introduce the random variables

$$\varepsilon = \frac{Y - m(X)}{\sigma(X)} \tag{2.3}$$

and

$$\varepsilon_0 = \frac{Y - m(X)}{\sigma_{\tilde{\theta}_0}(X)}, \tag{2.4}$$

where $\tilde{\theta}_0 \in \Theta$ denotes the parameter corresponding to the best approximation of the variance function σ^2 by elements of the class $\mathcal{M}$, that is

$$\tilde{\theta}_0 = \underset{\theta \in \Theta}{\operatorname{argmin}}\ E[\{(Y - m(X))^2 - \sigma_\theta^2(X)\}^2] \tag{2.5}$$

$$= \underset{\theta \in \Theta}{\operatorname{argmin}}\ E[\{\sigma^2(X) - \sigma_\theta^2(X)\}^2]. \tag{2.6}$$

Note that under H_0, $\tilde{\theta}_0$ equals θ_0, the true value of θ. Throughout this paper we assume that $\tilde{\theta}_0$ exists and is uniquely determined. If the null hypothesis

$$\sigma^2 \in \mathcal{M} = \{\sigma_\theta^2 \mid \theta \in \Theta\}$$

is satisfied, the random variables ε and ε_0 defined in (2.3) and (2.4) have the same distribution. In general these distributions are different and we obtain

$$P(\varepsilon_0 \leq y) \quad = \quad P\Big(\varepsilon \leq \frac{\sigma_{\tilde{\theta}_0}(X)}{\sigma(X)} y\Big). \tag{2.7}$$

The following result shows that the equality of the distributions of the random variables ε_0 and ε is equivalent to the hypothesis (2.1) of a specific parametric form for the conditional variance. The proof can be found in the Appendix.

Lemma 2.1 *Assume that all moments of the distribution of the error ε exist. The distributions of the random variables ε and ε_0 defined in (2.3) and (2.4) coincide if and only if the hypothesis $H_0 : \sigma^2 \in \mathcal{M} = \{\sigma_\theta^2 \mid \theta \in \Theta\}$ is satisfied.*

The construction of a test statistic for the hypothesis (2.1) follows exactly the same pattern, where we replace the unknown distributions of the random variables ε and ε_0 by appropriate estimates. For this let (X_i, Y_i), $i = 1, \ldots, n$, denote independent replications of (X, Y). We first estimate the distribution of the error ε in a nonparametric way. Define for any x in the support $[a, b]$ of the distribution F_X of X the estimates

$$\hat{m}(x) = \sum_{i=1}^{n} W_i(x, h) Y_i, \quad \hat{\sigma}^2(x) = \sum_{i=1}^{n} W_i(x, h)(Y_i - \hat{m}(X_i))^2, \tag{2.8}$$

where

$$W_i(x, h) = \frac{K\left(\frac{x - X_i}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right)} \tag{2.9}$$

are the Nadaraya-Watson weights [see Nadaraya (1964) or Watson (1964)], K is a known probability density function (kernel) and h is an appropriate bandwidth. This leads to

$$\hat{F}_\varepsilon(y) = n^{-1} \sum_{i=1}^{n} I(\hat{\varepsilon}_i \leq y), \tag{2.10}$$

as an estimate of the distribution function F_ε, where

$$\hat{\varepsilon}_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)}$$

are the nonparametric residuals. This estimator (modulo a slightly different variance estimator) has been recently proposed and studied by Akritas and Van Keilegom (2001). Next, under the null hypothesis H_0 we estimate the variance function $\sigma^2(x)$ by $\sigma_{\hat{\theta}}^2(x)$, where $\hat{\theta}$ is a minimizer (over $\theta \in \Theta$) of the expression

$$S_n(\theta) = n^{-1} \sum_{i=1}^{n} [(Y_i - \hat{m}(X_i))^2 - \sigma_\theta^2(X_i)]^2, \tag{2.11}$$

and we smooth the parametric estimate by the same bandwidth and kernel as for the nonparametric estimator $\hat{\sigma}^2(x)$, i.e.

$$\hat{\sigma}_0^2(x) = \sum_{i=1}^{n} W_i(x,h)\sigma_{\hat{\theta}}^2(X_i).$$

This leads to

$$\hat{F}_{\varepsilon 0}(y) = n^{-1}\sum_{i=1}^{n} I(\hat{\varepsilon}_{i0} \leq y), \tag{2.12}$$

as an estimator of the distribution function of the random variable ε_0 defined in (2.4), where

$$\hat{\varepsilon}_{i0} = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}_0(X_i)}$$

are the standardized residuals estimated under the null hypothesis H_0.
Under the null hypothesis (2.1) one expects not too large deviations between the empirical distribution functions $\hat{F}_{\varepsilon 0}$ and $\hat{F}_{\varepsilon}$. Consequently we propose the Kolmogorov-Smirnov type statistic

$$T_{KS} = n^{1/2} \sup_{-\infty<y<\infty} |\hat{F}_{\varepsilon}(y) - \hat{F}_{\varepsilon 0}(y)|$$

and the Cramér-von Mises type statistic

$$T_{CM} = n \int [\hat{F}_{\varepsilon}(y) - \hat{F}_{\varepsilon 0}(y)]^2 \, d\hat{F}_{\varepsilon}(y)$$

for testing the hypothesis (2.1) of a specific parametric form of the variance function in model (1.1).

3 Main results

In the following we will study some asymptotic properties of the these statistics under the null hypothesis (2.1) and under local (Pitman) alternatives of the form

$$H_1 : \sigma^2(\cdot) \equiv \sigma_{\theta_0}^2(\cdot) + n^{-1/2} r(\cdot), \tag{3.1}$$

for some function r (note that the null hypothesis H_0 is obtained for $n \to \infty$). We introduce the following additional notation:

$$\Omega \;=\; \left\{ E\left[\frac{\partial \sigma_\theta^2(X)}{\partial \theta_r}\Big|_{\theta=\theta_0} \frac{\partial \sigma_\theta^2(X)}{\partial \theta_s}\Big|_{\theta=\theta_0} \right] \right\}_{r,s=1,\ldots,p}, \tag{3.2}$$

and

$$\frac{\partial \sigma_\theta^2(x)}{\partial \theta} = \left(\frac{\partial \sigma_\theta^2(x)}{\partial \theta_r} \right)_{r=1,\ldots,p}$$

is the gradient of the variance function σ_θ^2 with respect to $\theta = (\theta_1, \ldots, \theta_p)'$ (here and throughout this paper we assume its existence). The assumptions mentioned in the results below can be found in the Appendix, as well as the proofs of Theorems 3.1 and 3.4.

