Some comments on specification tests in nonparametric absolutely regular processes

Holger Dette and Ingrid Spreckelsen
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany

July 12, 2001

Abstract

In this note several aspects of specification tests in nonparametric models driven by an absolutely regular process are discussed, which were recently proprosed in the literature. In particular we give a more detailed asymptotic analysis of tests based on kernel methods under fixed alternatives using a central limit theorem for U-statistics with n-dependent nondegenerate kernel. As a by-product it is demonstrated that several results regarding the asymptotic distribution of goodness-of-fit tests are incorrectly stated in the literature. Moreover, our result indicates that recent results on the asymptotic equivalence between nonparametric autoregression and nonparametric regression cannot be used for the asymptotic analysis of goodness-of-fit tests under fixed alternatives.

Keywords: nonparametric autoregressive model, goodness-of-fit test, absolute regular process, asymptotic equivalence.

1 Introduction

Let (X_t, Y_t) denote a strictly stationary discrete time stochastic process with $Y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^d$ and

$$Y_t = m(X_t) + \varepsilon_t, \tag{1.1}$$

where ε_t is an unobservable error with mean 0 and m is an unknown function. Several authors have discussed the problem of estimating the regression function [see e.g. Robinson (1983), Tjøstheim (1994) or Masry and Tjøstheim (1995) among many others]. It is well known that for a high dimensional predictor the regression function m cannot be estimated efficiently [see Stone (1985)] due to the so called curse of dimensionality and therefore parametric models are usually preferred to the purely nonparametric approach. On one hand the specific parameters in the model usually admit a direct interpretation, yet on the other hand the information available from the observations can be increased by the application of efficient or optimal procedures. However, misspecification of such a model may lead to serious errors in the subsequent data analysis and many authors recommend testing the goodness-of-fit of the postulated model [see Shillington (1979), or Neil and Johnson (1985) for early reference and Azzalini and Bowman (1993), Härdle and Mammen (1993) or Zheng (1996) for more recent reference in the context of independent observations (X_t, Y_t) in the model (1.1)]. More recently several authors have addressed the problem of testing the goodness-of-fit in the context of time series modeling [e.g. Hong-zhi and Bing (1991), Hjellvik and Tjøstheim (1995, 1996), Fan and Li (1996), Hjellvik, Yao and Tjøstheim (1998) or Koul and Stute (1999) among many others]. The proposed methods can be roughly divided into two classes: tests based on marked empirical processes and tests based on kernel methods.

In the present paper we will take a more careful look at the approach based on kernel methods with three objectives in mind. First we will give a detailed analysis of the corresponding tests under fixed alternatives. As pointed out by Dette and Munk (1998) the asymptotic distributional properties of a goodness-of-fit test under fixed alternatives are important for the analysis of the type II error and for the problem of testing precise hypotheses [see Berger and Delampady (1987) and Sellke, Bayarri and Berger (2001)].

Note that most authors discuss only asymptotic properties of the test statistic under the null hypothesis and local alternatives and prove consistency [e.g. Fan and Li (1999) or Kreiss, Neumann and Yao (1998)]. Some recent results regarding the asymptotic distribution of goodness-of-fit statistics for parametric model assumptions under fixed alternatives are given in Theorem 3.3 of Hjellvik, Yao and Tjøstheim (1998) who compared a nonparametric with a linear lag-k predictor for the construction of a test of linearity. Our second objective is to demonstrate that these results are incorrect and a substantially more sophisticated analysis is required to obtain the exact asymptotic distribution under fixed alternatives.

The third objective of this paper refers to the recent interesting and promising work of Neumann and Kreiss (1998), who showed the asymptotic equivalence between a nonparametric autoregression model and a nonparametric regression model with a nonrandom design. Although this equivalence can be used for deriving the asymptotic distribution of a modified L_2-distance between a nonparametric estimator of the regression function and its parametric counterpart under the null hypothesis [see Kreiss, Neumann and Yao (1998)], the results of the present paper show that this technique is unfortunately not applicable for the asymptotic analysis of the test under fixed alternatives. Our results suggest that differences exist between the nonparametric autoregression and the fixed design nonparametric regression model, which cannot be detected by the approach of Neumann and Kreiss (1998).

For the sake of brevity we will mainly concentrate on the problem of testing linearity and goodness-of-fit test, which was proposed by Zheng (1996) in the context of an i.i.d. sample and investigated by Fan and Li (1999) for absolutely regular processes. In comparision with Fan and Li we will use slightly different regularity assumptions, which will make the asymptotic analysis under fixed alternatives a little easier. In section 2 we will prove asymptotic normality of Fan and Li's (1999) statistic under fixed alternatives with a different rate of convergence as under the null hypothesis. The proof is based on a central limit theorem for U-statistics with a nondegenerate kernel, which will be proved in the appendix and can be seen as a complement to the central limit theorems for degenerate U-statistics obtained by Hjellvik, Yao and Tjøstheim (1998) and Fan and Li (1999). Section 3 discusses extensions and gives the corresponding results for the statistics considered by Kreiss, Neumann and Yao (1998), Hjellvik, Yao and Tjøstheim (1998) and Dette (1999). Finally, some of the proofs of the more technical results in Section 2 are given in an appendix in Section 4.

