Testing for a constant coefficient of variation in nonparametric regression

Holger Dette*, Mareen Marchlewski and Jens Wagener

Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, Germany

November 9, 2010

Abstract

In the common nonparametric regression model $Y_i = m(X_i) + \sigma(X_i)\epsilon_i$ we consider the problem of testing the hypothesis that the coefficient of the scale and location function is constant. The test is based on a comparison of the observations $Y_i/\hat{\sigma}(X_i)$ with their mean by a smoothed empirical process, where $\hat{\sigma}$ denotes the local linear estimate of the scale function. We show weak convergence of a centered version of this process to a Gaussian process under the null hypothesis and the alternative and use this result to construct a test for the hypothesis of a constant coefficient of variation in the nonparametric regression model. A small simulation study is also presented to investigate the finite sample properties of the new test.

Keywords and phrases: nonparametric regression, test for constant coefficient of variation, smoothed empirical process

AMS Subject Classification: 62G10, 62F35

1 Introduction

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ denote a bivariate sample of independent identically distributed observations corresponding to the nonparametric regression model

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

*Corresponding author. Email: holger.dette@rub.de

where $\epsilon_1, \ldots, \epsilon_n$ are independent identically distributed random variables with $E[\epsilon_i \mid X_i = t] = 0$ and $E[\epsilon_i^2 \mid X_i = t] = 1$ for all t. The functions $m(t) = E[Y_i \mid X_i = t]$ and $\sigma^2(t) = E[(Y_i - m(X_i))^2 \mid X_i = t]$ are called regression and variance function, respectively. In this paper we are interested in the problem of testing for a constant coefficient of variation in the nonparametric regression model (1.1), that is

$$H_0 : m(t) = c\sigma(t) \qquad a.e. \tag{1.2}$$

for some positive constant c. Several authors have discussed the problem of statistical inference under the assumption of a constant coefficient of variation. For example McCullagh and Nelder (1989) considered generalized linear models, Carroll and Ruppert (1988) investigated a parametric model with a constant coefficient of variation, while Eagleson and Müller (1997) considered the problem of nonparametric estimation of the regression function in a model where the standard deviation function is proportional to the regression function. The problem of testing the hypothesis (1.2) in a general nonparametric regression model has been recently discussed in Dette and Wieczorek (2009) and Dette et al. (2009). The last named authors investigated the difference between two empirical processes under the null hypothesis and the alternative and showed weak convergence. Because the distribution of the limiting process depends in a complicated way on the distribution of the random variable (X_1, Y_1) a bootstrap procedure is used to obtain critical values for the corresponding test. This test can detect alternatives converging to the null hypothesis with the rate $1/\sqrt{n}$. Dette and Wieczorek (2009) proposed a test which is based on an L^2-distance between an estimate of $m(t)/\sigma(t)$ under the null hypothesis and the alternative. The limiting distribution is a normal distribution, but the test is not able to detect alternatives converging to the null at the rate $1/\sqrt{n}$.

The purpose of the present paper is to propose a third test for the hypothesis (1.2) which on the one hand is based on a test statistic with a simple limiting distribution and on the other hand is able to detect alternatives converging to the null hypothesis at a parametric rate. Our approach is based on the fact that under the null hypothesis the conditional expectation of the random variables $Y_1/\sigma(X_1), \ldots, Y_n/\sigma(X_n)$ is constant and therefore equal to their mean. Consequently, we propose to consider a smoothed empirical process of the random variables $U_i - \bar{U}$ as the basis for our test, where $U_i = Y_i/\hat{\sigma}(X_i)$ and $\hat{\sigma}$ is the local linear estimate of the scale function. The test statistic is carefully defined in Section 2 where we also state the necessary assumptions for the asymptotic theory, which is presented in Section 3. Finally, a simulation study, which investigates the finite sample properties of the new test is presented in Section 4, while some of the very technical details are deferred to an Appendix in Sections 5 -7.

2 Preliminaries and a new test statistic

Let f denote the marginal density of X_1. We propose to base a test for the hypothesis (1.2) on the process

$$H_n(t) = \int_{-\infty}^{t} \frac{1}{n} \sum_{j=1}^{n} \omega(X_j) K_h(X_j - x)(U_j - \bar{U}) dx, \tag{2.1}$$

where the random variables U_j and $\bar{U}$ are defined by $U_j = \frac{Y_j}{\hat{\sigma}(X_j)}$ and $\bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i$, respectively, $\omega(\cdot)$ denotes a weight function, $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, $K(\cdot)$ is a kernel with compact support and h is a bandwidth converging to 0 with increasing sample size. This definition requires the specification of an estimate of the variance function $\hat{\sigma}$ and throughout this paper we use the local linear estimate

$$\hat{\sigma}^2(t) = \frac{1}{n} \sum_{i=1}^{n} W_i(t) \left(Y_i - \hat{m}(X_i)\right)^2, \tag{2.2}$$

where $\hat{m}$ is the local linear estimate of the regression function, i.e.

$$\hat{m}(t) = \frac{1}{n} \sum_{i=1}^{n} W_i(t) Y_i,$$

the weights $W_i(t)$ are defined by

$$W_i(t) = \frac{\tilde{K}_g(X_i - t)\left[s_{n,2}(t) - (t - X_i)s_{n,1}(t)\right]}{\frac{1}{n}\sum_{j=1}^{n} \tilde{K}_g(X_j - t)\left[s_{n,2}(t) - (t - X_j)s_{n,1}(t)\right]},$$

$$s_{n,l}(t) = \sum_{i=1}^{n} \tilde{K}_g(X_i - t)(t - X_i)^l, \quad l = 1, 2,$$

and $\tilde{K}_g(\cdot) = \frac{1}{g}\tilde{K}(\frac{\cdot}{g})$ [see Fan and Gijbels (1996)]. Here g and $\tilde{K}(\cdot)$ denote a further bandwidth and kernel function with compact support, respectively. Note that we use the same bandwidth in the weights for the estimates of the variance and regression function in order to keep the technical arguments simple. However, the results presented in this paper are correct for estimates $\hat{\sigma}^2$ and $\hat{m}$ with different bandwidths [see Ruppert et al. (1997) or Yu and Jones (2004) among others]. While the choice of these smoothing parameters is important for a good performance of the proposed test, it also worthwhile to mention that the procedure is relatively robust with respect to the choice of the bandwidth h (see the discussion in Section 4).

We illustrate our approach in Figure 1 showing the process $\{H_n(t)\}_{t\in[0,1]}$ under the null hypothesis $m(x) = \sigma(x) = 0.5(1+0.1x)$ and alternative $m(x) = 0.5(1+0.1x)$, $\sigma(x) = (1+0.1x+2\sqrt{x})$,

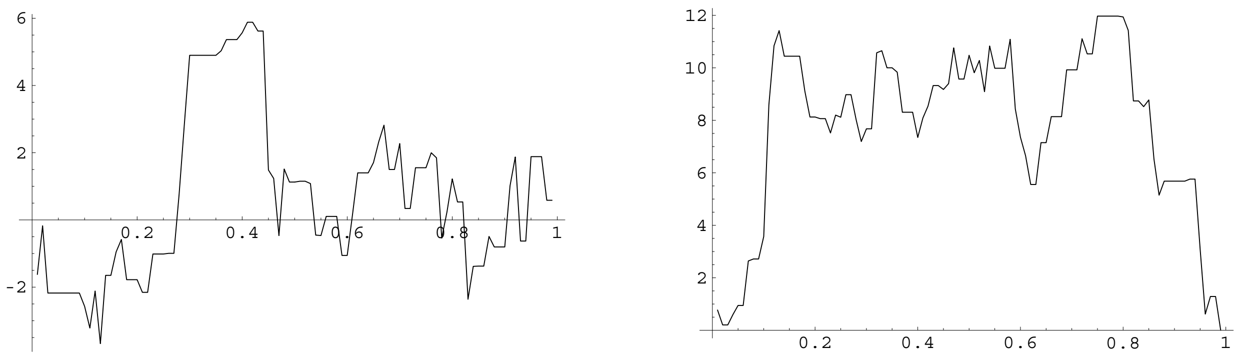

Figure 1: *The process $\{H_n(t)\}_{t\in[0,1]}$ under the null hypothesis (left panel) and alternative (right panel).*

where the sample size is $n = 200$, the bandwidth of the estimators have been chosen by cross validation and $h = 1/n^{0.9}$.

In order to prove our asymptotic theory we will require the following basic assumptions, where C denotes a generic constant which may have different values in different contexts.

(A1) The bandwidths g and h satisfy for $n \to \infty$

$$ng^2 \longrightarrow \infty, \quad ng^4 \longrightarrow 0, \quad h \longrightarrow 0, \quad nh \longrightarrow \infty$$

(A2) The density f of the predictors X_i has compact support, say $[0,1]$, is twice continuously differentiable on $(0,1)$ and $f(t) \geq C > 0$ for all $t \in [0,1]$.

(A3) The regression function $m : [0,1] \longrightarrow \mathbb{R}$ is twice continuously differentiable.

(A4) The variance function $\sigma^2 : [0,1] \longrightarrow \mathbb{R}$ is twice continuously differentiable and $\min_{t\in[0,1]} \sigma^2(t) \geq C > 0$.

(A5) The weight function ω is twice continuously differentiable and has compact support $[0,1]$.

(A6) The kernels K and $\tilde{K}$ are of order 2, Lipschitz continuous and have compact support, say $[-1,1]$.