Theorem 3.1 *Assume that the conditions (A1)-(A3) in the Appendix are satisfied. Then, under the null hypothesis H_0, the following stochastic expansion is valid:*

$$\begin{aligned}
&\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon 0}(y) \\
&= \frac{y f_\varepsilon(y)}{2} \Bigg\{ n^{-1} \sum_{i=1}^n \sigma^{-2}(X_i)[\{Y_i - m(X_i)\}^2 - \sigma^2(X_i)] \\
&\quad - \int \sigma^{-2}(x) \left(\frac{\partial \sigma_\theta^2(x)}{\partial \theta}\Big|_{\theta=\theta_0} \right)' dF_X(x)\, \Omega^{-1}\, n^{-1} \sum_{i=1}^n [\{Y_i - m(X_i)\}^2 - \sigma^2(X_i)] \frac{\partial \sigma_\theta^2(X_i)}{\partial \theta}\Big|_{\theta=\theta_0} \Bigg\} \\
&\quad + R_n(y),
\end{aligned}$$

where the random process $\{R_n(y)\}_{y \in I\!R}$ satisfies

$$\sup_{-\infty < y < \infty} |R_n(y)| = o_P(n^{-1/2}).$$

The first term on the right hand side of the above expansion is the result of estimating the variance function $\sigma^2(\cdot)$ nonparametrically by $\hat{\sigma}^2(\cdot)$, whereas the second term comes from the parametric estimator $\hat{\sigma}_0^2(\cdot)$. Note that the estimation of $m(\cdot)$ does not contribute to the main term of this expansion. This is because the same estimator of $m(\cdot)$ is used in the expressions of $\hat{\varepsilon}_i$ and $\hat{\varepsilon}_{i0}$ $(i = 1, \ldots, n)$, and hence the contribution of this estimator to the main term cancels out [see our proof in the Appendix]. The following Corollary is an immediate consequence of Theorem 3.1, since the main term in the above representation factorizes in a deterministic function only dependent on y and a sum of i.i.d. terms independent of y.

Corollary 3.2 *Assume that the conditions (A1)-(A3) in the Appendix are satisfied. Then, under the null hypothesis H_0, the process $\{n^{1/2}(\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon 0}(y))\}_{y \in I\!R}$ converges weakly to the process $\{y f_\varepsilon(y) W\}_{y \in I\!R}$, where W is a zero-mean normal random variable with variance*

$$Var(W) = \frac{1}{4} E\left[\varepsilon^4 - 1\right] E\left[\left(1 - \int \frac{\sigma_{\theta_0}^2(X)}{\sigma_{\theta_0}^2(x)} \Big(\frac{\partial \sigma_\theta^2(x)}{\partial \theta}\Big|_{\theta=\theta_0} \Big)' dF_X(x) \Omega^{-1} \frac{\partial \sigma_\theta^2(X)}{\partial \theta}\Big|_{\theta=\theta_0} \right)^2 \right].$$

We are now ready to establish the weak convergence of the test statistics T_{KS} and T_{CM}.

Corollary 3.3 *Assume that the conditions (A1)-(A3) in the Appendix are satisfied. Then, under the null hypothesis H_0,*

$$T_{KS} \xrightarrow{d} \sup_{-\infty<y<\infty} |y f_\varepsilon(y)|\, |W| \ ,$$

$$T_{CM} \xrightarrow{d} \int y^2 f_\varepsilon^2(y)\, dF_\varepsilon(y)\, W^2,$$

where the random variable W has been defined in Corollary 3.2.

The proof for the statistic T_{KS} follows from the continuous mapping theorem, while for T_{CM} the proof mimics almost exactly that of Corollary 3.3 in Van Keilegom, González-Manteiga and Sánchez-Sellero (2005).

Note that the limiting process in Corollary 3.2 has an extremely simple structure, as it factorizes in a deterministic function and a random variable independent of y. However, the deterministic factor depends on the unknown density of the error distribution, which may be difficult to estimate in many cases, although estimators for this density have been proposed and studied in the literature (see e.g. Van Keilegom and Veraverbeke (2002)). Therefore we propose the application of a smoothed bootstrap procedure for the calculation of the critical values (see our discussion in Section 4). We conclude this section considering the limiting behavior of the two test statistics under the local alternative H_1 and two illustrative examples.

Theorem 3.4 *Assume that the conditions (A1)-(A4) in the Appendix are satisfied. Then, under local alternatives of the form (3.1),*

$$T_{KS} \xrightarrow{d} \sup_{-\infty<y<\infty} |y f_\varepsilon(y)|\, |W+b|$$

$$T_{CM} \xrightarrow{d} \int y^2 f_\varepsilon^2(y)\, dF_\varepsilon(y)\, (W+b)^2,$$

where

$$\begin{aligned} b \ &= \ -\int \sigma_{\theta_0}^{-2}(x) \Big(\frac{\partial \sigma_\theta^2(x)}{\partial \theta}\Big|_{\theta=\theta_0} \Big)' dF_X(x)\, \Omega^{-1} \int r(x) \frac{\partial \sigma_\theta^2(x)}{\partial \theta}\Big|_{\theta=\theta_0} dF_X(x) \\ &\quad + \int \sigma_{\theta_0}^{-2}(x) r(x) dF_X(x). \end{aligned}$$

Example 3.1 In the important problem of testing for homoscedasticity (i.e. $H_0 : \sigma^2(\cdot) \equiv \theta$ for some $\theta > 0$), it follows that $\partial \sigma_\theta^2 / \partial \theta = 1$ and a straightforward calculation shows

that the main term of the asymptotic representation in Theorem 3.1 vanishes, i.e.

$$\sup_{-\infty<y<\infty} |\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon 0}(y)| = o_P(n^{-1/2}).$$

As a consequence the limit distribution in Corollary 3.2, 3.3 and 3.4 degenerates to a Dirac measure concentrated at the point 0. Hence, in the problem of testing for homoscedasticity, tests based on the process

$$\{n^{1/2}(\hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon 0}(y))\}_{y\in I\!R}$$

can detect local alternatives converging to the null hypothesis with a rate faster than root-n.

Example 3.2 The following example shows that one cannot expect that the statement of Example 3.1 is correct for more general hypotheses on the conditional variance. Consider for example the case $p = 1$ and $\sigma_\theta^2(x) = \theta k(x)$ for a given nonnegative function k. In this case the distribution of the random variable W is a centered normal with variance

$$\mathrm{Var}(W) = \frac{1}{4}E[\varepsilon^4 - 1]\mathrm{Var}[k^2(X)]/E[k^2(X)]^2,$$

which vanishes if and only the function k is constant.

