

# Testing for a constant coefficient of variation 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

Gabriele Wieczorek

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum, Germany

e-mail: gabriele.wieczorek@rub.de

November 19, 2007

## Abstract

In this paper we propose a new test for the hypothesis of a constant coefficient of variation in the common nonparametric regression model. The test is based on an estimate of the  $L^2$ -distance between the square of the regression function and variance function. We prove asymptotic normality of a standardized estimate of this distance under the null hypothesis and fixed alternatives and the finite sample properties of a corresponding bootstrap test are investigated by means of a simulation study. The results are applicable to stationary processes with the common mixing conditions and are used to construct tests for ARCH assumptions in financial time series.

Keywords and Phrases: stationary processes, nonparametric regression, constant coefficient of variation, multiplicative error structure, generalized nonparametric regression models.

## 1 Introduction

We consider the common nonparametric regression model

$$(1.1) \quad Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, 2, \dots, n,$$

where  $m$  denotes the regression function and  $\sigma^2$  the variance function and the random variables  $\varepsilon_i$  satisfy  $E[\varepsilon_i|X_i = x] = 0$  and  $E[\varepsilon_i^2|X_i = x] = 1$ . In many applications the variance can be assumed proportional to the squared mean which corresponds to the assumption of a constant coefficient of variation. Typical examples include models obtained by the logarithmic transformation from regression models with a multiplicative error structure [see Eagleson and Müller (1997)] or ARCH-type models [see Engle (1982)]. Several authors have discussed the problem of estimating and testing the regression function under the restriction that  $m$  and  $\sigma$  are proportional - see e.g. McCullagh and Nelder (1989), who considered generalized linear models, Carroll and Ruppert (1988), who considered a constant coefficient of variation with a parametric model, and Eagleson and Müller (1997), who investigated the common nonparametric regression model under the restriction that  $m = c\sigma$  for some constant  $c$ .

In the present paper we will develop a formal test for the hypothesis of a constant coefficient of variation in the nonparametric regression model (1.1), that is

$$(1.2) \quad H_0 : m(x) = c\sigma(x)$$

for some positive (but unknown constant)  $c$ . Besides the fact that this test can be used to check the assumptions for a statistical inference in a nonparametric regression model with a constant coefficient of variation, it can also be used as an indicator of a multiplicative error structure (if it is applied to the squares of the data) and an exponentially distributed response  $Y$  where  $E[Y|X = x] = \sqrt{\text{Var}[Y|X = x]} = m(x)$ . In Section 2 we introduce the test statistic and indicate possible applications. Section 3 contains our main results in the case of an i.i.d. sample  $\{X_i, Y_i\}_{i=1}^n$ . We prove asymptotic normality of a standardized version of the test statistic under the null hypothesis, local and fixed alternatives. In Section 4 we extend these results in the case of stationary time series with the common mixing properties and discuss an application to test for an ARCH(1) model. The finite sample properties of a bootstrap version of the new test are investigated in Section 5 and some of the technical details for the proofs of our main results are presented in the Appendix in Section 6.

## 2 Testing for a constant coefficient of variation in non-parametric regression

Numerous authors have considered testing various hypotheses regarding the mean and the variance function in the nonparametric regression model (1.1) [see e.g. Dette and Munk (2003) and the

references in this paper]. These hypotheses include parametric and semi parametric assumptions regarding the mean and variance function, but much less effort has been spent in investigating the relation between mean and variance in the nonparametric regression model (1.1). In the present paper we investigate the hypothesis (1.2) of a constant coefficient of variation using an estimate of the  $L^2$ -distance between the variance and squared regression function. Typical examples include multiplicative models of the form

$$Y_t = m(X_t)\eta_t$$

which can be written in the form (1.1) with  $\sigma(\cdot) = \sqrt{\text{Var}(\eta_t)}m(\cdot)$  and  $\varepsilon_t = (\eta_t - 1)/\sqrt{\text{Var}(\eta_t)}$ . Other examples include nonparametric ARCH models  $X_t = \sqrt{m(X_{t-1}^2)}\eta_t$ , for which the squared process corresponds to a multiplicative times series model.

To be precise let  $\{X_i, Y_i\}_{i=1}^n$  denote a bivariate sample of observations from the nonparametric regression model (1.1) with the same distribution and let  $\hat{m}$  and  $\hat{\sigma}^2$  denote two nonparametric estimates of the regression and variance function, respectively, which will be specified in the following section. For any positive  $c$  we define the statistic  $T_n(c)$  as

$$(2.1) \quad T_n(c) = \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \{c^2 Y_i^2 - (c^2 + 1)\hat{m}^2(X_i)\} w(X_i) \\ \times \{c^2 Y_j^2 - (c^2 + 1)\hat{m}^2(X_j)\} w(X_j),$$

where  $w$  denotes a weight function,  $K_g(\cdot) = \frac{1}{g}K(\cdot/g)$ ,  $K(\cdot)$  denotes a kernel and  $g$  is a bandwidth converging to 0 with increasing sample size. Note that the statistic of the form (2.1) has been considered before by Zheng (1996) for testing the parametric form of the regression function, by Dette (2002) for testing homoscedasticity, by Dette and von Lieres und Wilkau (2003) and Gozalo and Linton (2000) for testing additivity in a nonparametric regression model (1.1) with a multivariate predictor. If the estimate  $\hat{m}$  is consistent it is intuitively clear that for a large sample size

$$(2.2) \quad E[T_n(c)] \approx E[K_g(X_1 - X_2) \{c^2 \sigma^2(X_1)\varepsilon_1^2 - 2c^2 m(X_1)\sigma(X_1)\varepsilon_1 - m^2(X_1)\} \\ \times \{c^2 \sigma^2(X_2)\varepsilon_2^2 - 2c^2 m(X_2)\sigma(X_2)\varepsilon_2 - m^2(X_2)\}] \\ \approx E[f(X_i) \{c^2 \sigma^2(X_i) - m^2(X_i)\}^2 w^2(X_i)] \\ = E[\Delta_c^2(X_i) f(X_i) w^2(X_i)],$$

where  $f$  denotes the density of  $X$  and

$$(2.3) \quad \Delta_c(x) = m^2(x) - c^2 \sigma^2(x).$$

Note that  $E[\Delta_c^2(X_i)f(X_i)w^2(X_i)] = 0$  if and only if the null hypothesis (1.2) is satisfied. There exist a few cases, where the constant  $c$  in the statistic  $T_n(c)$  is known. For example in ARCH(1) models with standard normal distributed innovations  $\eta_t$  we have  $X_t^2 = a_0 + a_1X_{t-1}^2 + (a_0 + a_1X_{t-1}^2)(\eta_t^2 - 1)$ , which gives  $c = 1/\sqrt{2}$ . However, in most cases of practical interest the constant  $c$  has to be estimated from the data. For this purpose we consider the least squares problem

$$(2.4) \quad \hat{c}^2 = \arg \min_{c \in \mathbb{R}^{>0}} \sum_{i=1}^n (m^2(X_i) - c^2\sigma^2(X_i))^2 w(X_i) = \frac{\sum_{i=1}^n m^2(X_i)\sigma^2(X_i)w(X_i)}{\sum_{i=1}^n \sigma^4(X_i)w(X_i)}$$

and estimate the unknown quantities on the right hand side. We define the residuals

$$(2.5) \quad \hat{r}(X_i) = Y_i - \hat{m}(X_i), \quad (i = 1, \dots, n)$$

and the estimate

$$(2.6) \quad \hat{c}^2 = \frac{(1/n) \sum_{i=1}^n \hat{m}^2(X_i)\hat{r}^2(X_i)w(X_i)}{(1/n) \sum_{i=1}^n (\hat{\sigma}^2(X_i))^2 w(X_i)}.$$

Note that the squared residuals  $\hat{r}^2(\cdot)$  are used for estimating the variance function in the numerator of  $\hat{c}^2$  in order to avoid an additional bias caused by the use of the variance estimator  $\hat{\sigma}^2(\cdot)$  [see the proof of Theorem 3.2 in the Appendix].

It is intuitively clear that the expression  $\hat{c}^2$  estimates

$$(2.7) \quad c_0^2 = \frac{E[m^2(X)\sigma^2(X)w(X)]}{E[\sigma^4(X)w(X)]},$$

which coincides with the constant  $c^2$  if the null hypothesis (1.2) is satisfied and corresponds to the best  $L^2$ -approximation of  $m^2$  by functions of the form  $c^2\sigma^2$ , otherwise. Consequently the hypothesis of a constant coefficient of variation will be rejected for large values of the statistic  $T_n(\hat{c})$ .

In the following sections we specify the asymptotic properties of the statistics  $T_n(c)$ ,  $\hat{c}^2$  and  $T_n(\hat{c})$  if the local linear estimate [see Fan and Gijbels (1996)] is used for estimating the mean and variance function.