(A7) The conditional expectations $m_j(t) = E[\epsilon_1^j \mid X_1 = t]$ exist, are continuous for $j = 3,4$ and for all $1 \leq j \leq 8$ bounded, that is

$$\left| E[\epsilon_i^j \mid X_i = t] \right| \leq C < \infty, \;\; 1 \leq j \leq 8.$$

We will show in Section 5 and 7 that

$$H_n(t) - H_n^*(t) \xrightarrow{P} 0$$

for any $t \in [0,1]$, where $\{H_n^*(t)\}_{t\in[0,1]}$ denotes a deterministic function defined by

$$H_n^*(t) = \frac{n-1}{n} \int_0^1 f(x_1)c_t(x_1) \left(\frac{m(x_1)}{\sigma(x_1)} - \int_0^1 f(x_2)\frac{m(x_2)}{\sigma(x_2)}dx_2 \right) dx_1, \tag{2.3}$$

and

$$c_t(x_1) \quad = \quad \omega(x_1) \int_{-\infty}^t K_h(x_1 - u)du.$$

Note that

$$\lim_{n\to\infty} H_n^*(t) = H(t) = \int_0^t \varphi(x_1)dx_1$$

where

$$\varphi(x_1) = f(x_1)\omega(x_1) \left(\frac{m(x_1)}{\sigma(x_1)} - \int_0^1 f(x_2)\frac{m(x_2)}{\sigma(x_2)}dx_2 \right). \tag{2.4}$$

The following Lemma shows that the process $\{H(t)\}_{t\in[0,1]}$ vanishes if and only if the null hypothesis (1.2) of a constant coefficient of variation is satisfied.

Lemma 2.1 *If the assumptions (A2) and (A5) are satisfied, then the null hypothesis (1.2) of a constant coefficient of variation is satisfied if and only if* $H(t) = 0$ *almost everywhere on the interval* $[0,1]$.

Proof of Lemma 2.1. If the null hypothesis (1.2) is satisfied then there exists a $c > 0$, such that $m(t) = c\sigma(t)$ and

$$H(t) = 0$$

almost everywhere on the interval $[0,1]$. For the converse assume that $H(t) = 0$ almost everywhere on the interval $[0,1]$. This yields $H'(t) = 0$ almost everywhere on $[0,1]$, i.e.

$$0 = \frac{\partial}{\partial t} H(t) = \varphi(t)$$

almost everywhere on $[0,1]$, where we used the notation (2.4).

Observing that the functions f and ω do not vanish on the interval $[0,1]$ we obtain

$$\frac{m(t)}{\sigma(t)} - \int_0^1 f(x_2)\frac{m(x_2)}{\sigma(x_2)}dx_2 = 0$$

almost everywhere on the interval $[0,1]$, which proves the assertion of Lemma 2.1. $\Box$

Lemma 2.1 suggests to construct a statistical test for the hypothesis (1.2) on the basis of the process $\{H_n(t)\}_{t\in[0,1]}$. More precisely we propose to reject the null hypothesis of a constant coefficient of variation for large values of the Kolmogorov-Smirnov statistic

$$\hat{K}_n = \sup_{t\in[0,1]} \mid H_n(t) \mid$$

or the Cramér-von-Mises statistic

$$\hat{C}_n = \int_0^1 H_n^2(t) d\hat{F}_n(t),$$

where $\hat{F}_n(t) = \frac{1}{n}\sum_{j=1}^n I\{X_j \leq t\}$ denotes the empirical distribution function of $X_1,\dots,X_n$. In the following section we will study the asymptotic properties of the process $\{H_n(t)\}_{t\in[0,1]}$, which will be used to derive critical values for these tests.

3 Asymptotic properties

Throughout this paper $D[0,1]$ denotes the space of cadlag functions defined on the interval $[0,1]$ [see Billingsley (1999)]. Then the following result establishes weak convergence of the process $\{\sqrt{n}(H_n(t) - H_n^*(t))\}_{t\in[0,1]}$ in $D[0,1]$. A proof is given in Section 5

Theorem 3.1 *If the assumptions (A1)-(A7) are satisfied, then*

$$\{\sqrt{n}(H_n(t) - H_n^*(t))\}_{t\in[0,1]} \Rightarrow \{A(t)\}_{t\in[0,1]}$$

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

$$\begin{aligned} k(t,s) &= \int_0^{s\wedge t} \omega^2(u)f(u)du - \int_0^s \omega(u)f(u)du \int_0^t \omega(u)f(u)du \\ &\quad +\kappa(s,t) + 2\nu(s,t) + \mu(s,t) \end{aligned} \tag{3.1}$$

and the functions $\kappa(s,t)$, $\nu(s,t)$ and $\mu(s,t)$ are defined by

$$\begin{aligned} \kappa(s,t) &= \mathrm{Cov}(I_{\omega,t}(X_1)h(X_1), I_{\omega,s}(X_1)h(X_1)) \\ &\quad +\mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_1)h(X_3)) + \mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_2)h(X_3)) \\ &\quad +\mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_3)h(X_1)) + \mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_3)h(X_2)) \\ &\quad -\mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_1)h(X_1)) - \mathrm{Cov}(I_{\omega,t}(X_1)h(X_2), I_{\omega,s}(X_2)h(X_2)) \\ &\quad -\mathrm{Cov}(I_{\omega,t}(X_1)h(X_1), I_{\omega,s}(X_1)h(X_2)) - \mathrm{Cov}(I_{\omega,t}(X_1)h(X_1), I_{\omega,s}(X_2)h(X_1)), \end{aligned} \tag{3.2}$$

$$\begin{aligned}
\nu(s,t) &= -\frac{1}{2}\int_0^{s\wedge t}\omega^2(u)f(u)h(u)m_3(u)du \\
&\quad +\frac{1}{2}\int_0^t \omega(u)f(u)du\int_0^s \omega(u)f(u)h(u)m_3(u)du \\
&\quad +\frac{1}{2}\int_0^s \omega(u)f(u)du\int_0^t \omega(u)f(u)h(u)m_3(u)du \\
&\quad -\frac{1}{2}\int_0^s \omega(u)f(u)du\int_0^t \omega(u)f(u)du\int_0^1 f(u)h(u)m_3(u)du,
\end{aligned} \tag{3.3}$$

$$\begin{aligned}
\mu(s,t) &= \frac{1}{4}\int_0^{s\wedge t}\omega^2(u)f(u)h^2(u)(m_4(u)-1)du \\
&\quad -\frac{1}{4}\int_0^t \omega(u)f(u)du\int_0^s \omega(u)f(u)h^2(u)(m_4(u)-1)du \\
&\quad -\frac{1}{4}\int_0^s \omega(u)f(u)du\int_0^t \omega(u)f(u)h^2(u)(m_4(u)-1)du \\
&\quad +\frac{1}{4}\int_0^s \omega(u)f(u)du\int_0^t \omega(u)f(u)du\int_0^1 f(u)h^2(u)(m_4(u)-1)du,
\end{aligned} \tag{3.4}$$

respectively. Here the notation $I_{\omega,t}(X_1) = \omega(X_1)I\{X_1 \le t\}$ and $h(t) = m(t)/\sigma(t)$ is used.

Note that Theorem 3.1 holds under the null hypothesis and the alternative. However, in the case of a constant coefficient of variation the limiting process simplifies substantially.

Corollary 3.1 *If the assumptions of Theorem 3.1 and the null hypothesis (1.2) are satisfied with coefficient of variation $c > 0$, then, under the additional assumption of constant conditional moments $m_3(t) \equiv m_3$ and $m_4(t) \equiv m_4$ and $\omega(t) \equiv 1$, we have*

$$\{\sqrt{n}H_n(t)\}_{t\in[0,1]} \Rightarrow \sqrt{1 - cm_3 + \frac{c^2}{4}(m_4-1)}\{B\circ F\}_{t\in[0,1]}$$

in $D[0,1]$, where F denotes the distribution function of X_1 and B is a standard Brownian bridge.

From Corollary 3.1 we obtain by the continuous mapping theorem for the corresponding Kolmogorov-Smirnov and Cramér-von-Mises statistic

$$\frac{\sqrt{n}\hat{K}_n}{\sqrt{1 - cm_3 + \frac{c^2}{4}(m_4-1)}} \xrightarrow{D} \sup_{t\in[0,1]} \mid B(t) \mid, \tag{3.5}$$

$$\frac{n\hat{C}_n}{1 - cm_3 + \frac{c^2}{4}(m_4-1)} \xrightarrow{D} \int_0^1 B^2(t)dt. \tag{3.6}$$

This yields an asymptotic level α test by rejecting the null hypothesis of a constant coefficient of variation if

$$\sqrt{n}\hat{K}_n > \sqrt{1 - \hat{c}\hat{m}_3 + \frac{\hat{c}^2}{4}(\hat{m}_4 - 1)}k_{1-\alpha} \qquad \text{or} \qquad n\hat{C}_n > \left(1 - \hat{c}\hat{m}_3 + \frac{\hat{c}^2}{4}(\hat{m}_4 - 1)\right) c_{1-\alpha}, \quad (3.7)$$

where $\hat{\sigma}$, $\hat{m}_3$ and $\hat{m}_4$ are appropriate consistent estimates of the quantities c, m_3 and m_4, respectively, and $k_{1-\alpha}$ and $c_{1-\alpha}$ denote the $(1-\alpha)$-quantiles of the corresponding limiting distributions in (3.5) and (3.6), respectively. Note that the consistency of this test follows from Lemma 2.1 and Theorem 3.1 which shows that under the alternative we have $\sqrt{n}\hat{K}_n \xrightarrow{P} \infty$ and $n\hat{C}_n \xrightarrow{P} \infty$. Moreover, the test is able to detect alternatives converging to the null hypothesis with a rate $1/\sqrt{n}$.