4 Finite sample properties and a data example

4.1 A simulation study

In this section we study the finite sample properties of the new test and compare it with four other procedures, which are currently available in the literature for testing for homoscedasticity. We limit attention to the Cramér-von Mises test and demonstrate that in many cases the new procedure yields a larger power. Note that the asymptotic distribution of this test depends on the unknown density of the error ε. Although this can be estimated in principle, we decided to implement a smooth bootstrap version of the test. To be precise, first define

$$\varepsilon_i^* = \tilde{\varepsilon}_i + vN_i \quad i = 1, \ldots, n \tag{4.1}$$

where $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$ are a sample of i.i.d. observations with distribution function $\hat{F}_\varepsilon$, $N_1, \ldots, N_n$ are i.i.d. standard normal distributed random variables independent of $\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_n$, and v

is a smoothing parameter. In a next step we generate bootstrap data according to the model

$$Y_i^* = \hat{m}(X_i) + \sigma_{\hat{\theta}}(X_i)\varepsilon_i^*,$$

where $\sigma^2_{\hat{\theta}}(\cdot)$ is the estimate of the variance function under the null hypothesis (2.1), and calculate the corresponding Cramér-von Mises statistic T^*_{CM} from the bootstrap data. If B bootstrap replications have been performed and $T^{*(1)}_{CM} < \ldots < T^{*(B)}_{CM}$ denote the order statistics of the calculated bootstrap sample, the null hypothesis (2.1) is rejected if

$$T_{CM} > T^{*(\lfloor B(1-\alpha)\rfloor)}_{CM}. \tag{4.2}$$

$B = 100$ bootstrap replications are performed to calculate the rejection probabilities of the test (4.2) and 1000 simulation runs are used for each scenario. In order to investigate the impact of the bandwidth for the Nadaraya-Watson weights in the testing procedure we considered the two cases $h = 0.3$ and $h = 0.5$ and for the factor v in (4.1) we used $v = 0.2$.

Example 4.1. Our first example considers the problem of testing for homoscedasticity in the nonparametric regression model (1.1). In particular we consider the nonparametric regression models

$$\begin{array}{rrcllcl} (I) & m(x) & = & 1+\sin x & \sigma(x) & = & \sigma\exp(cx) \\ (II) & m(x) & = & 1+x & \sigma(x) & = & \sigma(1+c\cdot\sin(10x))^2 \\ (III) & m(x) & = & 1+x & \sigma(x) & = & \sigma(1+cx)^2, \end{array} \tag{4.3}$$

where the standard deviation is chosen as $\sigma = 0.5$ and the parameter c is given by $0, 0.5$ and 1.0. In Table 4.1 and 4.2 we show the simulated rejection probabilities for the sample sizes $n = 50$, 100, 200 and for the two bandwidths $h = 0.3$ and $h = 0.5$, respectively. Note that the case $c = 0$ corresponds to the null hypothesis of homoscedasticity and that the test beds defined by (4.3) were also considered in a simulation study by Dette and Munk (1998), who used an estimate of $\mathrm{Var}[\sigma^2(X)]$ as test for homoscedasticity. Thus our results are directly comparable with the corresponding rejection probabilities given in Table 1 of this reference. The models (I) and (III) have also been considered in Dette (2002), who proposed to estimate $\mathrm{Var}[\sigma^2(X)]$ by smoothing methods, see Table 1 and 2 in this reference. Francisco-Fernández and Vilar-Fernández (2005) investigated Liero's (2003) proposal to use an L^2-distance between a nonparametric and a constant estimate of the variance function and also considered the models (I) and (III) [see Table 3-4 in this

reference]. These authors propose a further test for homoscedasticity, which is based on a combination of a polynomial least squares fit to the squared nonparametric residuals with a test if this polynomial is constant. Because this test is not consistent in general, it is not included in our comparison.

We observe a rather precise approximation of the nominal level by the new test in all cases (see the rows labeled with $c = 0$ in Table 4.1 and 4.2). Moreover, for model (I) and (III) the impact of the choice of the bandwidth h seems to be less critical compared to the oscillating case (II). In this situation the power of the bootstrap test decreases substantially with an increasing bandwidth. A comparison with the results of Dette and Munk (1998) and Dette (2002) shows that in the test beds (I) and (III) the new test yields substantially larger power than the test proposed by these authors. For example in model (I) with $c = 1$ the power of the new test at level $\alpha = 10\%$ with $n = 50$ observations is 0.614 and 0.651 corresponding to the cases $h = 0.3$ and $h = 0.5$, respectively. This power is even not reached by the test of Dette and Munk (1998) for $n = 200$ observations, where the power is 0.458. On the other hand the power of Dette's (2002) test is 0.566 for $n = 50$ observations. For a comparison with the test of Francisco-Fernández and Vilar-Fernández (2005) we have to consider the sample size $n = 100$, because these authors did not present results for smaller sample sizes. In this case the power of the new procedure is 0.922 ($h = 0.3$) and 0.930 ($h = 0.5$), while the power of the test based on the L^2-distance is 0.945. This slight improvement can be explained by the fact that the test of Francisco-Fernández and Vilar-Fernández (2005) does not keep the 10% level (it is 0.129, while the level of the new test is 0.103). This phenomenon was observed by many authors for L^2-type statistics [see e.g. Fan and Linton (2003)].

The situation in the test bed (III) is very similar: the power of the new test obtained for 50 observations is 0.810 and 0.814 corresponding to the cases $h = 0.3$ and $h = 0.5$, respectively, and exceeds the power of the test of Dette and Munk (1998) with $n = 200$ observations (which is 0.598). The power of Dette's (2002) test is 0.774 for $n = 50$ observations. For the sample size $n = 100$ the test of Francisco-Fernández and Vilar-Fernández (2005) has power 0.995, while the new test yields the rejection probabilities 0.991 and 0.990 corresponding to the cases $h = 0.3$ and $h = 0.5$ (note again that the test of Francisco-Fernández and Vilar-Fernández (2005) exceeds the 10%-level by more than 25%). The other cases for model (I) and (III) show a similar superiority of the new procedure.

In the situation of an oscillating variance function considered in model (II) the situation

is different. Here the power of the new test depends sensitively on the bandwidth chosen for the regression estimate. If $h = 0.5$ the test of Dette and Munk (1998) is more powerful in most cases, while a smaller bandwidth $h = 0.3$ yields a substantially larger power than the test of Dette and Munk (1998).

		$n = 50$			$n = 100$			$n = 200$		
$h = 0.3$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
(I)	0	0.038	0.058	0.101	0.038	0.057	0.103	0.032	0.052	0.106
	0.5	0.136	0.173	0.246	0.292	0.352	0.471	0.564	0.638	0.736
	1	0.431	0.502	0.614	0.822	0.863	0.922	0.991	0.997	0.999
(II)	0	0.033	0.061	0.109	0.039	0.054	0.099	0.037	0.055	0.109
	0.5	0.219	0.279	0.414	0.497	0.600	0.740	0.949	0.968	0.992
	1	0.466	0.539	0.650	0.789	0.838	0.902	0.988	0.993	0.999
(III)	0	0.035	0.057	0.108	0.038	0.061	0.099	0.038	0.049	0.098
	0.5	0.285	0.351	0.471	0.631	0.700	0.797	0.948	0.967	0.984
	1	0.651	0.708	0.810	0.959	0.973	0.991	1.000	1.000	1.000

Table 4.1. *Simulated rejection probabilities of the bootstrap test with bandwidth* $h = 0.3$ *for the regression and variance functions in (4.3). The error has a standard normal distribution and the explanatory variable* X *has a uniform distribution on the interval* $[0, 1]$.