### 3 Main results

In order to state our main results we have to specify nonparametric estimates of the regression and variance function and several assumptions for the model (1.1). We begin with the definition of the estimates. For the regression function we use the local linear estimate [see Fan and Gijbels (1996)]

$$(3.1) \quad \hat{m}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i)s_{n,1}(x)] Y_i}{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i)s_{n,1}(x)]}$$

where  $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$ ,  $K(\cdot)$  is a kernel,  $h$  denotes a further bandwidth and

$$(3.2) \quad s_{n,l}(x) = \sum_{i=1}^n K_h(X_i - x)(x - X_i)^l \quad l = 1, 2.$$

Similarly, the estimate of the variance function is obtained by replacing the observations  $Y_i$  by the squared residuals  $\hat{r}^2(X_i)$  defined in (2.5) and is given by

$$(3.3) \quad \hat{\sigma}^2(x) = \frac{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i)s_{n,1}(x)] \hat{r}^2(X_i)}{\sum_{i=1}^n K_h(X_i - x) [s_{n,2}(x) - (x - X_i)s_{n,1}(x)]}.$$

For the sake of transparency we first assume that  $\{X_i, Y_i\}_{i=1}^n$  is a sample of independent identically distributed observations. A corresponding result in the time series context is given in the following section. Moreover, the same bandwidths are assumed for the calculation of the estimates of the regression and variance function for the sake of simple notation. The treatment of different bandwidths in these estimates does not cause additional difficulties (and in the simulation study presented in Section 5 we used in fact different bandwidths). Throughout this section we assume that the following assumptions are satisfied

- (A1) The density  $f$  is twice continuously differentiable on compact sets.
- (A2) The regression function  $m$  is four times continuously differentiable on compact sets.
- (A3) The variance function  $\sigma^2$  is positive and twice continuously differentiable on compact sets.
- (A4) The weight function  $w$  is twice continuously differentiable and has compact support contained in  $\{x|f(x) > 0\}$ .
- (A5) The kernel  $K$  is of order 2, and satisfies a Lipschitz condition.
- (A6) If  $n \rightarrow \infty$  the bandwidth  $g$  and  $h$  satisfy

$$(3.4) \quad h \sim n^{-1/5}, \quad g = o(h^2), \quad ng \rightarrow \infty.$$

- (A7) The function  $m_k(x) = E[\varepsilon^k|X = x]$  is continuous for  $k = 3, 4$  and for  $1 \leq k \leq 8$  uniformly bounded, that is

$$(3.5) \quad E[\varepsilon_t^k|X_t = x] \leq C < \infty, \quad k \leq 8.$$

- (A8) The regression and variance function satisfy

$$E[m(X)]^k < \infty \text{ for } k = 2, 4, \text{ and } E[\sigma^2(X)]^k < \infty \text{ for } k = 1, 2.$$

Our first result specifies the asymptotic distribution of the statistic  $T_n(c)$ , where the constant  $c$  in the hypothesis (1.2) is known. Roughly speaking the statistic  $T_n(c)$  is asymptotically normal distributed with different rates of convergence under the null hypothesis and alternative. The proof is complicated and therefore deferred to the Appendix.

**Theorem 3.1.** *Assume that the assumptions (A1) - (A7) are satisfied.*

(a) *Under the null hypothesis (1.2) we have*

$$(3.6) \quad n\sqrt{g} T_n(c) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_0^2),$$

where the asymptotic variance is given by

$$(3.7) \quad \mu_0^2 = 2 E[\{-1 + 4c^2 + 4cm_3(X) + m_4(X)\}^2 m^8(X) f(X) w^4(X)] \int K^2(u) du.$$

(b) *Under a fixed alternative  $H_1 : m \neq c\sigma$  we have*

$$(3.8) \quad \sqrt{n} (T_n(c) - E[T_n(c)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_1^2(c)),$$

where

$$(3.9) \quad E[T_n(c)] = E[\Delta_c^2(X) f(X) w^2(X)] + h^2 B(c) + o(h^2)$$

with  $\Delta_c$  defined in (2.3),  $\kappa_2 = \int u^2 K(u) du$  and

$$(3.10) \quad B(c) = 2(c^2 + 1) \kappa_2 E[\Delta_c(X) m(X) m''(X) f(X) w^2(X)].$$

The asymptotic variance is given by

$$\begin{aligned} \mu_1^2(c) &= 4 \text{Var}(\Delta_c^2(X) f(X) w^2(X)) + 16 E[\Delta_c^2(X) m^2(X) \sigma^2(X) f^2(X) w^4(X)] \\ &\quad + 4c^4 E[\Delta_c^2(X) \sigma^4(X) f^2(X) \{m_4(X) - 1\} w^4(X)] \\ &\quad - 16c^2 E[\Delta_c^2(X) m(X) \sigma^3(X) f^2(X) m_3(X) w^4(X)]. \end{aligned}$$

In most applications the value  $c$  in the hypothesis (1.2) is not known and has to be estimated from the data. The following results specify the asymptotic properties of the estimate  $\hat{c}^2$  defined in (2.6) and the test statistic  $T_n(\hat{c})$ .

**Theorem 3.2.** *If the assumptions (A1) - (A8) are satisfied, then*

$$(3.11) \quad \begin{aligned} \hat{c}^2 - E[\hat{c}^2] &= \frac{1}{n} \sum_{i=1}^n \left\{ \tau_1 \left( m^2(X_i) \sigma^2(X_i) w(X_i) \varepsilon_i^2 - E[m^2(X) \sigma^2(X) w(X)] \right) \right. \\ &\quad + 2\tau_1 m(X_i) \sigma^3(X_i) w(X_i) \varepsilon_i - \tau_2 \left( \sigma^4(X_i) w(X_i) - E[\sigma^4(X) w(X)] \right) \\ &\quad \left. - 2\tau_2 \sigma^4(X_i) w(X_i) \{ \varepsilon_i^2 - 1 \} \right\} + o_p \left( \frac{1}{\sqrt{n}} \right). \end{aligned}$$

Moreover,

$$(3.12) \quad \sqrt{n}(\hat{c}^2 - E[\hat{c}^2]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nu^2),$$

where

$$(3.13) \quad E[\hat{c}^2] = c_0^2 + h^2\Gamma + o(h^2)$$

and the constants  $\Gamma, \tau_1, \tau_2$  and  $\nu^2$  are given by

$$\begin{aligned} \Gamma &= \kappa_2 E[\sigma^2(X)\{\tau_1 m(X)m''(X) - \tau_2(\sigma^2(X))''\}w(X)], \\ \tau_1 &= \frac{1}{E[\sigma^4(X)w(X)]}, \\ \tau_2 &= \frac{E[m^2(X)\sigma^2(X)w(X)]}{E^2[\sigma^4(X)w(X)]}, \\ \nu^2 &= \tau_1^2 \text{Var}(m^2(X)\sigma^2(X)w(X)) + 4\tau_1^2 E[m^2(X)\sigma^6(X)w^2(X)] \\ &\quad + 4\tau_1^2 E[m^3(X)\sigma^5(X)m_3(X)w^2(X)] + \tau_2^2 \text{Var}(\sigma^4(X)w(X)) \\ &\quad + 4\tau_2^2 E[\sigma^8(X)\{m_4(X) - 1\}w^2(X)] - 2\tau_1\tau_2 \text{Cov}(m^2(X)\sigma^2(X)w(X), \sigma^4(X)w(X)) \\ &\quad - 4\tau_1\tau_2 E[m^2(X)\sigma^6(X)\{m_4(X) - 1\}w^2(X)] - 4\tau_1\tau_2 E[m(X)\sigma^7(X)m_3(X)w^2(X)]. \end{aligned}$$

We are now in a position to investigate weak convergence of the statistic  $T_n(\hat{c})$ , where the estimate  $\hat{c}^2$  is defined in (2.4). We begin with the asymptotic distribution under the null hypothesis (1.2). Interestingly, in this case the estimation of the scaling factor  $c$  has no influence on the asymptotic properties of the test statistic.

**Theorem 3.3.** *Assume that the assumptions (A1) - (A8) are satisfied. Under the null hypothesis (1.2) we have*

$$n\sqrt{g} T_n(\hat{c}) = n\sqrt{g} T_n(c) + o_p(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_0^2),$$

where the constant  $\mu_0^2$  is defined in (3.7).

Our final result in this section refers to the asymptotic properties of the statistic  $T_n(\hat{c})$  under the alternative. In this case there appears an additional term in the bias and variance of the test statistic, which is caused by the estimation of the scaling factor  $c$ . Recall that the constant  $c_0^2$  corresponds to the best  $L^2$ -approximation of  $m^2$  by functions of the form  $c^2\sigma^2$ .

**Theorem 3.4.** *Assume that the assumptions (A1) - (A8) are satisfied. Under a fixed alternative*

$$\varrho = E[\Delta_{c_0}(X)\sigma^2(X)f(X)w^2(X)] > 0$$

we have

$$\sqrt{n}(T_n(\hat{c}) - E[T_n(\hat{c})]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \omega_1^2),$$

where

$$E[T_n(\hat{c})] = E[\Delta_{c_0}^2(X)f(X)w^2(X)] + h^2(B(c_0) - 2\varrho\Gamma) + o(h^2),$$

and  $B(c_0)$  is a term in the bias of the statistic  $T_n(c_0)$ . The asymptotic variance  $\omega_1^2$  is given by

$$\omega_1^2 = \mu_1^2(c_0) + 4\varrho^2\nu^2 - 4\varrho v^2(c_0),$$

where  $\mu_1^2(c_0)$  is defined in Theorem 3.1(b),  $\nu^2$  corresponds to the asymptotic variance of  $\hat{c}^2$  in Theorem 3.2 and

$$\begin{aligned} v^2(c_0) &= 2\tau_1 E[\Delta_{c_0}(X)(m^2(X) - c_0^2\sigma^2(X)m_4(X))m^2(X)\sigma^2(X)f(X)w^3(X)] \\ &\quad - 2\tau_1 E[\Delta_{c_0}^2(X)f(X)w^2(X)]E[m^2(X)\sigma^2(X)w(X)] \\ &\quad - 4c_0^2\tau_1 E[\Delta_{c_0}(X)m(X)\sigma^5(X)f(X)m_3(X)w^3(X)] \\ &\quad - 2\tau_2 \text{Cov}(\Delta_{c_0}^2(X)f(X)w^2(X), \sigma^4(X)w(X)) \\ &\quad + 4c_0^2\tau_2 E[\Delta_{c_0}(X)\sigma^6(X)f(X)\{m_4(X) - 1\}w^3(X)] \\ &\quad + 4\tau_1 E[\Delta_{c_0}(X)m^3(X)\sigma^3(X)f(X)m_3(X)w^3(X)] \\ &\quad + 8\tau_1 E[\Delta_{c_0}(X)m^2(X)\sigma^4(X)f(X)w^3(X)] \\ &\quad - 8\tau_2 E[\Delta_{c_0}(X)m(X)\sigma^5(X)f(X)m_3(X)w^3(X)]. \end{aligned}$$

**Remark 3.5.** The term  $v^2(c_0)$  corresponds to the asymptotic covariance between the statistic  $T_n(c_0)$  and the estimate  $\hat{c}^2$  of  $c_0^2$ .