We conclude this section by presenting a corresponding result for the case of a fixed design. For this purpose we consider a triangular array of random variables

$$Y_{i,n} = m(x_{i,n}) + \sigma(x_{i,n})\epsilon_{i,n}, \quad i = 1, \dots, n, \tag{3.8}$$

where $E[\epsilon_{i,n}] = 0$ and $E[\epsilon_{i,n}^2] = 1$. In the model (3.8) $x_{i,n}, \dots, x_{n,n}$ denote fixed design points defined by

$$\frac{i - 0.5}{n} = \int_0^{x_{i,n}} f(t)dt = F(x_{i,n}), \tag{3.9}$$

where F is a distribution function with a positive density f on the interval $[0, 1]$ which is Hölder continuous of order $\gamma > 1/2$ [see Sacks and Ylvisaker (1966)]. The definition of the process $H_n(t)$ is given in (2.1), where the random variables X_i are replaced by $x_{i,n}$, $(i = 1, \dots, n)$. In this case we also obtain weak convergence of the process $\{\sqrt{n}(H_n(t) - H_n^*(t))\}_{t\in[0,1]}$ but with a different limiting process. The proof is similar to the proof of Theorem 3.1 and some details are indicated in the Appendix (see Section 6).

Theorem 3.2 *Consider the model* (3.8). *If the assumptions (A1)-(A7) and additionally the condition*

(A8) The functions $f, \omega, K, \tilde{K}$ are Hölder continuous of order $\gamma > 1/2$,

are satisfied (with the obvious modifications for a fixed design case), then we have as $n \to \infty$

$$\{\sqrt{n}(H_n(t) - H_n^*(t))\}_{t\in[0,1]} \Rightarrow \{G(t)\}_{t\in[0,1]}$$

in $D[0, 1]$, where $\{G(t)\}_{t\in[0,1]}$ is a Gaussian process with covariance kernel

$$\bar{k}(t, s) = \int_0^{s\wedge t} f(u)\omega^2(u)du - \int_0^s f(u)\omega(u)du \int_0^t f(u)\omega(u)du + 2\nu(t, s) + \mu(t, s).$$

Note that if the nullhypothesis (1.2) is satisfied this yields the same limiting process as in the random design case, as the additional term $\kappa(t,s)$ which appears in the covariance kernel $k(t,s)$ of Theorem 3.1 vanishes under the null hypothesis.

4 Finite sample properties

In this section we will study the finite sample properties of the Cramér-von-Mises test (3.7) and will also compare the new test with the method which has recently been proposed by Dette and Wieczorek (2009). Following these authors we have considered the model

$$m(t) = c(1+0.1t)\ , \quad \sigma(t) = (1+0.1t+\gamma\sqrt{t}),$$

with $c = 0.5, 1, 1.5$ and $\gamma = 0, 1, 2$, where the case $\gamma = 0$ corresponds to the null hypothesis (1.2) of a constant coefficient of variation. The variables X_i are independently uniformly distributed on the interval $[0,1]$ while the errors ϵ_i have a standard normal distribution. For the smoothing parameter h we used $h = n^{-0.9}$ and $h = n^{-0.5}$, while the smoothing parameter in the local linear estimates was chosen by least squares cross validation. As kernels K and $\tilde{K}$ we used the Epanechnikov kernel and the weight function is given by $\omega \equiv 1$. To calculate the critical values we estimated the squared coefficient of variation by the least squares estimate

$$\hat{c}^2 = \Big(\sum_{i=1}^n (\hat{m}(X_i))^2 (Y_i - \hat{m}(X_i))^2 \omega(X_i)\Big) / \Big(\sum_{i=1}^n (\hat{\sigma}^2(X_i))^2 \omega(X_i)\Big)$$

[see Dette and Wieczorek (2009)]. The third and fourth moments of the error variables were estimated according to Dette and Munk (1998) by

$$\begin{aligned}
\hat{m}_3 &= \hat{A}_{4,n} \cdot (\hat{S}_{4,n})^{-1} - 3, \\
\hat{m}_4 &= \Big(\hat{A}_{5,n} \cdot (\hat{S}_{6,n})^{-1}\Big)^{1/2}
\end{aligned}$$

with

$$\begin{aligned}
\hat{A}_{4,n} &= \frac{1}{2(n-1)} \sum_{j=2}^{n} R_j^4, \\
\hat{A}_{5,n} &= \frac{1}{36(n-5)} \sum_{j=3}^{n-3} (R_j - R_{j-1})^3 (R_{j+3} - R_{j+2})^3, \\
\hat{S}_{2k,n} &= \frac{1}{2^k(n-2k+1)} \sum_{j=2}^{n-2k+2} R_j^2 \dots R_{j+2k-2}^2, \quad k = 2, 3, \\
R_j &= Y_j^* - Y_{j-1}^*,
\end{aligned}$$

where Y_j^* is the observation belonging to $X_{(j)}$ and $X_{(1)} \leq \ldots \leq X_{(n)}$ denotes the order statistic of $X_1, \ldots, X_n$. In Table 1 we show the rejection probability of the Cramér-von-Mises test (3.7) for the sample sizes $n = 50, 100, 200$ on the basis of 1000 simulation runs. The results in brackets were calculated with bandwidth $h = n^{-0.5}$ and for the other results we used $h = n^{-0.9}$. A comparison of these results shows that both bandwidth choices yield similar rejection probabilities. This indicates that the procedure is not very sensitive to the choice of the bandwidth h.

The first part of Table 1 ($\gamma = 0$) shows the approximation of the nominal level of the test (3.7), which is rather good for sample sizes $n = 100$ and $n = 200$, while for the smaller sample size $n = 50$ the nominal level is slightly overestimated. The results for the power of the test are shown in the remaining part of Table 1 ($\gamma = 1, 2$). Although the deviation from the multiplicative structure is extremely small for $\gamma = 1$, because the predictor varies in the interval $[0, 1]$, we observe good results for all three coefficients c we studied. A visible increase in power is recognizable for increasing sample sizes n and increasing γ.

Comparing the results depicted in Table 1 with those of the bootstrap test presented by Dette and Wieczorek (2009), we see that both testing procedures yield similar results for the approximation of the nominal level. For the alternatives, however, the test developed in this paper yields higher relative rejection probabilities in almost all cases under consideration. Only for sample size $n = 200$ and coefficient of variation $c = 0.5$ the test proposed by Dette and Wieczorek (2009) has a better power.

Acknowledgements. The authors are grateful to Martina Stein who typed numerous versions of this paper with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG).

5 Proof of Theorem 3.1

In Section 7 we will show that the stochastic expansion

$$H_n(t) = \tilde{H}_n(t) + \bar{H}_n(t) + o_p(n^{-1/2}) \tag{5.1}$$

holds uniformly with respect to $t \in [0, 1]$, where the processes $\{\tilde{H}_n(t)\}_{t\in[0,1]}$ and $\{\bar{H}_n(t)\}_{t\in[0,1]}$ are defined by

$$\tilde{H}_n(t) = \frac{n-1}{n^2} \sum_{j=1}^{n} c_t(X_j) \frac{Y_j}{\sigma(X_j)} - \frac{1}{n^2} \sum_{j=1}^{n} c_t(X_j) \sum_{i \neq j} \frac{Y_i}{\sigma(X_i)}$$

γ	c	n = 50				n = 100				n = 200			
		2.5%	5%	10%	20%	2.5%	5%	10%	20%	2.5%	5%	10%	20%
	0.5	.035	.065	.113	.220	.027	.046	.096	.198	.027	.045	.105	.205
		(.038)	(.067)	(.123)	(.227)	(.035)	(.055)	(.092)	(.176)	(.025)	(.046)	(.096)	(.204)
0	1	.039	.070	.110	.203	.026	.046	.093	.192	.027	.052	.102	.198
		(.040)	(.059)	(.092)	(.181)	(.026)	(.049)	(.093)	(.195)	(.029)	(.053)	(.089)	(.184)
	1.5	.042	.061	.099	.195	.031	.051	.096	.188	.024	.049	.098	.194
		(.033)	(.050)	(.079)	(.183)	(.021)	(.041)	(.096)	(.196)	(.026)	(.044)	(.092)	(.194)
	0.5	.060	.095	.174	.297	.050	.082	.154	.260	.072	.109	.167	.290
		(.054)	(.100)	(.149)	(.245)	(.046)	(.088)	(.144)	(.253)	(.064)	(.099)	(.171)	(.297)
1	1	.088	.123	.197	.289	.105	.151	.216	.329	.131	.187	.264	.423
		(.085)	(.123)	(.193)	(.290)	(.096)	(.147)	(.224)	(.328)	(.135)	(.201)	(.282)	(.393)
	1.5	.122	.177	.236	.326	.143	.216	.292	.398	.241	.319	.423	.559
		(.107)	(.152)	(.209)	(.301)	(.135)	(.191)	(.281)	(.360)	(.209)	(.297)	(.393)	(.532)
	0.5	.067	.109	.176	.296	.074	.120	.172	.302	.087	.139	.192	.318
		(.067)	(.092)	(.137)	(.253)	(.066)	(.113)	(.164)	(.331)	(.075)	(.194)	(.181)	(.292)
2	1	.128	.173	.226	.323	.129	.183	.235	.395	.182	.254	.363	.504
		(.106)	(.149)	(.204)	(.305)	(.119)	(.174)	(.237)	(.381)	(.159)	(.222)	(.330)	(.463)
	1.5	.130	.205	.269	.382	.213	.287	.367	.495	.338	.424	.537	.668
		(.138)	(.186)	(.249)	(.342)	(.210)	(.277)	(.353)	(.469)	(.319)	(.424)	(.528)	(.640)