		$n = 50$			$n = 100$			$n = 200$		
$h = 0.5$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
(I)	0	0.024	0.046	0.100	0.038	0.058	0.103	0.032	0.054	0.109
	0.5	0.137	0.182	0.217	0.294	0.362	0.487	0.592	0.676	0.777
	1	0.438	0.523	0.651	0.794	0.862	0.930	0.992	0.995	0.997
(II)	0	0.032	0.055	0.097	0.033	0.050	0.111	0.035	0.055	0.103
	0.5	0.072	0.092	0.157	0.084	0.124	0.200	0.095	0.152	0.284
	1	0.184	0.237	0.305	0.294	0.349	0.427	0.470	0.525	0.630
(III)	0	0.029	0.039	0.085	0.040	0.060	0.112	0.037	0.056	0.103
	0.5	0.261	0.323	0.455	0.653	0.731	0.822	0.935	0.957	0.981
	1	0.637	0.708	0.814	0.959	0.976	0.990	0.999	1.000	1.000

Table 4.2. *Simulated rejection probabilities of the bootstrap test with bandwidth* $h = 0.5$ *for the regression and variance functions in (4.3). The error has a standard normal distribution and the explanatory variable* X *has a uniform distribution on the interval* $[0, 1]$.

	$n = 50$			$n = 70$			$n = 100$		
$h = 0.3$	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
(IV)	0.767	0.826	0.916	0.961	0.974	0.993	1.000	1.000	1.000
(V)	0.994	0.996	0.998	1.000	1.000	1000	1.000	1.000	1.000
(VI)	0.916	0.961	0.997	0.991	0.999	1.000	1.000	1.000	1.000

Table 4.3. *Simulated rejection probabilities of the bootstrap test with bandwidth $h = 0.3$ for the regression and variance functions in (4.4). The error has a standard normal distribution and the explanatory variable X has a uniform distribution on the interval* $[0, 1]$.

	$n = 50$			$n = 70$			$n = 100$		
$h = 0.5$	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
(IV)	0.718	0.811	0.902	0.943	0.968	0.996	1.000	1.000	1.000
(V)	0.989	0.991	0.999	1.000	1.000	1.000	1.000	1.000	1.000
(VI)	0.902	0.950	0.999	0.987	0.992	1.000	1.000	1.000	1.000

Table 4.4. *Simulated rejection probabilities of the bootstrap test with bandwidth $h = 0.5$ for the regression and variance functions in (4.4). The error has a standard normal distribution and the explanatory variable X has a uniform distribution on the interval* $[0, 1]$.

Example 4.2. Our second example refers to the work of Zhu, Fujikoshi and Naito (2001), who proposed functionals of a marked empirical process of the squared nonparametric residuals as test statistic in the problem of testing for homoscedasticity. In a simulation study they investigated the finite sample properties for the following regression and variance functions :

$$\begin{array}{rrcl rcl} (IV) & m(x) & = & 1 + 2x & \sigma(x) & = & 0.25 + x \\ (V) & m(x) & = & 1 + 2x & \sigma(x) & = & 0.25 + 4(x - 0.25)^2 \\ (VI) & m(x) & = & 1 + 2x & \sigma(x) & = & 0.25 \exp(x \log 5). \end{array} \tag{4.4}$$

In Table 4.3 and 4.4 we display the rejection probabilities of the new Cramér von Mises (bootstrap) test proposed in Section 3 for the sample sizes $n = 50$, $n = 70$ and $n = 100$. The two tables correspond again to the bandwidths $h = 0.3$ and $h = 0.5$, respectively. We observe no substantial differences between the two choices of the bandwidth. The results in the tables for sample size $n = 50$ and $n = 70$ and the 5% level are directly comparable with Table 2 in Zhu, Fujikoshi and Naito (2001). We observe that in model (IV) and (VI) the rejection probabilities of the test of Zhu, Fujikoshi and Naito (2001) are very similar

to those obtained by the new test. However, in model (V) the power of the new test is at least three times larger as the power of the test of Zhu, Fujikoshi and Naito (2001). Thus the new procedure performs reasonably well in the examples considered by these authors.

		$n = 50$			$n = 100$			$n = 200$		
$h = 0.3$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
	0	0.034	0.055	0.110	0.034	0.049	0.096	0.023	0.095	0.093
(VII)	0.5	0.200	0.248	0.337	0.368	0.440	0.544	0.662	0.723	0.827
	1	0.383	0.456	0.604	0.715	0.794	0.888	0.956	0.976	0.990

Table 4.5. *Simulated rejection probabilities of the bootstrap test with bandwidth $h = 0.3$ for the one-parametric hypothesis (4.5). The regression and variance functions are given by (4.6), the error has a standard normal distribution and the explanatory variable X has a uniform distribution on the interval* $[0, 1]$.

		$n = 50$			$n = 100$			$n = 200$		
$h = 0.5$	c	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
	0	0.025	0.042	0.087	0.021	0.043	0.089	0.021	0.046	0.087
(VII)	0.5	0.167	0.218	0.304	0.262	0.321	0.416	0.417	0.473	0.586
	1	0.348	0.424	0.578	0.690	0.752	0.821	0.919	0 .942	0.967

Table 4.6. *Simulated rejection probabilities of the bootstrap test with bandwidth $h = 0.5$ for the one-parametric hypothesis (4.5). The regression and variance functions are given by (4.6), the error has a standard normal distribution and the explanatory variable X has a uniform distribution on the interval* $[0, 1]$.

Example 4.3. Our final example considers the problem of testing for a one-parametric class of variance functions, that is

$$H_0 : \sigma^2(x) = 1 + \theta x^2 \tag{4.5}$$

for some $\theta \in I\!R$. For this we simulate data according to the model

$$(VII) \qquad m(x) \;=\; 1 + x \qquad \sigma(x) \;=\; 1 + 3x^2 + 2.5c\sin(2\pi x) \tag{4.6}$$

where the case $c = 0$ corresponds to the null hypothesis and the choices $c = 0.5, 1$ to two alternatives. The errors in model (1.1) are standard normal and the covariates follow again a uniform distribution on the interval $[0, 1]$. The rejection probabilities of the Cramér von

Mises test are displayed in Table 4.5 and 4.6 corresponding to the bandwidths $h = 0.3$ and $h = 0.5$, respectively. We observe a good approximation of the nominal level (for both choices of the bandwidth) and reasonable rejection probabilities under the two alternatives ($c = 0.5, 1$). In this case the power of the test depends again sensitively on the choice of the bandwidth.

4.2 Data example

We consider data on (log of) concentration of sulfate in rain at two towns (Coweeta and Lewiston) in North Carolina. The data were previously analyzed by Hall and Hart (1990), in the context of bootstrap tests for nonparametric analysis of covariance, where the covariate is 'amount of rainfall'. The sulfate concentration was available on a weekly basis over a five year period from 1979 to 1983. There were 220 weeks among all 261 weeks where data were available at Coweeta and 215 weeks where data were available at Lewiston. Hall and Hart (1990) restricted their analysis to the 189 weeks where data were available at both towns.