## 4 Further discussion

### 4.1 Asymptotic results for absolutely regular processes

The general nonparametric framework includes time series models. Typical examples are multiplicative models  $Z_t = \sigma_t\eta_t$ , where  $\sigma_t$  is a positive function of the past  $\{Z_{t-i} : i \geq 1\}$  and possibly of the past volatility  $\{\sigma_{t-i} : i \geq 1\}$ . For instance, defining  $\sigma_t$  by  $\sqrt{\vartheta_0 + \vartheta_1 Z_{t-1}^2}$  for  $\vartheta_i \geq 0$  we achieve the linear ARCH(1) model. Therefore our test can also be used as a preliminary step to identify certain time series. For this purpose it is necessary to extend the asymptotic results under a more general setup which includes both time series data and i.i.d. observations as special cases. For this purpose we need the following assumptions for some fixed  $\varepsilon \in (0, 1/2)$  and  $\xi > 2$ .

(M1) The process  $(X_i, Y_i)$  is absolutely regular, i.e.

$$\beta(k) = \sup_{s \in \mathbb{Z}} E[\sup\{|P(A|\mathcal{F}_{-\infty}^s) - P(A)|A \in \mathcal{F}_{s+k}^\infty\}] \rightarrow 0, \quad \text{as } k \rightarrow 0,$$

where  $\mathcal{F}_s^t$  is the  $\sigma$ -algebra generated by  $\{(X_l, Y_l) : s \leq l \leq t\}$ . Further,

$$\sum_{j=1}^{\infty} j^2 \beta^{\frac{\varepsilon}{1+\varepsilon}}(j) < \infty.$$

(M2) The innovations  $\varepsilon_t$  in the model (1.1) satisfy

$$E[\varepsilon_t | X_t, \mathcal{F}_{-\infty}^{t-1}(X, Y)] = E[\varepsilon_t | X_t] = 0,$$

and

$$\text{Var}(Y_t | X_t, \mathcal{F}_{-\infty}^{t-1}(X, Y)) = \sigma^2(x).$$

Further,  $E|\varepsilon_t|^k < \infty$  to the order  $k \leq 48\xi(1 + \varepsilon)$ .

(M3) The regression function  $m(\cdot)$  satisfies

$$\begin{aligned} E|m(X)|^k &< \infty \quad \text{for } k \leq 4(1 + \varepsilon) \quad \text{and} \\ E|m''(X)|^k &< \infty \quad \text{for } k \leq 20\xi(1 + \varepsilon), \end{aligned}$$

whereas the variance function  $\sigma^2(\cdot)$  fulfills

$$E|\sigma^2(X)|^k < \infty \quad \text{for } k \leq 12\xi(1 + \varepsilon).$$

Note that assumption (M3) contains assumption (A8) which is therefore omitted. Under the assumptions (A1) - (A7) together with (M1) - (M3) the asymptotic results for  $T_n(c)$ ,  $\hat{c}^2$  and  $T_n(\hat{c})$  can be established for strictly stationary,  $\beta$ -mixing processes  $\{X_i, Y_i\}_{i \in \mathbb{Z}}$ . The proof of the following results is obtained from the proof of the statements presented in Section 3 for the independent case using similar arguments as given by Dette und Spreckelsen (2004), where the authors investigate the asymptotic distribution of goodness-of-fit tests of linearity for absolutely regular processes. For the sake of brevity the details are omitted and we refer the interested reader to the PhD thesis of Wieczorek (2007). Moreover, we only state the results for the statistic  $T_n(\hat{c})$ . Note that under the null hypothesis the asymptotic distribution of  $T_n(\hat{c})$  under mixing assumptions coincides with the distribution for the i.i.d. case.

**Theorem 4.1** *Assume that the assumptions (A1) - (A7) and (M1) - (M3) are satisfied. Under the null hypothesis, we have*

$$n\sqrt{g} T_n(\hat{c}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_0^2),$$

where  $\mu_0^2$  is the asymptotic variance of  $T_n(c)$  defined in (3.7).

Our final theoretical result states the asymptotic properties of the statistic  $T_n(\hat{c})$  under fixed alternatives. Note that in this case the variance of the limit distribution contains the variance of the limit distribution for the i.i.d. case as well as additional covariances. For a precise statement of the result we introduce the notation  $E^\otimes$ , which denotes the expectation with respect to the product measure.

**Theorem 4.2.** *If the assumptions (A1) - (A7) and (M1) - (M3) are satisfied, then under a fixed alternative  $\varrho > 0$  we have*

$$\sqrt{n} (T_n(\hat{c}) - E^\otimes[T_n(\hat{c})]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\omega}_1^2).$$

In particular,

$$|E[T_n(\hat{c})] - E^\otimes[T_n(\hat{c})]| = o\left(\frac{1}{\sqrt{n}}\right),$$

where the mean  $E^\otimes[T_n(\hat{c})]$  and the constant  $\varrho$  are given in Theorem 3.4. The asymptotic variance is given by

$$(4.1) \quad \tilde{\omega}_1^2 = \tilde{\mu}_1^2(c_0) + 4\varrho^2 \tilde{\nu}^2 - 4\varrho \tilde{\nu}^2(c_0),$$

where  $\tilde{\mu}_1^2(c_0)$  denotes the asymptotic variance of  $T_n(c_0)$  defined by

$$\begin{aligned} \tilde{\mu}_1^2(c_0) = & \mu_1^2(c_0) + 8 \sum_{t=1}^{\infty} \text{Cov}(\Delta_{c_0}(X_1)[\Delta_{c_0}(X_1, \varepsilon_1) + 2m(X_1)\sigma(X_1)\varepsilon_1]f(X_1)w^2(X_1), \\ & \Delta_{c_0}^2(X_{1+t})f(X_{1+t})w^2(X_{1+t})), \end{aligned}$$

$\mu_1^2(c_0)$  is defined in Theorem 3.1(b). The term  $\tilde{\nu}^2$  in (4.1) corresponds to the asymptotic variance of the estimate  $\hat{c}^2$  given by

$$\begin{aligned} \tilde{\nu}^2 = & \nu^2 + 2 \sum_{t=1}^{\infty} \text{Cov}\left(2\tau_1 m(X_1)\sigma^3(X_1)w(X_1)\varepsilon_1 - \tau_2\sigma^4(X_1)w(X_1)\{2\varepsilon_1^2 - 1\}\right. \\ & \left. + \tau_1 m^2(X_1)\sigma^2(X_1)w(X_1)\varepsilon_1^2, \tau_1 m^2(X_{1+t})\sigma^2(X_{1+t})w(X_{1+t}) - \tau_2\sigma^4(X_1)w(X_1)\right), \end{aligned}$$

where  $\nu^2$  is given in Theorem 3.2 and  $\tilde{v}^2(c_0)$  corresponds to the asymptotic covariance between  $T_n(c_0)$  and  $\hat{c}^2$  defined by

$$\begin{aligned} \tilde{v}^2(c_0) &= \nu^2(c_0) + 2 \sum_{t=1}^{\infty} \text{Cov}(\Delta_{c_0}(X_1)f(X_1) \{ \Delta_{c_0}(V_1) + 2m(X_1)\sigma(X_1)\varepsilon_1 \} w^2(X_1), \\ &\quad \{ \tau_1 m^2(X_{1+t})\sigma^2(X_{1+t}) - \tau_2 \sigma^4(X_{1+t}) \} w(X_{1+t})) \\ &\quad + 2 \sum_{t=1}^{\infty} \text{Cov}(\Delta_{c_0}^2(X_{1+t})f(X_{1+t})w^2(X_{1+t}), \\ &\quad \{ \tau_1 m^2(X_1)\sigma^2(X_1)\varepsilon_1^2 + 2\tau_1 m(X_1)\sigma^3(X_1)\varepsilon_1 - \tau_2 \sigma^4(X_1)(2\varepsilon_1^2 - 1) \} w(X_1)), \end{aligned}$$

and  $\nu^2(c_0)$  is defined in Theorem 3.4.

**Remark 4.3.** It is worthwhile to mention that in the case where the stationary process is absolutely regular with a geometric rate, i.e.  $\beta(j) = O(\rho^j)$  for some  $\rho \in (0, 1)$ , the asymptotic covariance of the test statistic given in Theorem 4.2 coincides with the asymptotic covariance given in Theorem 3.4 for the independent case, that is:

$$\tilde{\mu}_1^2(c_0) = \mu_1^2(c_0), \quad \tilde{v}^2 = \nu^2, \quad \tilde{v}^2(c_0) = \nu^2(c_0).$$

**Remark 4.4.** The moment assumption (M3) is quite restrictive and limits the applicability of the test to many interesting time series models such as ARCH or GARCH models. One possible way to circumvent assumption (M3) is the introduction of an additional weight function in the estimates. As a consequence a slight modification of the estimates can be arranged in our testing procedure eliminating assumption (M3). The details can be found in Wieczorek (2007), and only the modification is mentioned for the sake of brevity. We introduce in a first step an additional weight function  $w^*$ , satisfying

(A9)  $w^*$  is twice continuously differentiable and has compact support contained in  $\{x | w(x) > 0\}$ .