Table 1: *Rejection probabilities of the Cramér-von-Mises test (3.7).*

and

$$
\begin{aligned}
\bar{H}_n(t) &= -\frac{n-1}{2n^3}\sum_{j=1}^{n}\sum_{k\neq j}^{n} c_t(X_j)h(X_j)\psi(X_k,X_j)(\epsilon_k^2-1) \\
&\quad +\frac{1}{2n^3}\sum_{j=1}^{n}\sum_{i\neq j}^{n}\sum_{k\neq i,j}^{n} c_t(X_j)h(X_i)\psi(X_k,X_i)(\epsilon_k^2-1),
\end{aligned}
$$

with

$$
\psi(x,y)=\frac{\sigma^2(x)\tilde{K}_g(x-y)}{\sigma^2(y)f(y)}.
$$

It is easy to see that

$$
E[\tilde{H}_n(t)] = H_n^*(t) \text{ and } E[\bar{H}_n(t)] = 0,
$$

where $H_n^*(t)$ is defined in (2.3). Consequently it is sufficient to establish the weak convergence

$$
\{\sqrt{n}(\tilde{H}_n(t)-H_n^*(t)+\bar{H}_n(t))\}_{t\in[0,1]} \Rightarrow \{A(t)\}_{t\in[0,1]}
$$

in $D[0,1]$. For the calculation of the asymptotic covariances we introduce the notation

$$
\begin{aligned}
A_\epsilon(t) &= \frac{1}{n\sqrt{n}}\sum_{i,j=1}^{n} c_t(X_j)\left(\epsilon_j-\epsilon_i\right), \\
A_{m,\sigma}(t) &= \frac{1}{n\sqrt{n}}\sum_{i,j=1}^{n} c_t(X_j)\left(h(X_j)-h(X_i)\right)
\end{aligned}
$$

and obtain by straightforward algebra the decomposition

$$
C_n(t) = \sqrt{n}(\tilde{H}_n(t)-H_n^*(t)+\bar{H}_n(t)) = A_\epsilon(t)+A_{m,\sigma}(t)-\sqrt{n}H_n^*(t)+\sqrt{n}\bar{H}_n(t). \quad (5.2)
$$

In the following discussion we will determine the covariance of $C_n(s)$ and $C_n(t)$. Straightforward calculations yield

$$
\begin{aligned}
\operatorname{Cov}(A_\epsilon(s),A_\epsilon(t)) &= E[A_\epsilon(s)A_\epsilon(t)] = \frac{n-1}{n}\big(E[c_s(X_1)c_t(X_1)]-E[c_s(X_1)]E[c_t(X_1)]\big) \\
&= \Big(\int_0^{s\wedge t}\omega^2(u)f(u)du-\int_0^s\omega(u)f(u)du\int_0^t\omega(u)f(u)du\Big)(1+o(1)), \\
\operatorname{Cov}(A_{m,\sigma}(s),A_{m,\sigma}(t)) &= \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k=1}^{n}\sum_{l\neq k}^{n}\Big\{\operatorname{Cov}(c_t(X_i)h(X_i),c_s(X_k)h(X_k)) \\
&\quad +\operatorname{Cov}(c_t(X_i)h(X_j),c_s(X_k)h(X_l))-\operatorname{Cov}(c_t(X_i)h(X_j),c_s(X_k)h(X_k))
\end{aligned}
$$

$$
\begin{aligned}
& -\mathrm{Cov}(c_t(X_i)h(X_i), c_s(X_k)h(X_l))\Big\} \\
= & \Big\{\mathrm{Cov}(c_t(X_1)h(X_1), c_s(X_1)h(X_1)) + \mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_1)h(X_3)) \\
& +\mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_2)h(X_3)) + \mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_3)h(X_1)) \\
& +\mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_3)h(X_2)) - \mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_1)h(X_1)) \\
& -\mathrm{Cov}(c_t(X_1)h(X_2), c_s(X_2)h(X_2)) - \mathrm{Cov}(c_t(X_1)h(X_1), c_s(X_1)h(X_2)) \\
& -\mathrm{Cov}(c_t(X_1)h(X_1), c_s(X_2)h(X_1))\Big\}(1+o(1)) \\
= & \kappa(s,t)(1+o(1)),
\end{aligned}
$$

where the function $\kappa(s,t)$ is defined in (3.2). Note that $\kappa(s,t) \equiv 0$ if the hypothesis (1.2) is satisfied. Finally we obtain for the remaining covariances

$$
\begin{aligned}
\mathrm{Cov}(A_\epsilon(t), \sqrt{n}\bar{H}_n(s)) = & -\frac{1}{2}E[c_t(X_1)c_s(X_2)h(X_2)\psi(X_1,X_2)\epsilon_1(\epsilon_1^2-1)] \\
& +\frac{1}{2}E[c_t(X_1)c_s(X_3)h(X_3)\psi(X_2,X_3)\epsilon_2(\epsilon_2^2-1)] \\
& +\frac{1}{2}E[c_t(X_1)c_s(X_2)h(X_3)\psi(X_1,X_3)\epsilon_1(\epsilon_1^2-1)] \\
& -\frac{1}{2}E[c_t(X_1)c_s(X_3)h(X_4)\psi(X_2,X_4)\epsilon_2(\epsilon_2^2-1)] + O(n^{-1}) \\
= & \Big(-\frac{1}{2}\int_0^{s\wedge t}\omega^2(u)f(u)h(u)m_3(u)du \\
& +\frac{1}{2}\int_0^t \omega(u)f(u)du \int_0^s \omega(u)f(u)h(u)m_3(u)du \\
& +\frac{1}{2}\int_0^s \omega(u)f(u)du \int_0^t \omega(u)f(u)h(u)m_3(u)du \\
& -\frac{1}{2}\int_0^s \omega(u)f(u)du \int_0^t \omega(u)f(u)du \int_0^1 f(u)h(u)m_3(u)du\Big)(1+o(1)) \\
= & \nu(s,t)(1+o(1)), \\
n\mathrm{Cov}(\bar{H}_n(t), \bar{H}_n(s)) = & \frac{1}{4}E[c_t(X_1)h(X_1)\psi(X_3,X_1)(\epsilon_3^2-1)^2 c_s(X_2)h(X_2)\psi(X_3,X_2)] \\
& -\frac{1}{4}E[c_t(X_1)h(X_1)\psi(X_4,X_1)(\epsilon_4^2-1)^2 c_s(X_2)h(X_3)\psi(X_4,X_3)] \\
& -\frac{1}{4}E[c_t(X_1)h(X_3)\psi(X_4,X_3)(\epsilon_4^2-1)^2 c_s(X_2)h(X_2)\psi(X_4,X_2)] \\
& +\frac{1}{4}E[c_t(X_1)h(X_3)\psi(X_5,X_3)(\epsilon_5^2-1)^2 c_s(X_2)h(X_4)\psi(X_5,X_4)] + O((ng)^{-1}) \\
= & \Big(\frac{1}{4}\int_0^{s\wedge t}\omega^2(u)f(u)h^2(u)(m_4(u)-1)du \\
& -\frac{1}{4}\int_0^t \omega(u)f(u)du \int_0^s \omega(u)f(u)h^2(u)(m_4(u)-1)du
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{4}\int_0^s \omega(u)f(u)du \int_0^t \omega(u)f(u)h^2(u)(m_4(u)-1)du \\
& +\frac{1}{4}\int_0^s \omega(u)f(u)du \int_0^t \omega(u)f(u)du \int_0^1 f(u)h^2(u)(m_4(u)-1)du\Big)(1+o(1)) \\
= \; & \mu(s,t)(1+o(1)).
\end{aligned}
$$

All other terms appearing in $\text{Cov}(C_n(t), C_n(s))$ vanish. Combining these results yields for the covariance kernel of the process $C_n(t)$ defined in (5.2)

$$
\text{Cov}(C_n(t), C_n(s)) = k(t,s)(1+o(1)),
$$

where the kernel k is defined in (3.1). In order to prove weak convergence of the finite dimensional distributions

$$
(C_n(t_1), \dots, C_n(t_k))^T \xrightarrow{D} (A(t_1), \dots, A(t_k))^T, \tag{5.3}
$$

we restrict ourselves to the case $k = 2$. For this purpose we use Theorem 1 in de Jong (1996), the Cramér-Wold device and introduce for $a_1, a_2 \in \mathbb{R}, \; t_1, t_2 \in [0,1]$ the notation

$$
\begin{aligned}
b(X_i) &= a_1 c_{t_1}(X_i) + a_2 c_{t_2}(X_i), \\
Z_{1,i} &= b(X_i)\frac{Y_i}{\sigma(X_i)} - E\left[b(X_i)\frac{Y_i}{\sigma(X_i)}\right], \\
Z_{2,i} &= b(X_i) - E[b(X_i)], \\
Z_{3,i} &= \frac{Y_i}{\sigma(X_i)} - E\left[\frac{Y_i}{\sigma(X_i)}\right], \\
Z_{4,i,j} &= \frac{1}{2}b(X_i)h(X_i)\psi(X_j,X_i) - E\left[\frac{1}{2}b(X_i)h(X_i)\psi(X_j,X_i) \mid X_j, \epsilon_j\right], \\
Z_{5,i} &= \epsilon_i^2 - 1, \\
Z_{6,i,j} &= h(X_i)\psi(X_j,X_i) - E\left[h(X_i)\psi(X_j,X_i) \mid X_j, \epsilon_j\right], \\
N_{1,j} &= \frac{n-2}{n}E\left[h(X_i)\psi(X_j,X_i) \mid X_j, \epsilon_j\right], \; i \neq j, \\
N_{2,j} &= \frac{n-1}{n}E\left[\frac{1}{2}b(X_i)h(X_i)\psi(X_j,X_i) \mid X_j, \epsilon_j\right], \; i \neq j.
\end{aligned}
$$