A crucial assumption in the applicability of their test for the comparison of two regression curves is a constant (not necessary equal) variance of the observations in both towns. In order to investigate whether this assumption is justified we applied the new bootstrap test to each data set, respectively. The bandwidth h was chosen by least squares cross validation, $b = 500$ bootstrap replications were used and we transformed the covariate (week) onto the interval $[0, 1]$ by dividing by 261. For the Lewiston data a p-value of 0.397 and for the Coweeta data a p-value of 0.418 were observed. Hence, Hall and Hart's (1990) assumption of constant variability in both groups is strictly supported by our test.

Appendix : Proofs

In what follows $\|\cdot\|$ denotes the Euclidean norm. For our main asymptotic results we require the following regularity conditions.

(*A*1) :

(*i*) As $n \to \infty$: $h \to 0$, $nh^4 \to 0$ and $nh^{3+2\delta}(\log h^{-1})^{-1} \to \infty$ for some $\delta > 0$.

(*ii*) K has compact support, $\int uK(u)\,du = 0$ and K is twice continuously differentiable.

(*A*2) :
(*i*) F_X is three times continuously differentiable and $\inf_{a\le x\le b} f_X(x) > 0$.
(*ii*) $F'(y|x)$ exists, is continuous in (x, y) for all x and y and $\sup_{x,y} |y^2 F'(y|x)| < \infty$, and the same holds for all other partial derivatives of $F(y|x)$ with respect to x and y up to order two.
(*iii*) $m(x)$ is twice continuously differentiable.
(*iv*) All partial derivatives up to order three of $\sigma^2_\theta(x)$ with respect to x and the components of θ exist and are continuous in (x, θ) for all x and θ. Moreover, $\inf_{a\le x\le b} \sigma(x) > 0$.
(*v*) $E(\varepsilon^4) < \infty$.

(*A*3) :
(*i*) Θ is a compact subspace of $I\!R^p$.
(*ii*) θ_0 is an interior point of Θ.
(*iii*) For all $\varepsilon > 0$, $\inf_{\|\theta-\theta_0\|>\varepsilon} E[(\sigma^2_\theta(X) - \sigma^2_{\theta_0}(X))^2] > 0$.
(*iv*) The matrix Ω defined in (3.2) is non-singular.

(*A*4) : $E[r^2(X)] < \infty$ and $r(x)$ is twice continuously differentiable for all x.

A.1 Proof of Lemma 2.1.

The necessary part follows from (2.7). In order to show that the equality of the distributions of the random variables is also sufficient for the hypothesis $\sigma^2 \in \mathcal{M} = \{\sigma^2_\theta \mid \theta \in \Theta\}$, we note that it follows from (2.7) that

$$E[\varepsilon^{2n}] = E[\varepsilon_0^{2n}] = E\Big[\Big(\frac{\sigma^2(X)}{\sigma^2_{\tilde\theta_0}(X)}\varepsilon^2\Big)^n\Big] = E[\varepsilon^{2n}]E\Big[\Big(\frac{\sigma^2(X)}{\sigma^2_{\tilde\theta_0}(X)}\Big)^n\Big]$$

for all $n \in I\!N$, which implies that

$$E\Big[\Big(\frac{\sigma^2(X)}{\sigma^2_{\tilde\theta_0}(X)}\Big)^n\Big] = 1$$

for all $n \in I\!N$. Therefore we obtain from Carleman's condition [see e.g. Feller (1966) p. 228] that the distribution of the random variable $\sigma^2(X)/\sigma^2_{\tilde\theta_0}(X)$ coincides with the distribution of the constant random variable $U \equiv 1$, that is $\sigma^2 \in \mathcal{M} = \{\sigma^2_\theta \mid \theta \in \Theta\}$. $\square$

A.2 Proof of Theorem 3.1 and 3.4.

The proof requires several auxiliary results.

Lemma A.1 *If the assumptions of Theorem 3.1 are satisfied, then the following stochastic expansion is valid :*

$$\int \sigma^{-2}(x)[\hat{\sigma}^2(x) - \sigma^2(x)]\, dF_X(x)$$
$$= n^{-1}\sum_{i=1}^{n} \sigma^{-2}(X_i)[\{Y_i - m(X_i)\}^2 - \sigma^2(X_i)] + o_P(n^{-1/2}).$$

Proof. With the notation $K_h(\cdot) = h^{-1}K(\cdot/h)$ we obtain the decomposition

$$\begin{aligned}
\int \sigma^{-2}(x)\hat{\sigma}^2(x)\, dF_X(x) &= n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)\hat{f}_X^{-1}(x)\sigma^{-2}(x)\, dF_X(x)\, Y_i^2 \\
&\quad -2n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)\hat{f}_X^{-1}(x)\sigma^{-2}(x)\, dF_X(x)\, Y_i\hat{m}(X_i) \\
&\quad +n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)\hat{f}_X^{-1}(x)\sigma^{-2}(x)\, dF_X(x)\, \hat{m}^2(X_i) \\
&= T_1 + T_2 + T_3, \qquad \text{(A.1)}
\end{aligned}$$

where the last equality defines the random variables T_1, T_2, T_3. For the first term T_1 in (A.1) we have

$$\begin{aligned}
T_1 &= n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)\sigma^{-2}(x)[Y_i^2 - E(Y^2|X=x)]\, dx \\
&\quad +n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)[f_X(x) - \hat{f}_X(x)]\hat{f}_X^{-1}(x)\sigma^{-2}(x)[Y_i^2 - E(Y^2|X=x)]\, dx \\
&\quad + \int \sigma^{-2}(x)E(Y^2|X=x)\, dF_X(x) \\
&= n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)[Y_i^2 - E(Y_i^2|X_i)] + E[\sigma^{-2}(X)E(Y^2|X)] + o_P(n^{-1/2}). \qquad \text{(A.2)}
\end{aligned}$$

Using a similar decomposition as for T_1 we obtain

$$\begin{aligned}
T_2 &= -2n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)Y_i[\hat{m}(X_i) - E\{\hat{m}(X_i)|\boldsymbol{X}\}] \\
&\quad -2n^{-1}\sum_{i=1}^{n}\int K_h(x-X_i)\hat{f}_X^{-1}(x)\sigma^{-2}(x)\, dF_X(x)\, Y_i E\{\hat{m}(X_i)|\boldsymbol{X}\} + o_P(n^{-1/2}) \\
&= T_{2,1} + T_{2,2}, \qquad \text{(A.3)}
\end{aligned}$$

where the last line defines the random variables $T_{2,i}$ $(i = 1, 2)$ and $\boldsymbol{X} = (X_1, \ldots, X_n)$ is the vector of covariates. The second term on the right hand side of (A.3) equals (using again a similar derivation as for the stochastic expansion of T_1)