Next, we propose a modified estimate of the regression function given by  $\hat{m}^*(x) = \hat{a}$ , where

$$(4.2) \quad (\hat{a}, \hat{b}) = \arg \min_{a,b} \sum_{i=1}^n \{Y_i - a - b(X_i - x)\}^2 w(X_i) K\left(\frac{X_i - x}{h}\right)$$

is the local linear estimate (additionally weighted by  $w$ ) of the regression function and its derivative. Note that the modified local linear regression estimator  $\hat{m}^*$  differs from the local linear

estimate  $\hat{m}$  in (3.1) by the introduction of the weight function  $w$  in (4.2). Similarly, we propose  $(\sigma^2)^*(x) = \hat{\alpha}$  as the modified estimate of the variance function, where

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} \sum_{i=1}^n \{(\hat{r}^*)^2(X_i) - \alpha - \beta(X_i - x)\}^2 w^*(X_i) K\left(\frac{X_i - x}{h}\right)$$

is the local linear estimate (weighted by the second weight function  $w^*$ ) based on the nonparametric residuals  $\hat{r}^*(X_i)$  defined by

$$\hat{r}^*(X_i) = Y_i - \hat{m}^*(X_i).$$

Based on the modified estimates of the regression function and the variance function the new test statistic is defined by

$$(4.3) \quad T_n^*(c) = \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \{c^2 Y_i^2 - (c^2 + 1)(\hat{m}^*)^2(X_i)\} w^*(X_i) \\ \times \{c^2 Y_j^2 - (c^2 + 1)(\hat{m}^*)^2(X_j)\} w^*(X_j).$$

In addition, we consider the modified least squares problem

$$(\hat{c}^2)^* = \arg \min_{c \in \mathbb{R}^{>0}} \sum_{i=1}^n (m^2(X_i) - c^2 \sigma^2(X_i))^2 (w^*)^3(X_i).$$

Therefore, we define the estimate of  $c^2$  by

$$(\hat{c}^2)^* = \frac{(1/n) \sum_{i=1}^n (\hat{m}^*)^2(X_i) (\hat{r}^*)^2(X_i) (w^*)^3(X_i)}{(1/n) \sum_{i=1}^n ((\hat{\sigma}^2)^*(X_i))^2 (w^*)^3(X_i)}.$$

As an immediate consequence of the modified definitions the asymptotic results in Theorem 4.1 and 4.2 can also be established for the test statistic  $T_n^*(\hat{c}^*)$ . No additional assumptions are needed, in particular, the introduction of the weight functions in the estimators avoids the assumption (M3) about the boundedness of the moments of the regression and variance function [see Wiecek (2007) for details]. This modification makes the test applicable to financial time series, as demonstrated in the following section.

## 4.2 Example: Application to financial time series

The hypothesis of the proportionality of the regression function  $m$  and the volatility function  $\sigma$  can also be used to test for a multiplicative model structure. In particular, the proposed test can be viewed as a preliminary step in time series analysis before applying other procedures such as specific testing procedures for ARCH or GARCH models. One important criterion in order

to establish all asymptotic results in such a context is assumption (M3). There the existence of bounds for the absolute moments of the regression function  $m$ , its second derivative  $m''$  and the variance function  $\sigma^2$  is required. But often financial time series do not satisfy this assumption. For instance, consider the linear ARCH(1) model

$$Z_t = \sqrt{\vartheta_0 + \vartheta_1 Z_{t-1}^2} \eta_t$$

for some constants  $\vartheta_0, \vartheta_1 \geq 0, \vartheta_1 < 1$ , where  $\eta_t$  has mean 0 and variance 1 and is independent of  $Z_{t-1}$  for all  $t$ . The squared ARCH(1) process can be written as

$$(4.4) \quad Z_t^2 = (\vartheta_0 + \vartheta_1 Z_{t-1}^2) + (\vartheta_0 + \vartheta_1 Z_{t-1}^2) \varepsilon_t,$$

where  $\varepsilon_t = \eta_t^2 - 1$ . Clearly, model (4.4) can be identified as a particular case of the general nonparametric regression model (1.1) by taking  $Y_t = Z_t^2$ ,  $X_t = Z_{t-1}^2$ ,  $m(X_t) = \vartheta_0 + \vartheta_1 X_t$  and  $\sigma(X_t) = c^{-1}(\vartheta_0 + \vartheta_1 X_t)$ . The scaling factor  $c$  is given by  $c^2 = (E[\eta^4] - 1)^{-1}$  and depends only on the error distribution.

For the ARCH(1) process the assumption (M3) can therefore be formulated in terms of the boundedness of absolute moments of  $Z_t$ . So it is important to know whether the stationary solution  $Z_t$  has moments of higher orders to apply the test. For stationary ARCH(p) processes with a symmetric error distribution, a necessary and sufficient condition for the existence of such moments has been given by Milhøj (1985). In particular, let  $m > 0$ , then the  $m$ th moment of an ARCH(1) model exists if and only if  $E[\vartheta_1 \eta_0^2]^m < 1$ . As an immediate consequence, one sees that in many cases ARCH processes do not have finite moments of higher orders.

In such cases we refer to Remark 4.4. In order to circumvent the assumption of existing high-order moments of  $Z_t$  we apply the (slightly) modified testing procedure. In particular, the identification of the regression function  $m$  and the variance function  $\sigma^2$  provides the assumptions (A2) and (A3) to be satisfied. Furthermore, from

$$E[\varepsilon^k | X = x] = c^k E[(\eta^2 - 1)^k | Z_{t-1}^2] = c^k E[(\eta^2 - 1)^k]$$

it follows that  $(\varepsilon_t)$  fulfills (A7) and (M2) if the innovations  $\eta_t$  satisfy certain moment conditions. If the assumptions (A1), (A4) - (A7), (A9) are satisfied and the ARCH(1) process  $(Z_t)$  fulfills the assumption (M1) the asymptotic normality under the null hypothesis of the corresponding test statistic  $T_n^*(\hat{c}^*)$  can be established, that is

$$n\sqrt{g} (T_n^*(\hat{c}^*) - E[T_n^*(\hat{c}^*)]) \xrightarrow{D} \mathcal{N}(0, (\mu_0^2)^*),$$

where the asymptotic variance  $(\mu_0^2)^*$  is given by

$$(\mu_0^2)^* = 1152 \int K^2(u) du \sum_{k=0}^8 \binom{n}{k} \vartheta_0^{n-k} \vartheta_1^k E[Z^{2k} f(Z^2) (w^*)^4(Z^2)].$$

## 5 Finite sample properties

In order to study the finite sample properties of the new test we have conducted a small simulation study. Because it is well known that the approximation of the nominal level by the normal distribution provided by Theorem 3.3 is not very accurate for moderate sample sizes, we do not recommend to estimate the asymptotic variance and bias and to compare the standardized statistic with the quantiles of a normal distribution. In contrary, we propose to use resampling methods. As an example we have implemented a smooth bootstrap procedure to obtain the critical values. For this purpose we estimate the regression and variance function by the local linear estimates defined in (3.1) and (3.3), respectively, and consider the standardized residuals

$$(5.1) \quad \eta_i = \frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \quad i = 1, \dots, n,$$

which are normalized to have mean 0 and variance 1, that is

$$(5.2) \quad \hat{\varepsilon}_i = \frac{\eta_i - \bar{\eta}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\eta_i - \bar{\eta})^2}} \quad i = 1, \dots, n.$$

The bootstrap errors are then defined as

$$(5.3) \quad \varepsilon_i^* = \tilde{\varepsilon}_i^* + vN_i,$$

where  $\tilde{\varepsilon}_1^*, \dots, \tilde{\varepsilon}_n^*$  are drawn randomly with replacement from the empirical distribution of the standardized residuals  $\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_n$  and  $N_1, \dots, N_n$  are i.i.d standard normal distributed random variables independent of the sample  $\mathcal{Y}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  and  $v=v_n$  is a smoothing parameter converging to 0 with increasing sample size. In the next step bootstrap data is generated according to the model

$$(5.4) \quad Y_i^* = \hat{c} \hat{\sigma}(X_i) + \hat{\sigma}(X_i) \varepsilon_i^* \quad i = 1, \dots, n,$$

where  $\hat{c}$  is the least squares estimate (2.6) obtained from the data corresponding to the range  $[5\%, 95\%]$  of the predictors. The test statistic  $T_n^*$  is calculated from the bootstrap data  $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$ . If B bootstrap replications have been performed, the null hypothesis (1.2) is rejected if

$$(5.5) \quad T_n > T_n^{*(\lfloor B(1-\alpha) \rfloor)},$$

where  $T_n^{*(1)} < \dots < T_n^{*(B)}$  denote the order statistic of the bootstrap sample. For the size of the bootstrap replications we chose  $B = 100$ , while 1000 simulation runs are performed for the calculation of the empirical level of this test. The sample sizes are given by  $n = 50, 100, 200$  and the smoothing parameters in the test statistic and the bootstrap procedure are chosen by

$g = n^{-1/2}$  and  $v = 0.1$ , respectively. The bandwidths for the estimation of the variance and regression function are chosen separately by least squares cross validation.