We consider for $v = (a_1, a_2)^T \in \mathbb{R}^2$ the random variable

$$
V(n) = v^T(C_n(t_1), C_n(t_2))^T / \tau_{t_1,t_2} = \sum_{I \subset \{1,\dots,n\}, |I| \le 3} W_I, \tag{5.4}
$$

where $\tau^2_{t_1,t_2} = \lim_{n\to\infty} v^T \text{Cov}(C_n(t_1), C_n(t_2))v$ denotes the asymptotic variance of the random variable $a_1 C_n(t_1) + a_2 C_n(t_2)$ and the last identity follows by a straightforward calculation using the notation

$$W_I = \begin{cases} \frac{n-1}{\tau_{t_1,t_2} n^{3/2}} \Big(Z_{1,i} - Z_{2,i} E[h(X_1)] - Z_{3,i} E[b(X_1)] - Z_{5,i} N_{2,i} \\ + \frac{1}{2} Z_{5,i} N_{1,i} E[b(X_1)] \Big), & I = \{i\} \\ \frac{-1}{\tau_{t_1,t_2} n^{3/2}} \Big(Z_{2,i} Z_{3,j} + Z_{2,j} Z_{3,i} + \frac{n-1}{n} (Z_{4,i,j} Z_{5,j} + Z_{4,j,i} Z_{5,i}) \\ - \frac{n-2}{2n} E[b(X_1)] (Z_{6,i,j} Z_{5,j} + Z_{6,j,i} Z_{5,i}) - \frac{1}{2} (Z_{2,i} Z_{5,j} N_{1,j} + Z_{2,j} Z_{5,i} N_{1,i}) \Big), & I = \{i,j\} \\ \frac{1}{2\tau_{t_1,t_2} n^{5/2}} \Big(Z_{2,i} Z_{6,j,k} Z_{5,k} + Z_{2,i} Z_{6,k,j} Z_{5,j} + Z_{2,j} Z_{6,i,k} Z_{5,k} + Z_{2,j} Z_{6,k,i} Z_{5,i} \\ + Z_{2,k} Z_{6,i,j} Z_{5,j} + Z_{2,k} Z_{6,j,i} Z_{5,i} \Big), & I = \{i,j,k\}. \end{cases}$$

For a set I let $\mathcal{F}_I := \sigma\{(X_i, \epsilon_i), i \in I\}$ denote the sigma field generated by $\{(X_i, \epsilon_i) | \ i \in I\}$, then a straightforward calculation shows $E[W_I \mid \mathcal{F}_J] = 0$ whenever $I \not\subseteq J$. If the index set I contains only one element, say $I = \{i\}$, we have

$$\begin{aligned} \operatorname{Var}(W_{\{i\}}) &= \frac{(n-1)^2}{\tau^2_{t_1,t_2} n^3} E\left[Z^2_{1,i} + Z^2_{2,i} E^2\left[h(X_1)\right] + Z^2_{3,i} E^2[b(X_1)] + Z^2_{5,i} N^2_{2,i} \right. \\ &\quad + \frac{1}{4} Z^2_{5,i} N^2_{1,i} E^2[b(X_1)] - 2 Z_{1,i} Z_{2,i} E\left[h(X_1)\right] - 2 Z_{1,i} Z_{3,i} E[b(X_1)] \\ &\quad - 2 Z_{1,i} Z_{5,i} N_{2,i} + Z_{1,i} Z_{5,i} N_{1,i} E[b(X_1)] + 2 Z_{2,i} Z_{3,i} E\left[h(X_1)\right] E[b(X_1)] \\ &\quad + 2 Z_{2,i} Z_{5,i} N_{2,i} E[h(X_1)] - Z_{2,i} Z_{5,i} N_{1,i} E[h(X_1)] E[b(X_1)] \\ &\quad + 2 Z_{3,i} Z_{5,i} N_{2,i} E[b(X_1)] - Z_{3,i} Z_{5,i} N_{1,i} E^2[b(X_1)] \\ &\quad \left. - Z^2_{5,i} N_{2,i} N_{1,i} E[b(X_1)] \right] \\ &= O(n^{-1}). \end{aligned}$$

Similarly, if $I = \{i, j\}$ contains two elements it follows

$$\begin{aligned} \operatorname{Var}(W_{\{i,j\}}) &= \frac{2}{\tau^2_{t_1,t_2} n^3} E\left[Z^2_{2,i} Z^2_{3,j} + \frac{(n-1)^2}{n^2} Z^2_{4,i,j} Z^2_{5,j} + \frac{(n-2)^2}{4n^2} Z^2_{6,j,i} Z^2_{5,i} E^2[b(X_1)] \right. \\ &\quad + \frac{1}{4} Z^2_{2,i} Z^2_{5,j} N^2_{1,j} + 2 Z_{2,i} Z_{3,i} Z_{2,j} Z_{3,j} - \frac{(n-1)(n-2)}{n^2} Z_{4,i,j} Z_{6,i,j} Z^2_{5,j} E[b(X_1)] \\ &\quad - \frac{n-1}{n} Z_{4,i,j} Z_{2,i} Z^2_{5,j} N_{1,j} + \frac{n-2}{2n} Z_{6,i,j} Z_{2,i} Z^2_{5,j} N_{1,j} E[b(X_1)] \\ &\quad + \frac{n-2}{2n} Z_{6,j,i} Z_{2,j} Z^2_{5,i} N_{1,i} E[b(X_1)] + 2 \frac{n-1}{n} Z_{2,i} Z_{3,j} Z_{4,i,j} Z_{5,j} \\ &\quad \left. - \frac{n-2}{n} Z_{2,i} Z_{3,j} Z_{6,i,j} Z_{5,j} E[b(X_1)] - Z^2_{2,i} Z_{3,j} Z_{5,j} N_{1,j} \right] \\ &= O(n^{-3}), \end{aligned}$$

and if $I = \{i, j, k\}$ contains three elements we obtain

$$\operatorname{Var}(W_{\{i,j,k\}}) = \frac{3}{2\tau^2_{t_1,t_2} n^5} E\left[Z^2_{2,i} Z^2_{6,j,k} Z^2_{5,k} + Z_{2,i} Z_{6,j,k} Z_{2,j} Z_{6,i,k} Z^2_{5,k} \right] = O(n^{-5}).$$

This yields

$$\max_{1\le i\le n}\Big\{\sum_I \mathrm{Var}(W_I)\mid i\in I\Big\}\to 0$$

as $n\to\infty$ and establishes the first condition of de Jong's (1996) theorem. We will show in Section 7 that the second condition of this theorem

$$\sum_{(I,J,K,L)} \mid E[W_IW_JW_KW_L]\mid\to 0 \tag{5.5}$$

is also satisfied, where the summation is performed over all index sets I,J,K,L of $\{1,\dots,n\}$ which are connected. By Theorem 1 of de Jong (1996) we therefore obtain for the statistic $V(n)$ defined in (5.4) weak convergence, i.e.

$$V(n)\xrightarrow{D}\mathcal{N}(0,1)$$

as $n\to\infty$, and the Cramér-Wold device yields assertion (5.3).

The proof of Theorem 3.1 will be completed by showing tightness of the process $\{C_n(t)\}_{t\in[0,1]}$. For this property we use Theorem 13.5 in Billingsley (1999) and prove that the condition

$$E[(C_n(t)-C_n(s))^2(C_n(r)-C_n(t))^2]\le C(r-s)^2 \tag{5.6}$$

holds for all $0\le s<t<r\le 1$ and some positive constant $C>0$. The assertion of Theorem 3.1 then follows, because the processes $\{C_n(t)\}_{t\in[0,1]}$ and $\{\sqrt{n}(H_n(t)-H_n^*(t))\}_{t\in[0,1]}$ have the same asymptotic distribution by (5.1).

For a proof of (5.6) we use the representation

$$C_n(t)-C_n(s) \;=\; \frac{n-1}{n\sqrt{n}}\sum_{j=1}^n Z_{t,s,j}+\frac{1}{\sqrt{n}}\sum_{j=1}^n M_{t,s,j}$$

with

$$\begin{aligned}
Z_{t,s,j} &= L_{t,s}(X_j)-E[L_{t,s}(X_j)]\\
L_{t,s}(X_j) &= c_{t,s}(X_j)\left(\frac{Y_j}{\sigma(X_j)}-\frac{1}{n-1}\sum_{i\neq j}\frac{Y_i}{\sigma(X_i)}\right),\\
M_{t,s,j} &= c_{t,s}(X_j)\Big(\frac{1}{2n^2}\sum_{i\neq j}\sum_{k\neq i,j} h(X_i)\psi(X_k,X_i)(\epsilon_k^2-1)-\frac{n-1}{2n^2}\sum_{k\neq j} h(X_j)\psi(X_k,X_j)(\epsilon_k^2-1)\Big),\\
c_{t,s}(X_j) &= \omega(X_j)\int_s^t K_h(X_j-u)du \qquad 0\le s,t\le 1.
\end{aligned}$$