$$\begin{aligned} T_{2,2} &= -\frac{2}{n}\sum_{i=1}^{n}\int K_h(x - X_i)\hat{f}_X^{-1}(x)\sigma^{-2}(x)\, dF_X(x)\, Y_i m(X_i) + o_P(n^{-1/2}) \\ &= -\frac{2}{n}\sum_{i=1}^{n}\sigma^{-2}(X_i)[Y_i m(X_i) - m^2(X_i)] - 2E[\sigma^{-2}(X)m^2(X)] + o_P(n^{-1/2}), \end{aligned} \tag{A.4}$$

whereas the first term can be written as

$$\begin{aligned} T_{2,1} &= -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h(X_i - X_j)f_X^{-1}(X_i)\sigma^{-2}(X_i)Y_i(Y_j - m(X_j)) + o_P(n^{-1/2}) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} U_{nij} + o_P(n^{-1/2}) \ , \end{aligned}$$

where the random variables U_{nij} are defined by

$$U_{nij} = -2K_h(X_i - X_j)f_X^{-1}(X_i)\sigma^{-2}(X_i)Y_i(Y_j - m(X_j)).$$

Now a standard argument shows that

$$\begin{aligned} T_{2,1} &= n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}[U_{nij} - E\{U_{nij}|X_i, Y_i\} - E\{U_{nij}|X_j, Y_j\} + E\{U_{nij}\}] \\ &\quad + n^{-1}\sum_{i=1}^{n} E\{U_{ni1}|X_i, Y_i\} + n^{-1}\sum_{j=1}^{n} E\{U_{n1j}|X_j, Y_j\} - n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} E\{U_{nij}\} \\ &\quad + o_P(n^{-1/2}) \\ &= \sum_{k=1}^{4} A_{nk} + o_P(n^{-1/2}) \ , \end{aligned} \tag{A.5}$$

where the last line defines the random variables A_{nk} $(k = 1, \ldots, 4)$. For A_{n1}, note that $E[A_{n1}] = 0$ and hence, by Chebyshev's inequality, for any $K > 0$,

$$\begin{aligned} &P(|A_{n1}| > Kn^{-1}E\{U_{n12}^*(U_{n12}^* + U_{n21}^*)\}^{1/2}) \\ &\qquad \le K^{-2}n^2 E\{U_{n12}^*(U_{n12}^* + U_{n21}^*)\}^{-1}E(A_{n1}^2) \\ &\qquad = K^{-2}n^{-2}E\{U_{n12}^*(U_{n12}^* + U_{n21}^*)\}^{-1}\sum_{i,j}\sum_{l,m} E(U_{nij}^* U_{nlm}^*), \end{aligned} \tag{A.6}$$

where $U_{nij}^* = U_{nij} - E\{U_{nij}|X_i, Y_i\} - E\{U_{nij}|X_j, Y_j\} + E\{U_{nij}\}$. Since $E[U_{nij}^*] = 0$, the terms for which both i and j are different from l and m are zero. The terms for which

either i or j equals l or m and the other differs from l and m, are also zero, because, for example when $i = l$ and $j \neq m$,

$$E[U^*_{nij} E\{U^*_{nim} | X_i, Y_i, X_j, Y_j\}] = 0.$$

Thus, only the $2n^2$ terms for which (i, j) equals (l, m) or (m, l) are non-zero. Hence, (A.6) is bounded by K^{-2}, which can be made arbitrarily small for K large enough. It now follows that $A_{n1} = O_P(n^{-1}h^{-1/2}) = o_P(n^{-1/2})$, since it is easily seen that

$$E\{U^*_{n12}(U^*_{n12} + U^*_{n21})\} = O(h^{-1}).$$

Next, note that $A_{n2} = A_{n4} = 0$, and that

$$A_{n3} = -2n^{-1} \sum_{i=1}^{n} \sigma^{-2}(X_i) m(X_i)(Y_i - m(X_i)) + o_P(n^{-1/2}).$$

Hence it follows from (A.3)

$$T_2 \;=\; -\frac{4}{n} \sum_{i=1}^{n} \sigma^{-2}(X_i) m(X_i)[Y_i - m(X_i)] - 2E[\sigma^{-2}(X) m^2(X)] + o_P(n^{-1/2}). \quad \text{(A.7)}$$

It remains to consider the term T_3 in the decomposition (A.1), for which a similar argument as used for the stochastic expansion of the term T_1 yields

$$\begin{aligned} T_3 \;=\; & n^{-1} \sum_{i=1}^{n} \int K_h(x - X_i) \sigma^{-2}(x) [\hat{m}^2(X_i) - E\{\hat{m}(X_i) | \boldsymbol{X}\}^2] \, dx \qquad \text{(A.8)} \\ & + n^{-1} \sum_{i=1}^{n} \int K_h(x - X_i) \hat{f}_X^{-1}(x) \sigma^{-2}(x) E\{\hat{m}(X_i) | \boldsymbol{X}\}^2 \, dF_X(x) + o_P(n^{-1/2}). \end{aligned}$$

The second term of (A.8) can be written as $E[\sigma^{-2}(X) m^2(X)] + o_P(n^{-1/2})$, whereas the first term equals (using similar techniques as for the stochastic expansion of the term $T_{2,1}$)

$$\begin{aligned} & n^{-3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} K_h(X_i - X_j) K_h(X_i - X_k) f_X^{-2}(X_i) \sigma^{-2}(X_i) [Y_j Y_k - m(X_j) m(X_k)] \\ & + o_P(n^{-1/2}) \\ & = 2n^{-1} \sum_{i=1}^{n} \sigma^{-2}(X_i) m(X_i)[Y_i - m(X_i)] + o_P(n^{-1/2}). \qquad \text{(A.9)} \end{aligned}$$

It now follows from (A.1), (A.2), (A.7) and (A.9) that

$$\begin{aligned} & \int \sigma^{-2}(x) \hat{\sigma}^2(x) \, dF_X(x) \\ & = n^{-1} \sum_{i=1}^{n} \sigma^{-2}(X_i)[Y_i^2 - E(Y_i^2 | X_i)] + E[\sigma^{-2}(X) E(Y^2 | X)] \end{aligned}$$

$$
\begin{aligned}
&-2n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)m(X_i)[Y_i-m(X_i)]-E[\sigma^{-2}(X)m^2(X)]+o_P(n^{-1/2})\\
&=n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)[Y_i^2-E(Y^2|X_i)-2m(X_i)\{Y_i-m(X_i)\}+\sigma^2(X_i)]+o_P(n^{-1/2})\\
&=n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)[Y_i-m(X_i)]^2+o_P(n^{-1/2}).
\end{aligned}
$$

□

Lemma A.2 *If the assumptions of Theorem 3.1 are satisfied, then under the null hypothesis H_0,*

$$\hat{\theta}-\theta_0\to_P 0.$$

Proof. Recall the definition S_n in (2.11) and define

$$S_0(\theta)=E\{\sigma^4_{\theta_0}(X)\}E(\varepsilon^4-1)+E\{[\sigma^2_\theta(X)-\sigma^2_{\theta_0}(X)]^2\}, \tag{A.10}$$