Our first example considers the model

$$(5.6) \quad m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x),$$

where  $c = 0.5, 1, 1.5$ . The predictors  $X_1, \dots, X_n$  are independent identically distributed following a uniform distribution on the interval  $[0,1]$ , while the errors  $\varepsilon_1, \dots, \varepsilon_n$  have a standard normal distribution. The first part of Table 1 shows the approximation of the nominal level, which is rather accurate for sample sizes larger than  $n = 100$ . In a second step we study the power of the test and consider the models

$$(5.7) \quad m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x + \sqrt{x}),$$

$$(5.8) \quad m(x) = c(1 + 0.1x); \quad \sigma(x) = (1 + 0.1x + 2\sqrt{x}).$$

The corresponding results are depicted in the lower part of Table 1. For the model (5.7) we observe a moderate increase in power, which corresponds to intuition. Because the predictor varies in the interval  $[0,1]$ , the deviation from a multiplicative structure is extremely small for model (5.7). On the other hand, the alternative model (5.8) is detected with larger power, which is also reflected by rather high simulated rejection probabilities.

$n$		50				100				200			
	$c \setminus \alpha$	2.5 %	5 %	10 %	20 %	2.5 %	5 %	10 %	20 %	2.5 %	5 %	10 %	20 %
(5.6)	0.5	.028	.056	.104	.229	.035	.063	.108	.211	.032	.052	.099	.207
	1.0	.042	.051	.105	.210	.031	.049	.099	.194	.033	.054	.104	.204
	1.5	.051	.069	.123	.241	.041	.061	.109	.202	.038	.055	.096	.205
(5.7)	0.5	.043	.067	.144	.267	.051	.074	.148	.269	.122	.187	.281	.442
	1.0	.041	.069	.136	.264	.044	.094	.168	.281	.105	.144	.233	.368
	1.5	.064	.091	.141	.247	.077	.109	.166	.273	.114	.152	.208	.341
(5.8)	0.5	.049	.092	.167	.283	.066	.097	.213	.368	.097	.155	.261	.421
	1.0	.073	.122	.215	.362	.092	.156	.297	.464	.177	.266	.388	.554
	1.5	.063	.107	.203	.378	.123	.187	.281	.442	.233	.316	.424	.559

Table 1: *Simulated rejection probabilities of the bootstrap test (5.5), for three nonparametric regression models, where the first line corresponds to a multiplicative model.*

Our second example investigates the performance of the bootstrap test in the context of stationary time series. To this end we consider two models corresponding to the null hypothesis, that is

$$(5.9) \quad X_t = (1 + 0.1 X_{t-1}) + (1 + 0.1 X_{t-1})\varepsilon_t$$

$$(5.10) \quad X_t = \sin(1 + 0.5 X_{t-1}) + \sin(1 + 0.5 X_{t-1})\varepsilon_t$$

and two models corresponding to the alternatives of no multiplicative structure, i.e.

$$(5.11) \quad X_t = (1 + 0.1 X_{t-1}) + 0.5\sqrt{|X_{t-1}|} \varepsilon_t$$

$$(5.12) \quad X_t = \sin(1 + 0.5 X_{t-1}) + \cos(1 + 0.5 X_{t-1})\varepsilon_t$$

where the innovations are again independent standard normal distributed. The corresponding results are displayed in Table 2. We observe a reasonable approximation of the nominal level for the two models corresponding to the null hypothesis. On the other hand, the two alternatives in (5.11) and (5.12) are detected with reasonable power.

$n$	50				100				200			
$\alpha$	2.5 %	5 %	10 %	20 %	2.5 %	5 %	10 %	20 %	2.5 %	5 %	10 %	20 %
(5.9)	.029	.047	.097	.217	.023	.048	.089	.187	.024	.048	.097	.189
(5.10)	.038	.057	.109	.201	.035	.054	.092	.191	.036	.057	.109	.205
(5.11)	.053	.077	.161	.295	.074	.092	.182	.314	.113	.156	.237	.395
(5.12)	.084	.117	.189	.299	.097	.133	.212	.321	.129	.176	.289	.417

Table 2: Simulated rejection probabilities of the bootstrap test (5.5) for four nonparametric autoregressive time series models. The models (5.9) and (5.10) correspond to the null hypothesis of a multiplicative model, while models (5.11) and (5.12) represent the alternative.

## 6 Appendix: proofs

### 6.1 Proof of Theorem 3.1.

A straightforward calculation gives the decomposition

$$(6.1) \quad T_n(c) = (c^2 + 1)^2 T_{1n} - 2(c^2 + 1)\{2c^2 T_{2n} - T_{3n}(c)\} + T_{4n}(c) - 4c^2\{T_{5n}(c) - c^2 T_{6n}\},$$

with

$$\begin{aligned}
T_{1n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \delta_n(X_j) w(X_j), \\
T_{2n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) m(X_j) \sigma(X_j) w(X_j) \varepsilon_j, \\
T_{6n} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \sigma(X_i) w(X_i) \varepsilon_i m(X_j) \sigma(X_j) w(X_j) \varepsilon_j, \\
T_{3n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j), \\
T_{4n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \Delta_c(X_i, \varepsilon_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j), \\
T_{5n}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \Delta_c(X_i, \varepsilon_i) w(X_i) m(X_j) \sigma(X_j) w(X_j) \varepsilon_j,
\end{aligned}$$

where we have used the notation

$$(6.2) \quad \Delta_c(X_i, \varepsilon_i) = m^2(X_i) - c^2 \sigma^2(X_i) \varepsilon_i^2$$

$$(6.3) \quad \delta_n(X_i) = \hat{m}^2(X_i) - m^2(X_i).$$

At the end of the proof we will show that the terms  $T_{1n}$  and  $T_{2n}$  are asymptotically negligible under the null hypothesis and under fixed alternatives, that is

$$(6.4) \quad n\sqrt{g} T_{jn} \xrightarrow{p} 0, \quad j = 1, 2.$$

We now have to distinguish the case of the null hypothesis and alternative.

**Proof of Theorem 3.1(a).** Note that the statistic  $T_{3n}(c)$  reduces under the null hypothesis to

$$\begin{aligned}
(6.5) \quad T_{3n}(c) &\stackrel{H_0}{=} \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) m^2(X_j) w(X_j) \{1 - \varepsilon_j^2\}. \\
&\stackrel{H_0}{=} 2\tilde{T}_{3n}^{(1)} + \tilde{T}_{3n}^{(2)}
\end{aligned}$$

with

$$\begin{aligned}
\tilde{T}_{3n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \tilde{\delta}_n(X_i) w(X_i) m^2(X_j) w(X_j) \{1 - \varepsilon_j^2\}, \\
\tilde{T}_{3n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \tilde{\delta}_n^2(X_i) w(X_i) m^2(X_j) w(X_j) \{1 - \varepsilon_j^2\},
\end{aligned}$$

where we use the notation

$$(6.6) \quad \tilde{\delta}_n(x) = \hat{m}(x) - m(x).$$

An application of Lemma 2 in Yao and Tong (2000) and a straightforward calculation [similar to the proof of the estimate (6.4) given at the end of this section] yields

$$(6.7) \quad \tilde{T}_{3n}^{(1)} = O_p\left(\frac{1}{nh}\right), \quad \tilde{T}_{3n}^{(2)} = O_p\left(\frac{h}{n}\right).$$

Observing the conditions on the bandwidths we therefore have

$$(6.8) \quad T_{3n}(c) \stackrel{H_0}{=} o_p\left(\frac{1}{n\sqrt{g}}\right).$$

We note again that this estimate holds only under the null hypothesis.

We now obtain from (6.4) and (6.8) under the null hypothesis  $H_0 : m(\cdot) = c \sigma(\cdot)$  that

$$(6.9) \quad T_n(c) \stackrel{H_0}{=} T_{4n}(c) - 4c^2T_{5n}(c) + 4c^4T_{6n} + o_p\left(\frac{1}{n\sqrt{g}}\right),$$

and the assertion follows if the weak convergence can be established for the statistic

$$(6.10) \quad \tilde{T}_n(c) = T_{4n}(c) - 4c^2T_{5n}(c) + 4c^4T_{6n} = \frac{2}{n(n-1)} \sum_{i < j} h_n(V_i, V_j),$$

where  $V_i = (X_i, \varepsilon_i)$ ,

$$\begin{aligned} h_{ij} &= h_n(V_i, V_j) = K_g(X_i - X_j)\pi(V_i)w(X_i)\pi(V_j)w(X_j), \\ \pi(V_i) &= m^2(X_i)\{1 - \varepsilon_i^2 - 2c\varepsilon_i\}. \end{aligned}$$

For a proof of weak convergence of  $\tilde{T}_n(c)$  we note that the kernel  $h_n(\cdot, \cdot)$  is degenerate, i.e.