Note that $E[Z_{t,s,j}] = E[M_{t,s,j}] = 0$ and that the random variables $Z_{t,s,j}$ and $M_{t,s,j}$ $(j = 1, \ldots, n)$ are not independent. For $0 \le s < t < r \le 1$ we have

$$\begin{aligned} & E\left[(C_n(t) - C_n(s))^2 (C_n(r) - C_n(t))^2\right] \qquad (5.7) \\ = \; & \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \sum_{l=1}^{n} E\,[Z_{t,s,i} Z_{t,s,j} Z_{r,t,k} Z_{r,t,l} + 2 Z_{t,s,i} Z_{t,s,j} Z_{r,t,k} M_{r,t,l} + Z_{t,s,i} Z_{t,s,j} M_{r,t,k} M_{r,t,l} \\ & + 2 Z_{t,s,i} M_{t,s,j} Z_{r,t,k} Z_{r,t,l} + 4 Z_{t,s,i} M_{t,s,j} Z_{r,t,k} M_{r,t,l} + 2 Z_{t,s,i} M_{t,s,j} M_{r,t,k} M_{r,t,l} \\ & + M_{t,s,i} M_{t,s,j} Z_{r,t,k} Z_{r,t,l} + 2 M_{t,s,i} M_{t,s,j} Z_{r,t,k} M_{r,t,l} + M_{t,s,i} M_{t,s,j} M_{r,t,k} M_{r,t,l}]\,. \end{aligned}$$

A straightforward but tedious calculation shows that the expression above is a sum over $O(n^2)$ nonzero summands. The nonzero summands are products of expectations of the form

$$E\left[c_{t,s}^i(X_1) c_{r,t}^k(X_1) (h(X_1))^j (\psi(X_1, X_2))^l \epsilon_2^q\right] \quad (i, k = 0, 1, 2,\; j, l = 0, 1, 2, 3, 4,\; q = 2, \ldots, 8)$$

with bounded coefficients. Moreover, both terms $c_{t,s}(X_1)$ and $c_{r,t}(X_1)$ appear in each product exactly two times but not necessarily in the same expectation. Typical products are of the form

$$E\Big[c_{t,s}^2(X_1) c_{r,t}^2(X_1) h(X_1)\Big] \quad \text{or} \quad E\Big[c_{t,s}^2(X_1) h(X_1)\Big]\; E\Big[c_{r,t}^2(X_1) \psi(X_1, X_2)\Big].$$

By the continuity of $f(\cdot)$, $\tilde{K}(\cdot)$, $m(\cdot)$ and $\sigma(\cdot)$ all expectations of the form

$$E\left[(h(X_1))^k (\psi(X_1, X_2))^l\right]$$

yield a rate $O(g^{-l+1})$ if $l = 2, 3, 4$ or $O(1)$ if $l = 0, 1$. In the cases where $l = 2, 3, 4$ we obtain by the equality of indices an additional factor $1/n^{l-1}$. We now discuss the expectations of different types separately. If $i = 1$, $k = 0$ it follows (observing that the kernel K has compact support)

$$\left| E\Big[c_{t,s}(X_1) (h(X_1))^j (\psi(X_1, X_2))^l\Big] \right| \le C(t - s) \le C(r - s)$$

for some positive constant C $(j, l = 0, \ldots, 4)$. If $i = 2$ and $k = 0$ we have

$$\begin{aligned} & E\Big[c_{t,s}^2(X_1) (h(X_1))^j (\psi(X_1, X_2))^l\Big] \\ = \; & \int_0^1 \int_s^t \int_{-u/h}^{(1-u)/h} f(x) f(zh + u) \omega^2(zh + u) (h(zh + u))^j (\psi(zh + u, x))^l K(z) \\ & \int_{z+(u-t)/h}^{z+(u-s)/h} K(w)\, dw\, dz\, du\, dx \\ & \le C(t - s) \le C(r - s). \end{aligned}$$

Similar calculations yield for the cases $i = 0$, $k = 2$; $i = k = 1$ and $i = 2$, $k = 1$

$$\left| E\Big[c_{t,s}^i(X_1) c_{r,t}^k(X_1) (h(X_1))^j (\psi(X_1, X_2))^l\Big] \right| \le C(t - s) \le C(r - s).$$

In the previous cases there always exists a second factor in the corresponding product, which yields the estimate $C(r-s)^2$. The case $i = k = 2$ requires a different argument, because there is no additional second expectation which yields a further factor $(r-s)$. However, in this case at least four indices are the same, so that there appears an additional factor $1/n$ in the corresponding terms and we obtain

$$
\begin{aligned}
& \frac{1}{n}\left|E\left[c_{t,s}^2(X_1)c_{r,t}^2(X_1)(h(X_1))^j(\psi(X_1,X_2))^l\right]\right| \\
= \; & \frac{1}{n}\left|\int_0^1\int_0^1\left(\int_s^t K_h(x_1-u)du\right)^2\left(\int_t^r K_h(x_1-u)du\right)^2 w(x_1)^4h(x_1)^j\psi(x_1,x_2)^l f(x_1)f(x_2)dx_1dx_2\right| \\
\le \; & \frac{1}{n}\int_0^1\int_0^1\left(\int_{(x_1-s)/h}^{(x_1-t)/h}K(v)dv\right)^2\left|\left(\int_{(x_1-r)/h}^{(x_1-t)/h}K(v)dv\right)\left(\int_t^r K_h(x_1-u)du\right)w(x_1)^4h(x_1)^j\psi(x_1,x_2)^l\right| \\
& \times f(x_1)f(x_2)dx_1dx_2 \\
\le \; & \frac{\tilde{C}(r-t)}{nh}\int_0^1\int_t^r\int_{-u/h}^{(1-u)/h}\left|K(x)w(u+hx)^4h(u+hx)^j\psi(u+hx,x_2)^l\right| f(u+hx)f(x_2)dxdudx_2 \\
\le \; & C(r-t)^2 \le C(r-s)^2.
\end{aligned}
$$

Therefore every summand in (5.7) can be bounded by $C(r-s)^2$, which proves tightness of the process $\{C_n(t)\}_{t\in[0,1]}$ and completes the proof of Theorem 3.1.

6 Appendix: proof of Theorem 3.2

Note that similar arguments as given in the proof of Theorem 3.1 yield a stochastic expansion

$$
\begin{aligned}
& \sqrt{n}(H_n(t)-H_n^*(t)) = A_\epsilon(t)+\sqrt{n}\bar{H}_n(t)+o_p(1) \\
= \; & \sqrt{n}\Big(\frac{n-1}{n^2}\sum_{j=1}^n c_t(x_{j,n})\Big(\epsilon_{j,n}-\frac{1}{n-1}\sum_{i\neq j}\epsilon_{i,n}\Big)\Big) \\
& +\sqrt{n}\Big(\frac{1}{n}\sum_{j=1}^n c_t(x_{j,n})\Big(\frac{1}{2n^2}\sum_{i\neq j}^n\sum_{k\neq i,j}^n h(x_{i,n})\psi(x_{k,n},x_{i,n})(\epsilon_{k,n}^2-1) \\
& -\frac{n-1}{2n^2}\sum_{k\neq j}^n h(x_{j,n})\psi(x_{k,n},x_{j,n})(\epsilon_{k,n}^2-1)\Big)\Big)+o_p(1),
\end{aligned}
$$

uniformly with respect to $t\in[0,1]$. Therefore it follows for the covariance of $C_n(s)$ and $C_n(t)$

$$
\begin{aligned}
\mathrm{Cov}(C_n(s),C_n(t)) \; = \; & \mathrm{Cov}(A_\epsilon(s),A_\epsilon(t))+\mathrm{Cov}(A_\epsilon(s),\sqrt{n}\bar{H}_n(t)) \\
& +\mathrm{Cov}(\sqrt{n}\bar{H}_n(s),A_\epsilon(t))+n\mathrm{Cov}(\bar{H}_n(s),\bar{H}_n(t)),
\end{aligned}
$$

and similar calculations as given for the proof of Theorem 3.1 yield

$$
\begin{aligned}
\mathrm{Cov}(A_\epsilon(s), A_\epsilon(t)) &= \frac{1}{n}\sum_{j=1}^{n} c_s(x_{j,n})c_t(x_{j,n}) - \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}^{n} c_s(x_{j,n})c_t(x_{k,n}) + o(1)\\
&= \int_0^{s\wedge t} f(u)\omega^2(u)du - \int_0^{s} f(u)\omega(u)du \int_0^{t} f(u)\omega(u)du + o(1),\\
\mathrm{Cov}(\sqrt{n}\bar{H}_n(s), A_\epsilon(t)) &= \frac{-1}{2n^2}\sum_{j=1}^{n}\sum_{i\neq j}^{n} c_t(x_{j,n})c_s(x_{i,n})h(x_{i,n})\psi(x_{j,n}, x_{i,n})E[\epsilon_{j,n}^3]\\
&\quad +\frac{1}{2n^3}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i\neq j}^{n} c_t(x_{j,n})c_s(x_{k,n})h(x_{i,n})\psi(x_{j,n}, x_{i,n})E[\epsilon_{j,n}^3]\\
&\quad +\frac{1}{2n^3}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k\neq i}^{n} c_t(x_{j,n})c_s(x_{k,n})h(x_{k,n})\psi(x_{i,n}, x_{k,n})E[\epsilon_{i,n}^3]\\
&\quad -\frac{1}{2n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\sum_{k\neq i}^{n}\sum_{l\neq i}^{n} c_t(x_{j,n})c_s(x_{k,n})h(x_{l,n})\psi(x_{i,n}, x_{l,n})E[\epsilon_{i,n}^3]\\
&= \nu(t,s) + o(1),\\
n\mathrm{Cov}(\bar{H}_n(s), \bar{H}_n(t)) &= \frac{1}{4n^3}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i\neq k}^{n} c_t(x_{j,n})c_s(x_{i,n})h(x_{j,n})h(x_{i,n})\\
&\qquad \times \psi(x_{k,n}, x_{j,n})\psi(x_{k,n}, x_{i,n})(E[\epsilon_{k,n}^4]-1)\\
&\quad -\frac{1}{4n^4}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{l,i\neq k}^{n} c_t(x_{j,n})c_s(x_{l,n})h(x_{j,n})h(x_{i,n})\psi(x_{k,n}, x_{j,n})\\
&\qquad \times \psi(x_{k,n}, x_{i,n})(E[\epsilon_{k,n}^4]-1)\\
&\quad -\frac{1}{4n^4}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{l,i\neq k}^{n} c_t(x_{l,n})c_s(x_{j,n})h(x_{j,n})h(x_{i,n})\psi(x_{k,n}, x_{j,n})\\
&\qquad \times \psi(x_{k,n}, x_{i,n})(E[\epsilon_{k,n}^4]-1)\\
&\quad +\frac{1}{4n^5}\sum_{j,i=1}^{n}\sum_{k\neq i,j}^{n}\sum_{l,q\neq k}^{n} c_t(x_{j,n})c_s(x_{l,n})h(x_{i,n})h(x_{q,n})\psi(x_{k,n}, x_{i,n})\\
&\qquad \times \psi(x_{k,n}, x_{q,n})(E[\epsilon_{k,n}^4]-1)\\
&= \mu(t,s) + o(1),
\end{aligned}
$$

where we have used (3.9) in the last steps, and ν and μ are defined in (3.3) and (3.4), respectively. The proof now follows by similar arguments as given in Section 5, which are omitted for the sake of brevity.