It follows from Theorem 5.7 in van der Vaart (1998, p. 45) that it suffices to show that

$$
\begin{aligned}
&\sup_\theta|S_n(\theta)-S_0(\theta)|\to_P 0\\
&\inf_{\|\theta-\theta_0\|>\varepsilon}S_0(\theta)>S_0(\theta_0)\ \forall\varepsilon>0.
\end{aligned}
$$

The latter follows from assumption (A3)(iii), whereas for the former, note that

$$\sup_\theta|S_n(\theta)-S_0(\theta)|\le\sup_\theta|S_n(\theta)-\tilde{S}_n(\theta)|+\sup_\theta|\tilde{S}_n(\theta)-S_0(\theta)|, \tag{A.11}$$

where the random variable $\tilde{S}_n$ is given by

$$\tilde{S}_n(\theta)=n^{-1}\sum_{i=1}^{n}[(Y_i-m(X_i))^2-\sigma^2_\theta(X_i)]^2.$$

From the uniform consistency of $\hat{m}(\cdot)$ it follows that the first term on the right hand side of (A.11) is $o_P(1)$. For the second term, apply e.g. Theorem 2 in Jennrich (1969). □

Lemma A.3 *If the assumptions of Theorem 3.1 are satisfied, then under the null hypothesis H_0,*

$$\hat{\theta}-\theta_0=\Omega^{-1}\,n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\theta_0}+o_P(n^{-1/2}).$$

Proof. First note that $\hat{\theta} - \theta_0 = o_P(1)$ by Lemma A.2. Hence, it follows from assumption $(A3)(ii)$ that $\hat{\theta}$ is an interior point of Θ for n large enough. Now write

$$\frac{\partial S_n(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}} - \frac{\partial S_n(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = \frac{\partial^2 S_n(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_{1n}} (\hat{\theta} - \theta_0),$$

for some θ_{1n} between $\hat{\theta}$ and θ_0. Then, we obtain that

$$\hat{\theta} - \theta_0 = -\left(\frac{\partial^2 S_n(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_{1n}}\right)^{-1} \frac{\partial S_n(\theta)}{\partial \theta}\Big|_{\theta=\theta_0}, \tag{A.12}$$

where

$$\frac{\partial S_n(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = -2n^{-1} \sum_{i=1}^{n} [(Y_i - \hat{m}(X_i))^2 - \sigma^2_{\theta_0}(X_i)] \frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_0}. \tag{A.13}$$

In order to prove that $\hat{m}(X_i)$ can be replaced by $m(X_i)$ in the above expression, we can follow the same type of arguments as presented in the proof of Lemma A.1, the main difference being that the weights are now all equal to n^{-1}. It can be shown in this way that the left hand side of (A.13) equals

$$\frac{\partial S_n(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = -2n^{-1} \sum_{i=1}^{n} [(Y_i - m(X_i))^2 - \sigma^2_{\theta_0}(X_i)] \frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_0} + o_P(n^{-1/2}).$$

Finally, we obtain from the continuity of $\frac{\partial \sigma^2_\theta}{\partial \theta \partial \theta'}$ and Lemma A.2,

$$\begin{aligned}
&\frac{\partial^2 S_n(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_{1n}} \\
&= \frac{2}{n} \sum_{i=1}^{n} \frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_{1n}} \left(\frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_{1n}}\right)' - \frac{2}{n} \sum_{i=1}^{n} [(Y_i - \hat{m}(X_i))^2 - \sigma^2_{\theta_{1n}}(X_i)] \frac{\partial^2 \sigma^2_\theta(X_i)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_{1n}} \\
&= \frac{2}{n} \sum_{i=1}^{n} \frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_0} \left(\frac{\partial \sigma^2_\theta(X_i)}{\partial \theta}\Big|_{\theta=\theta_0}\right)' - \frac{2}{n} \sum_{i=1}^{n} [(Y_i - m(X_i))^2 - \sigma^2_{\theta_0}(X_i)] \frac{\partial^2 \sigma^2_\theta(X_i)}{\partial \theta \partial \theta'}\Big|_{\theta=\theta_0} \\
&\quad + o_P(1) \\
&= 2\Omega + o_P(1),
\end{aligned}$$

and the assertion of Lemma A.3 follows. □

Proof of Theorem 3.1. From the proof of Theorem 1 in Akritas and Van Keilegom (2001) (hereafter called AVK) it follows that

$$\begin{aligned}
\hat{F}_\varepsilon(y) - F_\varepsilon(y) &= n^{-1} \sum_{i=1}^{n} I(\varepsilon_i \le y) - F_\varepsilon(y) \\
&\quad + f_\varepsilon(y) \int \sigma^{-1}(x)[y\{\hat{\sigma}(x) - \sigma(x)\} + \hat{m}(x) - m(x)]\, dF_X(x) + R_{n1}(y),
\end{aligned} \tag{A.14}$$

where

$$\sup_{y\in I\!R} |R_{n1}(y)| = o_P(n^{-1/2}).$$

Note that AVK assume that the regression function m and the variance function σ^2 are L-functionals that depend on a certain score function, say J. It is easy to show that the representation (A.14) can be extended to the choice $J \equiv 1$, which leads to the conditional mean and variance that we consider in this paper (it suffices to replace Propositions 3–5 in AVK by their analogues for the estimators of the conditional mean and variance).
We will now construct a similar representation for the differences $\hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y)$. We will do this by showing that Theorem 1 in AVK can be adapted to the case where $\hat{\sigma}(x)$ is replaced by $\hat{\sigma}_0(x)$. It can be easily seen that Propositions 3–5 in AVK continue to hold true when $\hat{\sigma}$ is replaced by $\hat{\sigma}_0$ (use assumption (A2)(iv) and the fact that $\hat{\theta} - \theta_0 = O_P(n^{-1/2})$).
We can now follow exactly the same derivation as for the above representation, and find in this way that

$$\begin{aligned} \hat{F}_{\varepsilon 0}(y) - F_\varepsilon(y) &= n^{-1}\sum_{i=1}^n I(\varepsilon_i \le y) - F_\varepsilon(y) \\ &+ f_\varepsilon(y)\int \sigma^{-1}(x)[y\{\hat{\sigma}_0(x) - \sigma(x)\} + \hat{m}(x) - m(x)]\, dF_X(x) + R_{n2}(y), \end{aligned} \tag{A.15}$$

where

$$\sup_{y\in I\!R} |R_{n2}(y)| = o_P(n^{-1/2}).$$

Hence, under the null hypothesis we obtain

$$\begin{aligned} \hat{F}_\varepsilon(y) - \hat{F}_{\varepsilon 0}(y) &= y f_\varepsilon(y)\int \sigma^{-1}(x)[\hat{\sigma}(x) - \hat{\sigma}_0(x)]\, dF_X(x) + o_P(n^{-1/2}) \\ &= \frac{y f_\varepsilon(y)}{2}\int \sigma^{-2}(x)[\hat{\sigma}^2(x) - \hat{\sigma}_0^2(x)]\, dF_X(x) + o_P(n^{-1/2}) \end{aligned} \tag{A.16}$$