$$(6.11) \quad E[h_n(V_i, V_j)|V_i] = E[h_n(V_i, V_j)|V_j] = 0 \quad a.s.$$

and apply Theorem 2.1 in de Jong (1987). For this purpose we calculate the variance as

$$\begin{aligned} \sigma_n^2 &= \text{Var}(\tilde{T}_n(c)) = \frac{2^2}{n^2(n-1)^2} \sum_{i < j} E[h_{ij}^2] \\ &= \frac{2}{n(n-1)} E[K_g^2(X_i - X_j)m^4(X_i)w^2(X_i)m^4(X_j)E[a^2(\varepsilon_i)|X_i]E[a^2(\varepsilon_j)|X_j]w^2(X_j)] \\ &= \frac{\mu_0^2}{n(n-1)g} + o\left(\frac{1}{n^2g}\right), \end{aligned}$$

where we have used the notation

$$(6.12) \quad a(\varepsilon_i) = 1 - \varepsilon_i^2 - 2c\varepsilon_i$$

and  $\mu_0^2$  is defined in (3.7). Observing that by this calculation  $\sigma_{ij}^2 := E[h_{ij}^2] = \frac{1}{g} \frac{\mu_0^2}{2} \cdot (1 + o(1))$  we therefore have

$$\sigma_n^{-2} \max_{i=1}^n \left( \frac{1}{n^2(n-1)^2} \sum_{1 \leq j \leq n} \sigma_{ij}^2 \right) = O\left(\frac{1}{n}\right),$$

which proves the first assumption in de Jong's (1987) Theorem 2.1. In order to establish the second assumption we derive the decomposition [see de Jong (1987)]

$$(6.13) \quad E[\tilde{T}_n^4(c)] = G_I + 6G_{II} + 12G_{III} + 24G_{IV} + 6G_V,$$

where

$$\begin{aligned} G_I &= \frac{2^4}{n^4(n-1)^4} \sum_{i < j} E[h_{ij}^4], \\ G_{II} &= \frac{2^4}{n^4(n-1)^4} \sum_{i < j < k} E[h_{ij}^2 h_{ik}^2] + E[h_{ji}^2 h_{jk}^2] + E[h_{ki}^2 h_{kj}^2], \\ G_{III} &= \frac{2^4}{n^4(n-1)^4} \sum_{i < j < k} E[h_{ij}^2 h_{ki} h_{kj}] + E[h_{ik}^2 h_{ji} h_{jk}] + E[h_{kj}^2 h_{ij} h_{ik}], \\ G_{IV} &= \frac{2^4}{n^4(n-1)^4} \sum_{i < j < k < l} E[h_{ij} h_{ik} h_{lj} h_{lk}] + E[h_{ij} h_{il} h_{kj} h_{kl}] + E[h_{ik} h_{il} h_{jk} h_{jl}], \\ G_V &= \frac{2^4}{n^4(n-1)^4} \sum_{i < j < k < l} E[h_{ij}^2 h_{kl}^2] + E[h_{ik}^2 h_{jl}^2] + E[h_{il}^2 h_{jk}^2]. \end{aligned}$$

It is easy to see that [recall the notation of  $a(\varepsilon_i)$  in (6.12)]

$$\begin{aligned} E[h_{12}^4] &= \frac{1}{g^4} \int \int K_g^4(x_1 - x_2) m^8(x_1) m^8(x_2) E[a^4(\varepsilon_1) | X_1 = x_1] E[a^4(\varepsilon_2) | X_2 = x_2] f(x_1) f(x_2) \\ &\quad w^4(x_1) w^4(x_2) dx_1 dx_2 \\ &= O\left(\frac{1}{g^3}\right), \end{aligned}$$

which gives

$$G_I = \frac{2^4}{n^4(n-1)^4} \sum_{i < j} E[h_{ij}^4] = O\left(\frac{1}{n^6 g^3}\right).$$

The other terms are estimated similarly, i.e.  $G_{II} = O(\frac{1}{n^5 g^2})$ ,  $G_{III} = O(\frac{1}{n^5 g^2})$ ,  $G_{IV} = O(\frac{1}{n^4 g})$  and it follows from (6.13) that

$$(6.14) \quad E[\tilde{T}_n^4(c)] = 6G_V + o\left(\frac{1}{n^4 g^2}\right).$$

On the other hand a straightforward calculation shows

$$\frac{1}{n^4(n-1)^4} \sum_{i < j} \sigma_{ij}^4 = O\left(\frac{1}{n^6 g^3}\right), \quad \frac{1}{n^4(n-1)^4} \sum_{i < j < k} \sigma_{ij}^2 \sigma_{ik}^2 = O\left(\frac{1}{n^5 g^2}\right),$$

and we obtain

$$\sigma_n^4 = 2G_V + o\left(\frac{1}{n^4 g^2}\right),$$

which proves the second assumption in Theorem 2.1 of de Jong (1987), that is

$$E[\tilde{T}_n^4(c)] - 3\sigma_n^4 = o\left(\frac{1}{n^4 g^2}\right).$$

Now this theorem yields asymptotic normality of  $n\sqrt{g} \tilde{T}_n(c)$ , i.e.  $n\sqrt{g} \tilde{T}_n(c) \xrightarrow{D} \mathcal{N}(0, \mu_0^2)$ , and the assertion of Theorem 3.1(a) follows from (6.9) and (6.10).

**Proof of Theorem 3.1(b).** We first note that under a fixed alternative the statistic  $T_{3n}(c)$  is not negligible. More precisely, we show at the end of the proof

$$(6.15) \quad T_{3n}(c) - E[T_{3n}(c)] = 2T_{5n}(c) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

where

$$E[T_{3n}(c)] = h^2 \kappa_2 E[\Delta_c(X) m(X) m''(X) f(X) w^2(X)] + o(h^2).$$

We obtain for the statistic  $T_{6n}$  in (6.1)

$$\begin{aligned} \text{Var}(T_{6n}) &= \frac{2}{n^2(n-1)^2} \sum_{i \neq j} E[K_g^2(X_i - X_j) m^2(X_i) \sigma^2(X_i) w^2(X_i) m^2(X_j) \sigma^2(X_j) w^2(X_j)] \\ &= \frac{2}{n(n-1)g} E[m^4(x) \sigma^4(x) f(x) w^4(x)] \left( \int K^2(u) du \right) + o\left(\frac{1}{n^2 g}\right), \end{aligned}$$

which gives

$$(6.16) \quad \sqrt{n} T_{6n} \xrightarrow{P} 0$$

(note that the expectation of  $T_{6n}$  vanishes). A similar calculation yields

$$(6.17) \quad T_{4n}(c) - E[T_{4n}(c)] = \frac{2}{n} \sum_{i=1}^n \Delta_c(X_i, \varepsilon_i) \Delta_c(X_i) f(X_i) w^2(X_i) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$T_{5n}(c) = \frac{1}{n} \sum_{i=1}^n \Delta_c(X_i) m(X_i) \sigma(X_i) f(X_i) w^2(X_i) \varepsilon_i + o_p\left(\frac{1}{\sqrt{n}}\right)$$

and  $E[T_{4n}(c)] = E[\Delta_c^2(X)f(X)w^2(X)] + O(g^2)$ . Consequently we obtain from (6.1), (6.4), (6.16) and (6.17) with

$$E[T_n(c)] = E[\Delta_c^2(X)f(X)w^2(X)] + h^2B(c) + o(h^2)$$

the following stochastic expansion under a fixed alternative

$$(6.18) \quad \begin{aligned} T_n(c) - E[T_n(c)] &= (T_{4n}(c) - E[\Delta_c^2(X)f(X)w^2(X)]) + 4T_{5n}(c) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= Z_n(c) + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where the random variable  $Z_n(c)$  is defined by

$$(6.19) \quad \begin{aligned} Z_n(c) &= \frac{2}{n} \sum_{i=1}^n \left[ \Delta_c(X_i, \varepsilon_i) \Delta_c(X_i) f(X_i) w^2(X_i) - E[\Delta_c^2(X) f(X) w^2(X)] \right. \\ &\quad \left. + 2m(X_i) \sigma(X_i) \Delta_c(X_i) f(X_i) w^2(X_i) \varepsilon_i \right], \end{aligned}$$

and the assertion of Theorem 3.1(b) follows by a standard application of the central limit theorem verifying Ljapunoff's condition and observing that the dominating term on the right hand side of (6.18) has expectation 0 and variance  $\frac{\mu_1^2(c)}{n}$ .

The proof of Theorem 3.1 is now completed showing the remaining estimates (6.4) and (6.15).

**Proof of the estimate (6.4).** We consider exemplarily the case  $j = 1$  (the other case is treated by similar arguments) and obtain the decomposition

$$(6.20) \quad T_{1n} = 4T_{1n}^{(1)} + 4T_{1n}^{(2)} + T_{1n}^{(3)}$$

with

$$\begin{aligned} T_{1n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \tilde{\delta}_n(X_i) w(X_i) m(X_j) \tilde{\delta}_n(X_j) w(X_j), \\ T_{1n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \tilde{\delta}_n(X_i) w(X_i) \tilde{\delta}_n^2(X_j) w(X_j), \\ T_{1n}^{(3)} &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \tilde{\delta}_n^2(X_i) w(X_i) \tilde{\delta}_n^2(X_j) w(X_j) \end{aligned}$$

[recall the definition of  $\tilde{\delta}_n(x)$  in (6.6)]. The terms  $T_{1n}^{(\ell)}$  are all treated similarly and we consider again only the case  $\ell = 1$ . With the notation

$$(6.21) \quad d_k(x) = \frac{K_h(X_k - x)}{f(x)}$$

and Lemma 2 in Yao and Tong (2000) it follows

$$(6.22) \quad T_{1n}^{(1)} = \left( T_{1n}^{(1.1)} + \kappa_2 T_{1n}^{(1.2)} + \frac{\kappa_2^2}{4} T_{1n}^{(1.3)} \right) \{1 + o_p(1)\}$$

with

$$\begin{aligned} T_{1n}^{(1.1)} &= \frac{1}{n^3(n-1)} \sum_{i \neq j, k, l} K_g(X_i - X_j) d_k(X_i) m(X_i) w(X_i) d_l(X_j) m(X_j) w(X_j) \sigma(X_k) \varepsilon_k \sigma(X_l) \varepsilon_l, \\ T_{1n}^{(1.2)} &= \frac{h^2}{n^2(n-1)} \sum_{i \neq j, k} K_g(X_i - X_j) d_k(X_i) m(X_i) w(X_i) \sigma(X_k) \varepsilon_k m(X_j) m''(X_j) w(X_j), \\ T_{1n}^{(1.3)} &= \frac{h^4}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) m''(X_i) w(X_i) m(X_j) m''(X_j) w(X_j). \end{aligned}$$