7 Appendix: auxiliary results

7.1 Proof of the stochastic expansion (5.1)

Recall the definition $h(t) = m(t)/\sigma(t)$. A Taylor expansion of the function $1/\sqrt{\hat{\sigma}^2(X_j)}$ yields

$$H_n(t) = \sum_{\ell=1}^{12} H_{n,\ell}(t),$$

where the quantities $H_{n,\ell}(t)$ are defined by

$$\begin{aligned}
H_{n,1}(t) &= \frac{n-1}{n^2}\sum_{j=1}^n c_t(X_j)h(X_j), \quad H_{n,2}(t) = -\frac{1}{n}\sum_{j=1}^n c_t(X_j)\frac{1}{n}\sum_{i\neq j}^n h(X_i) \\
H_{n,3}(t) &= \frac{n-1}{n^2}\sum_{j=1}^n c_t(X_j)\epsilon_j, \quad H_{n,4}(t) = -\frac{1}{n}\sum_{j=1}^n c_t(X_j)\frac{1}{n}\sum_{i\neq j}^n \epsilon_i \\
H_{n,5}(t) &= -\frac{n-1}{2n^2}\sum_{j=1}^n c_t(X_j)\frac{m(X_j)}{\sigma^3(X_j)}\left(\hat{\sigma}^2(X_j)-\sigma^2(X_j)\right) \\
H_{n,6}(t) &= \frac{1}{2n^2}\sum_{j=1}^n c_t(X_j)\sum_{i\neq j}^n \frac{m(X_i)}{\sigma^3(X_i)}\left(\hat{\sigma}^2(X_i)-\sigma^2(X_i)\right) \\
H_{n,7}(t) &= -\frac{n-1}{2n^2}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\left(\hat{\sigma}^2(X_j)-\sigma^2(X_j)\right) \\
H_{n,8}(t) &= \frac{1}{2n^2}\sum_{j=1}^n c_t(X_j)\sum_{i\neq j}^n \frac{\epsilon_i}{\sigma^2(X_i)}\left(\hat{\sigma}^2(X_i)-\sigma^2(X_i)\right) \\
H_{n,9}(t) &= \frac{3(n-1)}{8n^2}\sum_{j=1}^n c_t(X_j)\frac{m(X_j)}{\xi_j^{5/2}}\left(\hat{\sigma}^2(X_j)-\sigma^2(X_j)\right)^2 \\
H_{n,10}(t) &= -\frac{3}{8n^2}\sum_{j=1}^n c_t(X_j)\sum_{i\neq j}^n \frac{m(X_i)}{\xi_i^{5/2}}\left(\hat{\sigma}^2(X_i)-\sigma^2(X_i)\right)^2 \\
H_{n,11}(t) &= \frac{3(n-1)}{8n^2}\sum_{j=1}^n c_t(X_j)\frac{\sigma(X_j)\epsilon_j}{\xi_j^{5/2}}\left(\hat{\sigma}^2(X_j)-\sigma^2(X_j)\right)^2 \\
H_{n,12}(t) &= -\frac{3}{8n^2}\sum_{j=1}^n c_t(X_j)\sum_{i\neq j}^n \frac{\sigma(X_i)\epsilon_i}{\xi_i^{5/2}}\left(\hat{\sigma}^2(X_i)-\sigma^2(X_i)\right)^2,
\end{aligned}$$

and the random variables ξ_i satisfy $\mid \xi_i - \sigma^2(X_i) \mid \leq \mid \hat{\sigma}^2(X_i) - \sigma^2(X_i) \mid$ ($i = 1, \ldots, n$). In the following we show that all terms $H_{n,i}$, $i = 7, \ldots, 12$, in this stochastic expansion are of order $o_p(n^{-1/2})$, where we restrict ourselves exemplarily to the random variable $H_{n,7}(t)$. The other terms $H_{n,8}, \ldots, H_{n,12}$ can be treated by similar (but sometimes tedious) arguments.

Observing the definition of the local linear estimate of the variance function in (2.2) we have

$$
\begin{aligned}
H_{n,7}(t) &= -\frac{n-1}{2n^3}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\sum_{k=1}^n W_k(X_j)(m(X_k)-\hat{m}(X_k))^2 \\
&\quad -\frac{n-1}{n^3}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\sum_{k=1}^n W_k(X_j)\sigma(X_k)\epsilon_k(m(X_k)-\hat{m}(X_k)) \\
&\quad -\frac{n-1}{2n^2}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\Big(\frac{1}{n}\sum_{k=1}^n W_k(X_j)\sigma^2(X_k)\epsilon_k^2-\sigma^2(X_j)\Big) \\
&= H_{n,71}(t)+H_{n,72}(t)+H_{n,73}(t), \qquad (7.1)
\end{aligned}
$$

where the last identity defines the terms $H_{n,7i}(x)$ in an obvious manner ($i=1,2,3$). From the estimate $\max_{i=1,\dots,n} \mid \hat{m}(X_i)-m(X_i)\mid = O_p((ng)^{-1/2}+g^2)$ [see Yao and Tong (2000)] we get

$$
\begin{aligned}
\mid H_{n,71}(t)\mid &\leq \max_{i=1,\dots,n}(\hat{m}(X_i)-m(X_i))^2\,\frac{1}{n^2}\sum_{j=1}^n\sum_{k=1}^n \mid c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}W_k(X_j)\mid = O_p\left((ng)^{-1}+\frac{g^2}{\sqrt{ng}}+g^4\right) \\
&= o_p\left(n^{-1/2}\right)
\end{aligned}
$$

uniformly with respect to $t\in[0,1]$. Similarly, using the definition of the local linear estimate for the regression function it follows

$$
\begin{aligned}
H_{n,72}(t) &= -\frac{n-1}{n^3}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\sum_{k=1}^n W_k(X_j)\sigma(X_k)\epsilon_k\Big(m(X_k)-\frac{1}{n}\sum_{l=1}^n W_l(X_k)m(X_l)\Big) \\
&\quad +\frac{n-1}{n^4}\sum_{j=1}^n c_t(X_j)\frac{\epsilon_j}{\sigma^2(X_j)}\sum_{k=1}^n W_k(X_j)\sigma(X_k)\epsilon_k\sum_{l=1}^n W_l(X_k)\sigma(X_l)\epsilon_l \\
&= H_{n,721}(t)+H_{n,722}(t)
\end{aligned}
$$

with an obvious definition of $H_{n,72\ell}$ ($\ell=1,2$). For the second moments of these random variables we have

$$
\begin{aligned}
E[H^2_{n,721}(t)] &= \frac{(n-1)^2}{n^6}\sum_{i=1}^n\sum_{j=1}^n E\Big[c_t(X_i)c_t(X_j)\frac{\epsilon_i\epsilon_j}{\sigma^2(X_i)\sigma^2(X_j)}\sum_{k=1}^n\sum_{q=1}^n W_k(X_i)W_q(X_j)\sigma(X_k)\epsilon_k\sigma(X_q)\epsilon_q \\
&\quad \times\Big(m(X_k)-\frac{1}{n}\sum_{l=1}^n W_l(X_k)m(X_l)\Big)\Big(m(X_q)-\frac{1}{n}\sum_{l=1}^n W_l(X_q)m(X_l)\Big)\Big] \\
&= O(n^{-2}g^2)=o(n^{-1}), \\
E[H^2_{n,722}(t)] &= \frac{(n-1)^2}{n^8}\sum_{i=1}^n\sum_{j=1}^n E\Big[c_t(X_i)c_t(X_j)\frac{\epsilon_i\epsilon_j}{\sigma^2(X_i)\sigma^2(X_j)} \\
&\quad \times\sum_{k=1}^n\sum_{q=1}^n W_k(X_i)W_q(X_j)\sigma(X_k)\sigma(X_q)\epsilon_k\epsilon_q
\end{aligned}
$$

$$\times \sum_{l=1}^{n}\sum_{p=1}^{n} W_l(X_k)W_p(X_q)\sigma(X_l)\sigma(X_p)\epsilon_l\epsilon_p\Big]$$
$$= O(n^{-3}g^{-2}) = o\left(n^{-1}\right),$$

uniformly with respect to $t \in [0,1]$, where we used the facts

$$| \, m(X_k) - \frac{1}{n}\sum_{l=1}^{n} W_l(X_k)m(X_l) \, |= O\left(g^2\right),$$