uniformly in y. On the other hand it follows from Lemma A.1 that

$$\begin{aligned} &\int \sigma^{-2}(x)[\hat{\sigma}^2(x) - \sigma^2(x)]\, dF_X(x) \\ &\quad = n^{-1}\sum_{i=1}^n \sigma^{-2}(X_i)[\{Y_i - m(X_i)\}^2 - \sigma^2(X_i)] + o_P(n^{-1/2}), \end{aligned} \tag{A.17}$$

whereas Lemma A.3 and a Taylor expansion yield

$$\begin{aligned} &\int \sigma^{-2}(x)[\hat{\sigma}_0^2(x) - \sigma^2(x)]\, dF_X(x) \\ &= n^{-1}\sum_{i=1}^n \int \sigma^{-2}(x) K_h(x - X_i)\{\sigma^2_{\hat{\theta}}(x) - \sigma^2_{\theta_0}(x)\} dx + o_P(n^{-1/2}) \end{aligned}$$

$$= n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)\{\sigma^2_{\hat\theta}(X_i)-\sigma^2_{\theta_0}(X_i)\}+o_P(n^{-1/2})$$

$$= n^{-1}\sum_{i=1}^{n}\sigma^{-2}(X_i)\left(\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\theta_0}\right)'\Omega^{-1}\,n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\theta_0}$$
$$+o_P(n^{-1/2})$$

$$= \int\sigma^{-2}(x)\left(\frac{\partial\sigma^2_\theta(x)}{\partial\theta}\Big|_{\theta=\theta_0}\right)'dF_X(x)\,\Omega^{-1}\,n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\theta_0}$$
$$+o_P(n^{-1/2}). \tag{A.18}$$

The assertion of Theorem 3.1 now follows from (A.16) - (A.18). □

Proof of Theorem 3.4. In a similar way as in the proof of Lemmas A.1 and A.2 in Van Keilegom, González-Manteiga and Sánchez-Sellero (2005), it can be shown that under the alternative hypothesis H_1,

$$\begin{aligned}\hat\theta-\theta_0 &= \Omega^{-1}\,n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2_{\tilde\theta_{0n}}(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\tilde\theta_{0n}}\\ &\quad+\Omega^{-1}n^{-1/2}\int r(x)\frac{\partial\sigma^2_\theta(x)}{\partial\theta}\Big|_{\theta=\theta_0}dF_X(x)+o_P(n^{-1/2}),\end{aligned}$$

where

$$\begin{aligned}\tilde\theta_{0n} &= \operatorname*{argmin}_{\theta\in\Theta}\,E[\{(Y-m(X))^2-\sigma^2_\theta(X)\}^2]\\ &= \operatorname*{argmin}_{\theta\in\Theta}\,E[\{\sigma^2(X)-\sigma^2_\theta(X)\}^2].\end{aligned}$$

The proof parallels that of Theorem 3.1, except for the term considered in (A.18). For this expression we obtain under local alternatives

$$\int\sigma^{-2}(x)[\hat\sigma^2_0(x)-\sigma^2(x)]dF_X(x)$$
$$= \int\sigma^{-2}_{\theta_0}(x)\Big(\frac{\partial\sigma^2_\theta(x)}{\partial\theta}\Big|_{\theta=\theta_0}\Big)'dF_X(x)\Omega^{-1}\,n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2_{\tilde\theta_{0n}}(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}\Big|_{\theta=\tilde\theta_{0n}}$$
$$+\int\sigma^{-2}_{\theta_0}(x)\Big(\frac{\partial\sigma^2_\theta(x)}{\partial\theta}\Big|_{\theta=\theta_0}\Big)'dF_X(x)\Omega^{-1}n^{-1/2}\int r(x)\frac{\partial\sigma^2_\theta(x)}{\partial\theta}\Big|_{\theta=\theta_0}dF_X(x)$$
$$-n^{-1/2}\int\sigma^{-2}_{\theta_0}(x)r(x)dF_X(x)+o_P(n^{-1/2}).$$

It is easy to show that the asymptotic distribution of the random variable

$$n^{-1}\sum_{i=1}^{n}[\{Y_i-m(X_i)\}^2-\sigma^2_{\tilde\theta_{0n}}(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}|_{\theta=\tilde\theta_{0n}}$$

under H_1 equals the asymptotic distribution of

$$n^{-1}\sum_{i=1}^{n}[\{Y_i - m(X_i)\}^2 - \sigma^2_{\theta_0}(X_i)]\frac{\partial\sigma^2_\theta(X_i)}{\partial\theta}|_{\theta=\theta_0}$$

under H_0. The rest of the proof is similar as in the proofs of Theorem 3.1 and Corollaries 3.2 and 3.3. Hence, the assertion of Theorem 3.4 follows. □

References

Akritas, M.G. and Van Keilegom, I. (2001). Nonparametric estimation of the residual distribution. *Scand. J. Statist.*, **28**, 549–568.

Atkinson, A. C. (1985). *Plots, transformations and regression : an introduction to graphical methods of diagnostic.* Oxford : Clarendon Pr.

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

Carroll, R. and Ruppert, D. (1988). *Transformation and weighting in regression.* Monographs on Statistics and Applied Probability. New York, Chapman and Hall.

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

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

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

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

Eubank, R.L. and Thomas, W. (1993). Detecting Heteroscedasticity in Nonparametric Regression. *J. Roy. Statist. Soc. Ser. B*, **55**, 145–155.

Fan, Y.,and Linton, O. (2003). Some higher-order theory for a consistent non-parametric model specification test. *J. Stat. Plann. Inference* **109**, 125-154.

Feller, W. (1966). *An introduction to probability theory and its applications. Vol II.* Wiley Series in Probability and Mathematical Statistics. New York.

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

Hall, P. and Hart, J.D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.*, **85**, 1039–1049.

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

Jennrich, R.I. (1969). Asymptotic properties of non-linear least-squares estimators. *Ann. Math. Statist.*, **40**, 633–643.

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

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

Müller, H.G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.*, **15**, 610–625.

Nadaraya, E.A. (1964). On estimating regression. *Theory Probab. Appl.*, **15**, 134–137.

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

van der Vaart, A.W. (1998). *Asymptotic statistics*. Cambridge Univ. Press. Cambridge.

Van Keilegom, I. and Veraverbeke, N. (2002). Density and hazard estimation in censored regression models. *Bernoulli*, **8**, 607–625.

Van Keilegom , I., González-Manteiga, W. and Sánchez-Sellero, C. (2005). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Discussion Paper 0405, Institute of Statistics, Université catholique de Louvain* (http://www.stat.ucl.ac.be/ISpub/ISdp.html).

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā, Ser. A*, **26**, 359–372.

Zhu, L., Fujikoshi, Y. and Naito, K. (2001). Heteroscedasticity checks for regression models. *Science in China (Series A)*, **44**, 1236–1252.