The terms  $T_{1n}^{(1.k)}$ ,  $k = 1, 2, 3$ , can now be treated by calculating expectation and variance. For example,

$$\begin{aligned} & E[T_{1n}^{(1.1)}] \\ &= \frac{1}{n^3(n-1)} \sum_{i \neq j, k} E [K_g(X_i - X_j) d_k(X_i) m(X_i) w(X_i) d_k(X_j) m(X_j) w(X_j) \sigma^2(X_k)] + O\left(\frac{1}{n^2 h^2}\right) \\ &= \frac{1}{n} \int K_g(x_i - x_j) K_h(x_k - x_i) K_h(x_k - x_j) \\ &\quad \times m(x_i) w(x_i) m(x_j) w(x_j) \sigma^2(x_k) f(x_k) dx_i dx_j dx_k + O\left(\frac{1}{n^2 h^2}\right) \\ &= \frac{1}{nh} \int m^2(x) \sigma^2(x) f(x) w^2(x) dx \int K^2(w) dw + o\left(\frac{1}{nh}\right) + O\left(\frac{1}{n^2 h^2}\right). \\ &= O\left(\frac{1}{nh}\right) + O\left(\frac{1}{n^2 h^2}\right) = O\left(\frac{1}{nh}\right). \end{aligned}$$

For the calculation of the variance of  $T_{1n}^{(1.1)}$  we introduce the notation  $V_i = (X_i, \varepsilon_i)$ ,

$$t_{1n}^{(1.1)}(V_i, V_j, V_k, V_l) = K_g(X_i - X_j) d_k(X_i) m(X_i) w(X_i) \sigma(X_k) \varepsilon_k d_l(X_j) m(X_j) w(X_j) \sigma(X_l) \varepsilon_l$$

and obtain

$$(6.23) \quad \begin{aligned} \text{Var}(T_{1n}^{(1.1)}) &\leq E[(T_{1n}^{(1.1)})^2] \\ &= \frac{1}{n^6(n-1)^2} \sum_{i \neq j, k, l} \sum_{p \neq q, r, s} E \left[ t_{1n}^{(1.1)}(V_i, V_j, V_k, V_l) t_{1n}^{(1.1)}(V_p, V_q, V_r, V_s) \right] = O\left(\frac{1}{n^2 h^2}\right), \end{aligned}$$

where we have used the fact, that the sum in (6.23) is dominated by those expectations, for which two indices of  $k, l, r, s$  coincide. Markov's inequality and the assumptions on the bandwidth now

yield  $T_{1n}^{(1.1)} = o_p(n^{-1}g^{-1/2})$ . The terms  $T_{1n}^{(1.2)}$  and  $T_{1n}^{(1.3)}$  in (6.22) are treated similarly, which implies  $T_{1n}^{(1)} = o_p(n^{-1}g^{-1/2})$ . A similar argument for the statistics  $T_{1n}^{(2)}$  and  $T_{1n}^{(3)}$  yields assertion (6.4) for  $j = 1$ . The second case  $j = 2$  is treated in the same way.

**Proof of the estimate (6.15).** With the representation

$$\delta_n(x) = \hat{m}^2(x) - m^2(x) = 2m(x)\tilde{\delta}_n(x) + \tilde{\delta}_n^2(x)$$

[recall the definition of  $\tilde{\delta}(x)$  in (6.6)] we have

$$(6.24) \quad T_{3n}(c) = 2T_{3n}^{(1)}(c) + T_{3n}^{(2)}(c)$$

with

$$\begin{aligned} T_{3n}^{(1)}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) \tilde{\delta}_n(X_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j), \\ T_{3n}^{(2)}(c) &= \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \tilde{\delta}_n^2(X_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j). \end{aligned}$$

A similar calculation as used in the proof of (6.4) yields

$$(6.25) \quad T_{3n}^{(2)}(c) = O_p\left(\frac{1}{nh}\right) = o_p\left(\frac{1}{\sqrt{n}}\right).$$

For the first term in (6.24) we introduce a further decomposition applying Lemma 2 in Yao and Tong (2000) and obtain

$$(6.26) \quad T_{3n}^{(1)}(c) = \left(T_{3n}^{(1.1)}(c) + \frac{\kappa_2}{2} T_{3n}^{(1.2)}\right) \{1 + o_p(1)\}$$

with [recall the definition of  $d_k(\cdot)$  in (6.21)]

$$\begin{aligned} T_{3n}^{(1.1)}(c) &= \frac{1}{n^2(n-1)} \sum_{i \neq j, k} K_g(X_i - X_j) d_k(X_i) m(X_i) w(X_i) \sigma(X_k) \varepsilon_k \Delta_c(X_j, \varepsilon_j) w(X_j), \\ T_{3n}^{(1.2)} &= \frac{h^2}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) m(X_i) m''(X_i) w(X_i) \Delta_c(X_j, \varepsilon_j) w(X_j). \end{aligned}$$

A straightforward but tedious calculation shows that  $T_{3n}^{(1.1)}(c)$  and  $T_{5n}(c)$  are asymptotically equivalent, that is

$$(6.27) \quad E[(T_{3n}^{(1.1)}(c) - T_{5n}(c))^2] = o\left(\frac{1}{n}\right).$$

For the second term we obtain by a Taylor expansion

$$(6.28) \quad E[T_{3n}^{(1.2)}] = h^2 E[\Delta_c(X) m(X) m''(X) f(X) w^2(X)] + o\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\text{Var}(T_{3n}^{(1,2)}) = o\left(\frac{1}{n}\right),$$

which yields

$$(6.29) \quad \sqrt{n} \left( T_{3n}^{(1,2)} - E[T_{3n}^{(1,2)}] \right) = o_p(1).$$

Combining (6.24) - (6.29) and noting that  $E[T_{3n}(c)] = \kappa_2 E[T_{3n}^{1,2}(c)] + o(h^2)$  now establishes (6.15) and completes the proof of Theorem 3.1.  $\square$

## 6.2 Proof of Theorem 3.2.

Note that  $\hat{c}^2 = \frac{R_1}{R_2}$  where

$$R_1 := \frac{1}{n} \sum_{i=1}^n \hat{m}^2(X_i) \hat{r}^2(X_i) w(X_i), \quad R_2 := \frac{1}{n} \sum_{i=1}^n (\hat{\sigma}^2(X_i))^2 w(X_i).$$

We now apply Lemma 2 in Yao and Tong (2000) and Lemma 1 in Fan and Yao (1998), which give asymptotic representations of  $\hat{m}(x) - m(x)$  and  $\hat{\sigma}^2(x) - \sigma^2(x)$ . A straightforward but tedious calculation yields

$$(6.30) \quad R_1 - E[R_1] = \frac{1}{n} \sum_{i=1}^n \left\{ \left( m^2(X_i) \sigma^2(X_i) w(X_i) \varepsilon_i^2 - E[m^2(X) \sigma^2(X) w(X)] \right) \right. \\ \left. + 2 m(X_i) \sigma^3(X_i) w(X_i) \varepsilon_i \right\} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

$$(6.31) \quad R_2 - E[R_2] = \frac{1}{n} \sum_{i=1}^n \left\{ \left( \sigma^4(X_i) w(X_i) - E[\sigma^4(X) w(X)] \right) \right. \\ \left. + 2 \sigma^4(X_i) w(X_i) \{ \varepsilon_i^2 - 1 \} \right\} + o_p\left(\frac{1}{\sqrt{n}}\right)$$

with

$$E[R_1] = E[m^2(X) \sigma^2(X) w(X)] + h^2 \kappa_2 E[m(X) m''(X) \sigma^2(X) w(X)] + o(h^2), \\ E[R_2] = E[\sigma^4(X) w(X)] + h^2 \kappa_2 E[\sigma^2(X) (\sigma^2(X))'' w(X)] + o(h^2).$$

With this representation and the notations  $r_1 := E[m^2(X) \sigma^2(X) w(X)]$  and  $r_2 := E[\sigma^4(X) w(X)]$  it is easy to see that

$$\frac{1}{R_2} = \frac{1}{r_2} \left\{ 1 - \frac{R_2 - r_2}{r_2} + o\left(\frac{1}{\sqrt{n}}\right) + o_p\left(\frac{1}{\sqrt{n}}\right) \right\},$$

which implies

$$(6.32) \quad \hat{c}^2 - c_0^2 = \frac{R_1 - r_1}{r_2} - \frac{R_1(R_2 - r_2)}{r_2^2} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where  $c_0^2$  is defined in (2.7). Observing (6.30) we have

$$\begin{aligned} R_1(R_2 - r_2) &= E[R_1](R_2 - r_2) + (R_1 - E[R_1])(R_2 - r_2) \\ &= E[R_1](R_2 - E[R_2]) + E[R_1](E[R_2] - r_2) + O_p\left(\frac{h^2}{\sqrt{n}}\right) \\ &= r_1(R_2 - E[R_2]) + E[R_1](E[R_2] - r_2) + O_p\left(\frac{h^2}{\sqrt{n}}\right), \end{aligned}$$

and we obtain

$$\begin{aligned} \hat{c}^2 - c_0^2 &= \frac{R_1 - E[R_1]}{r_2} + \frac{E[R_1] - r_1}{r_2} - \frac{r_1(R_2 - E[R_2])}{r_2^2} - \frac{E[R_1](E[R_2] - r_2)}{r_2^2} + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= h^2\Gamma + o(h^2) + \tilde{Z}_n + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where the random variable  $\tilde{Z}_n$  is defined as

$$(6.33) \quad \begin{aligned} \tilde{Z}_n &= \frac{1}{n} \sum_{i=1}^n \left\{ \tau_1 \left( m^2(X_i) \sigma^2(X_i) w(X_i) \varepsilon_i^2 - E[m^2(X) \sigma^2(X) w(X)] \right) + 2\tau_1 m(X_i) \sigma^3(X_i) w(X_i) \varepsilon_i \right. \\ &\quad \left. - \tau_2 \left( \sigma^4(X_i) w(X_i) - E[\sigma^4(X) w(X)] \right) - 2\tau_2 \sigma^4(X_i) w(X_i) \{ \varepsilon_i^2 - 1 \} \right\}. \end{aligned}$$