$| \, W_k(X_k) \, |= O\left(g^{-1}\right)$ (almost surely) and $ng^2 \to \infty$, as $n \to \infty$. This yields $H_{n,72}(t) = o_p\left(n^{-1/2}\right)$ uniformly with respect to $t \in [0,1]$. Finally we provide a corresponding estimate for the remaining term $H_{n,73}$ in the decomposition (7.1). For this purpose we consider its second moment

$$\begin{aligned}
E[H^2_{n,73}(t)] &= \frac{(n-1)^2}{4n^3} E\Big[c_t^2(X_1)\frac{\epsilon_1^2}{\sigma^4(X_1)}\Big(\frac{1}{n}\sum_{k=2}^{n} W_k(X_1)\sigma^2(X_k)\epsilon_k^2 - \sigma^2(X_1)\Big)^2\Big] \\
&\quad + \frac{(n-1)^2}{2n^4} E\Big[c_t^2(X_1)\frac{\epsilon_1^4}{\sigma^2(X_1)}W_1(X_1)\Big(\frac{1}{n}\sum_{k=2}^{n} W_k(X_1)\sigma^2(X_k)\epsilon_k^2 - \sigma^2(X_1)\Big)\Big] \\
&\quad + \frac{(n-1)^2}{4n^5} E\Big[c_t^2(X_1)\epsilon_1^6 W_1^2(X_1)\Big] \\
&\quad + \frac{(n-1)^3}{4n^5} E\Big[c_t(X_1)c_t(X_2)\epsilon_1^3\epsilon_2^3 W_1(X_1)W_2(X_2)\Big] \\
&= O\left(n^{-1}g^2 + (ng)^{-2}\right) = o\left(n^{-1}\right),
\end{aligned}$$

using the fact

$$| \, \sigma^2(X_k) - \frac{1}{n}\sum_{l=1}^{n} W_l(X_k)\sigma^2(X_l)\epsilon_l^2 \, |= O\left(g^2\right)$$

(almost surely). This gives $H_{n,73}(t) = o_p\left(n^{-1/2}\right)$ uniformly with respect to $t \in [0,1]$, and shows that the random variable $H_{n,7}(t)$ is of order $o_p(n^{-1/2})$. For the terms $H_{n,5}(t)$ and $H_{n,6}(t)$ we use the decomposition

$$\begin{aligned}
\hat{\sigma}^2(X_j) - \sigma^2(X_j) &= \frac{1}{nf(X_j)}\sum_{i=1}^{n} \tilde{K}_g(X_i - X_j)\{\sigma^2(X_i)(\epsilon_i^2 - 1) - 2\sigma(X_i)\epsilon_i(\hat{m}(X_i) - m(X_i)) \\
&\quad + (\hat{m}(X_i) - m(X_i))^2\} + \frac{g^2k_2}{2}(\sigma^2)''(X_j) + R(X_j)
\end{aligned}$$

where $R(X_j) = O\left(gn^{-1/2}\right)$ (almost surely) and $k_2 = \int u^2\tilde{K}(u)du$ [see Fan and Yao (1998)]. This yields by similar arguments as used for the term $H_{n,7}(t)$

$$H_{n,5}(t) = -\frac{n-1}{2n^3}\sum_{j=1}^{n}\sum_{k\neq j}^{n} c_t(X_j)\frac{m(X_j)\sigma^2(X_k)}{\sigma^3(X_j)}\frac{\tilde{K}_g(X_k - X_j)}{f(X_j)}(\epsilon_k^2 - 1) + o_p\left(n^{-1/2}\right)$$

and

$$H_{n,6}(t) = \frac{1}{2n^3}\sum_{j=1}^{n}\sum_{i\neq j}^{n}\sum_{k\neq i,j}^{n} c_t(X_j)\frac{m(X_i)\sigma^2(X_k)}{\sigma^3(X_i)}\frac{\tilde{K}_g(X_k-X_i)}{f(X_i)}(\epsilon_k^2-1)+o_p\left(n^{-1/2}\right),$$

which completes the proof of the stochastic expansion (5.1).

7.2 Proof of (5.5)

For a proof we essentially have to distinguish 15 different cases of connected subsets of $\{1,\dots,n\}$, where there are no elements contained in only one set among I,J,K,L (otherwise the independence would yield $E[W_IW_JW_KW_L]=0$). We begin with the case where all sets I,J,K,L are singletons, which implies $I=J=K=L=\{i\}$ for some $i\in\{1,\dots,n\}$. This yields

$$\begin{aligned} E\left[W_{\{i\}}^4\right] &= \frac{(n-1)^4}{\tau_{t_1,t_2}^4 n^6}E\Big[\Big(Z_{1,i}-Z_{2,i}E\Big[h(X_1)\Big]-Z_{3,i}E\Big[b(X_1)\Big]-Z_{5,i}N_{2,i}\\ &\quad +\frac{1}{2}Z_{5,i}N_{1,i}E\Big[b(X_1)\Big]\Big)^4\Big]=O\left(n^{-2}\right).\end{aligned}$$

If there is exactly one set, say L, with two elements, the sets are connected if and only if $I,J,K\subset L$. In this case there exist two cases of the type $L=\{i,j\}$ and $I=J=K=\{i\}$ or $L=\{i,j\}$, $I=J=\{i\}$ and $K=\{j\}$. We only consider the last named case (the other one yields $E[W_IW_JW_KW_L]=0$, as the index j is only contained in one of the sets), which gives

$$\begin{aligned} &E\left[W_{\{i\}}^2W_{\{j\}}W_{\{i,j\}}\right]\\ =\;& \frac{-(n-1)^3}{\tau_{t_1,t_2}^4 n^6}E\Big[\Big(Z_{1,i}-Z_{2,i}E[h(X_1)]-Z_{3,i}E[b(X_1)]-Z_{5,i}N_{2,i}+\frac{1}{2}Z_{5,i}N_{1,i}E[b(X_1)]\Big)^2\\ &\times\Big(Z_{1,j}-Z_{2,j}E[h(X_1)]-Z_{3,j}E[b(X_1)]-Z_{5,j}N_{2,j}+\frac{1}{2}Z_{5,j}N_{1,j}E[b(X_1)]\Big)\\ &\times\Big(Z_{2,i}Z_{3,j}+Z_{2,j}Z_{3,i}+\frac{n-1}{n}(Z_{4,i,j}Z_{5,j}+Z_{4,j,i}Z_{5,i})-\frac{n-2}{2n}E[b(X_1)](Z_{6,i,j}Z_{5,j}+Z_{6,j,i}Z_{5,i})\\ &-\frac{1}{2}(Z_{2,i}Z_{5,j}N_{1,j}+Z_{2,j}Z_{5,i}N_{1,i})\Big)\Big]\\ =\;& O\left(n^{-3}\right).\end{aligned}$$

In the case of exactly one set with three elements and all other sets containing only one element, we just have to consider the case $L=\{i,j,k\}$, $I=\{i\}$, $J=\{j\}$ and $K=\{k\}$, which yields $E\left[W_{\{i\}}W_{\{j\}}W_{\{k\}}W_{\{i,j,k\}}\right]=O\left(n^{-4}\right)$. The case of exactly two sets with two elements and two sets with one element also gives the rate $O\left(n^{-4}\right)$. If two sets are singletons, one set contains two and one set three elements, we consider exemplarily the case $L=\{i,j,k\}$, $K=\{i,j\}$, $I=J=\{k\}$, which gives $E\left[W_{\{i,j,k\}}W_{\{i,j\}}W_{\{k\}}^2\right]=O\left(n^{-5}\right)$. All other scenarios of this case yield the same rate. The case where two sets are singletons and two sets contain three elements and the case where one set is a singleton, two sets contain two elements and one set contains three elements both give a rate $O(n^{-6})$. If one set is a singleton, one set contains two and two sets contain three elements, the rate is $O(n^{-7})$, and if two sets contain two elements and two sets contain three elements we get the rate $O(n^{-8})$. In the case where there exist precisely three or four sets with two elements and the remaining set is a singleton we obtain by similar arguments the rates $O(n^{-5})$ and $O(n^{-6})$, respectively. When three sets contain two elements and one set contains three elements the rate is $O(n^{-7})$. Finally, when there exist exactly three or four sets with three elements, we get a rate $O(n^{-8})$ if the remaining set is a singleton, or $O(n^{-9})$ if the remaining set contains two elements, and $O(n^{-10})$, respectively. Counting the number of nonzero summands in each case completes the proof of the assertion (5.5).

References

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

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

de Jong, P. (1996). A central limit theorem with applications to random hypergraphs. *Random Structures and Algorithms*, 8(2):105–120.

Dette, H. and Munk, A. (1998). Testing heteroscedasticity in nonparametric regression. *Journal of the Royal Statistical Society, Ser. B*, 60:693–708.

Dette, H., Pardo-Fernández, J. C., and Van Keilegom, I. (2009). Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics*, 36:782–799.

Dette, H. and Wieczorek, G. (2009). Testing for a constant coefficient of variation in nonparametric regression. *Journal of Statistical Theory and Practice*, 3(3):587–612.

Eagleson, G. K. and Müller, H. G. (1997). Transformations for smooth regression models with multiplicative errors. *Journal of the Royal Statistical Society, Ser. B.*, 59(1):173–189.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85(3):645–660.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models. 2nd ed.* Chapman & Hall, London.

Ruppert, D., Wand, M. P., Holst, U., and Hössjer, O. (1997). Local polynomial variance function estimation. *Technometrics*, 39:262–273.

Sacks, J. and Ylvisaker, N. D. (1966). Designs for regression problems with correlated errors. *Annals of Mathematical Statistics*, 37:66–89.

Yao, Q. and Tong, H. (2000). Nonparametric estimation of ratios of noise to signal in stochastic regression. *Statistica Sinica*, 10:751–770.

Yu, K. and Jones, M. C. (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association*, 99(465):139–144.