The assertion finally follows from the central limit theorem and a straightforward but tedious calculation of the variance  $n\text{Var}(\tilde{Z}_n) = \nu^2$ .  $\square$

### 6.3 Proof of Theorem 3.3.

The assertion of the theorem follows, if the estimate

$$(6.34) \quad T_n(c) - T_n(\hat{c}) = o_p\left(\frac{1}{n\sqrt{g}}\right)$$

can be established. For this purpose we introduce the analogue of the decomposition (6.1), and obtain

$$(6.35) \quad \begin{aligned} T_n(\hat{c}) &= (\hat{c}^2 + 1)^2 T_{1n} - 4\hat{c}^2(\hat{c}^2 + 1) T_{2n} + 2(\hat{c}^2 + 1) T_{3n}(\hat{c}) + T_{4n}(\hat{c}) - 4\hat{c}^2 T_{5n}(\hat{c}) + 4\hat{c}^4 T_{6n} \\ &= 2(\hat{c}^2 + 1) T_{3n}(\hat{c}) + T_{4n}(\hat{c}) - 4\hat{c}^2 T_{5n}(\hat{c}) + 4\hat{c}^4 T_{6n} + o_p\left(\frac{1}{n\sqrt{g}}\right), \end{aligned}$$

where the last equality follows from (6.4) and Theorem 3.2. For the third term in (6.35) we have

$$T_{3n}(\hat{c}) = T_{3n}(c_0) - (\hat{c}^2 - c_0^2) T_{3n}^{(a)},$$

where

$$T_{3n}^{(a)} = \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \delta_n(X_i) w(X_i) \sigma^2(X_j) w(X_j) \varepsilon_j^2 = O_p(h^2) + O_p\left(\frac{1}{nh}\right)$$

[this estimate can be shown by similar arguments as presented in the proof of the estimate (6.4)].  
Consequently we obtain from Theorem 3.2

$$(6.36) \quad T_{3n}(\hat{c}) - T_{3n}(c_0) = o_p\left(\frac{1}{n\sqrt{g}}\right).$$

The corresponding estimates

$$(6.37) \quad T_{4n}(\hat{c}) - T_{4n}(c_0) \stackrel{H_0}{=} o_p\left(\frac{1}{n\sqrt{g}}\right),$$

$$(6.38) \quad \hat{c}^2 T_{5n}(\hat{c}) = c_0^2 T_{5n}(c_0) + o_p\left(\frac{1}{n\sqrt{g}}\right),$$

$$(6.39) \quad (\hat{c}^2)^2 T_{6n} = c_0^4 T_{6n} + O_p\left(\frac{h^2}{n\sqrt{g}}\right)$$

are proved by similar arguments, which are not given here for the sake of brevity. Note that the estimates (6.38) and (6.39) hold also under a fixed alternative. Recalling

$$c_0^2 \stackrel{H_0}{=} c^2$$

and combining (6.35) - (6.38) with Theorem 3.2 yields

$$(6.40) \quad T_n(\hat{c}) \stackrel{H_0}{=} T_{4n}(c) - 4c^2 T_{5n}(c) + 4c^4 T_{6n} + o_p\left(\frac{1}{n\sqrt{g}}\right) \stackrel{H_0}{=} T_n(c) + o_p\left(\frac{1}{n\sqrt{g}}\right)$$

where the last equality follows from the proof of Theorem 3.1(a). This proves the first identity in Theorem 3.3, while the remaining part follows again from Theorem 3.1.  $\square$

## 6.4 Proof of Theorem 3.4.

Recalling the decomposition (6.35) and (6.38), (6.39) we obtain

$$(6.41) \quad T_n(\hat{c}) = 2(\hat{c}^2 + 1)T_{3n}(\hat{c}) + T_{4n}(\hat{c}) - 4c_0^2 T_{5n}(c_0) + 4c_0^4 T_{6n} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

and it remains to investigate the asymptotic properties of  $T_{3n}(\hat{c})$  and  $T_{4n}(\hat{c})$  under the fixed alternative. For this we note that (6.36) and the obvious estimate  $T_{3n}(c_0) = o_p(1)$  yield

$$(6.42) \quad (\hat{c}^2 + 1)T_{3n}(\hat{c}) - (c_0^2 + 1)E[T_{3n}(c_0)] = 2(c_0^2 + 1)T_{5n}(c_0) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

The corresponding estimate for the term  $T_{4n}(\hat{c})$  is more difficult. Note first that

$$(6.43) \quad T_{4n}(\hat{c}) = T_{4n}(c_0) - 2(\hat{c}^2 - c_0^2)T_{4n}^{(a)}(c_0) + (\hat{c}^2 - c_0^2)^2 T_{4n}^{(b)},$$

where

$$T_{4n}^{(a)}(c_0) \stackrel{H_1}{=} \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) (m^2(X_i) - c_0^2 \sigma^2(X_i) \varepsilon_i^2) w(X_i) \sigma^2(X_j) w(X_j) \varepsilon_j^2,$$

$$T_{4n}^{(b)} \stackrel{H_1}{=} \frac{1}{n(n-1)} \sum_{i \neq j} K_g(X_i - X_j) \sigma^2(X_i) w(X_i) \varepsilon_i^2 \sigma^2(X_j) w(X_j) \varepsilon_j^2.$$

A standard calculation yields  $T_{4n}^{(b)} = O_p(1)$ , which shows that the second term in (6.43) is of order  $o_p(\frac{1}{\sqrt{n}})$ . From Theorem 3.2. it follows

$$(\hat{c}^2 - c_0^2) T_{4n}^{(a)}(c_0) = \varrho(\hat{c}^2 - E[\hat{c}^2]) + h^2 \Gamma E[\Delta_{c_0}(X) \sigma^2(X) f(X) w^2(X)] + o(h^2) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

where  $\Delta_{c_0}(X)$  is defined in (2.3). Combining this with (6.43) therefore yields

$$\begin{aligned} T_n(\hat{c}) - E[T_n(\hat{c})] &= T_{4n}(c_0) - E[T_{4n}(c_0)] + 4T_{5n}(c_0) - 2\varrho(\hat{c}^2 - E[\hat{c}^2]) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &= Z_n(c_0) - 2\varrho\tilde{Z}_n + o_p\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where the random variables  $Z_n(c_0)$  and  $\tilde{Z}_n$  are defined in (6.19) and (6.33), respectively, and

$$E[T_n(\hat{c})] = E[\Delta_{c_0}^2(X) f(X) w^2(X)] + h^2(B(c_0) - 2\varrho\Gamma) + o(h^2).$$

The variances of  $Z_n(c_0)$  and  $\tilde{Z}_n$  have been determined in the proof of Theorem 3.1 and 3.2 and a straightforward calculation gives  $n \text{Cov}(Z_n(c_0), \tilde{Z}_n) = v^2(c_0)$ , where  $v^2(c_0)$  is defined in Theorem 3.4. The assertion now follows from a standard application of the central limit theorem.  $\square$

## Acknowledgements.

The work of the authors was supported by the Sonderforschungsbereich 475, Komplexitätsreduktion in multivariaten Datenstrukturen (Teilprojekt B1) and an NIH grant award IR01GM072876:01A1. The work of G. Wieczorek was additionally supported by a grant of the Ruhr-University Research School.

## References

R.J. Carroll, D. Ruppert (1988). *Transformation and weighting in regression*. Chapman and Hall, London.

- H. Dette (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *J. Stat. Plann. Inference*, 103, (1-2), 311-329.
- H. Dette, A. Munk (2003). Some methodological aspects of validation of models in nonparametric regression. *Statistica Neerlandica*, 57(2), 207-244.
- H. Dette, C. von Lieres und Wilkau (2003). On a test for constant volatility in continuous time financial models. *Financ. Stoch.* 7, 363-384.
- H. Dette, I. Spreckelsen (2004). Some comments on specification tests in nonparametric absolutely regular processes. *J. Time Ser. Anal.* 25. No. 2. 159-172.
- G.K. Eagleson, H.G. Müller (1997). Transformations for smooth regression models with multiplicative errors. *J. R. Statist. Soc. B.*, 59(1), 173-189.
- R.F. Engle (1982). Autoregressive heteroscedasticity of estimates of the variance of U.K. inflation. *Econometrica*, 50(4), 987-1007.
- J. Fan, I. Gijbels (1996). *Local polynomial modelling and its applications*. Chapman and Hall, London.
- J. Fan, Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3), 645-660.
- P. Gozalo, O. Linton (2000). Local nonlinear least squares: using parametric information in nonparametric regression. *Journal of Econometrics*, 99, 63-106.
- P. de Jong (1987). A central limit theorem for generalized quadric forms. *Probability Theory and Related Fields*, 75, 261-277.
- P. Mc Cullagh, J. Nelder (1989). *Generalized Linear Models*. 2<sup>nd</sup> ed. Chapman and Hall, London.
- A. Miløj (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics*, 12, 281-292.
- G. Wieczorek (2007). Tests auf multiplikative Struktur in strikt stationären Prozessen. *PhD-Thesis, Ruhr University Bochum (in German)*.
- Q. Yao, H. Tong (2000). Nonparametric estimation of ratios of noise to signal in stochastic regression. *Statistica Sinica*, 10, 751-770.
- J.X. Zheng (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75, 263-289.