2 A central limit theorem for nondegenerate U-statistics and its application to model checks

Consider the strictly stationary discrete time stochastic process defined in (1.1) and let

$$g_1, \dots, g_p : \mathbb{R}^d \to \mathbb{R}$$

denote linearly independent regression functions. We are interested in testing the hypothesis of linearity, i.e.

$$H_0 : m(x) = \sum_{j=1}^{p} \vartheta_j g_j(x) \quad \forall\, x \in \text{supp}(\pi) \tag{2.1}$$

for some $\vartheta_1, \dots, \vartheta_p \in \mathbb{R}$, where π is the density of the stationary distribution of X_i. Let $g(x) = (g_1(x), \dots, g_p(x))^\mathsf{T}$ be the vector of regression functions and

$$\hat{\vartheta}_n = \Big(\frac{1}{n}\sum_{i=1}^{n} g(X_i) g^\mathsf{T}(X_i)\Big)^{-1} \frac{1}{n}\sum_{i=1}^{n} g(X_i) Y_i \tag{2.2}$$

denote the least squares estimator of the unknown parameter $\vartheta = (\vartheta_1 \dots, \vartheta_p)^\mathsf{T}$. Following Zheng (1996) we define the test statistic for the hypothesis of linearity by

$$T_n = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \hat{\varepsilon}_i \hat{\varepsilon}_j, \tag{2.3}$$

where $\hat{\varepsilon}_i = Y_i - \hat{\vartheta}_n^\mathsf{T} g(X_i)$ are the residuals obtained under the assumption of linearity, K is a kernel function and h is a bandwidth. A straightforward calculation shows that

$$T_n = (T_{1n} - 2\,T_{2n} + T_{3n}) \;+\; (2\,T_{4n} - 2\,T_{5n} + T_{6n}), \tag{2.4}$$

where

$$
\begin{aligned}
T_{1n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \varepsilon_i \, \varepsilon_j \\
T_{2n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \varepsilon_i \, \delta_j \\
T_{3n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \delta_i \, \delta_j \qquad (2.5) \\
T_{4n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \varepsilon_i \, \Delta_j \\
T_{5n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \delta_i \, \Delta_j \\
T_{6n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\Big(\frac{X_i - X_j}{h}\Big) \Delta_i \Delta_j,
\end{aligned}
$$

$\delta_i = \hat{\vartheta}_n^{\mathsf{T}} g(X_i) - \vartheta_0^{\mathsf{T}} g(X_i)$, $\Delta_i = \Delta(X_i) = m(X_i) - \vartheta_0^{\mathsf{T}} g(X_i)$ and ϑ_0 ist the unique value minimizing

$$E[m(X_t) - \vartheta_0^{\mathsf{T}} g(X_t)]^2,$$

that is

$$\vartheta_0 = E[g(X_t) g^{\mathsf{T}}(X_t)]^{-1} E[m(X_t) g(X_t)]. \qquad (2.6)$$

Note that under the hypothesis of linearity (2.1) we have $\Delta_i = 0\,;\ i = 1, \ldots, n$ and consequently $T_{4n} = T_{5n} = T_{6n} = 0$. Our first result reviews the asymptotic properties of the statistic T_n under the null hypothesis. For this purpose we need the following assumptions for some fixed $\varepsilon > 0$.

(A1) The process $(V_t)_{t \in \mathbb{Z}} = (X_t, \varepsilon_t)_{t \in \mathbb{Z}}$ in the model (1.1) is strictly stationary and absolutely regular with mixing coefficients $\beta(k)$ satisfying

$$\sum_{k=1}^{\infty} k^2 \beta^{\frac{\varepsilon}{1+\varepsilon}}(k) < \infty. \qquad (2.7)$$

The density π of X_t has compact support and is r-times continuously differentiable for some $r \geq 2$.

(A2) The innovations ε_t in the model (1.1) satisfy

$$E(\varepsilon_t | X_t, \mathcal{F}_1^{t-1}(X, Y)) = E(\varepsilon_t | X_t) = 0, \qquad (2.8)$$

$$\mathrm{Var}(Y_t \,|\, X_t = x, \mathcal{F}_1^{t-1}(X, Y)) = E(\varepsilon_t^2 | X_t = x) =: \sigma^2(x), \qquad (2.9)$$

where $\mathcal{F}_1^{t-1}(X,Y)$ denotes the σ-algebra generated by the sequence $(X_j,Y_j)_{j=1}^{t-1}$ and the function σ^2 is twice continuously differentiable.

(A3) The kernel K used in the definiton of the test statistic T_n in (2.3) is symmetric, bounded, of order $r \geq 2$ with compact support, such that

$$\int K^2(u)\,du < \infty.$$

The regression functions $g_1,\dots,g_p$ are orthonormal with respect to the stationary density π and r-times continuously differentiable.

(A4) The bandwidth h used in the definition of the test statistic T_n in (2.3) satisfies as $n \to \infty$

$$h \to 0, \quad nh^d \to \infty, \quad nh^{2r} \to 0. \tag{2.10}$$

Our first result shows that under the null hypothesis of linearity the dominating term in T_n is the statistic T_{1n}, which is asymptotically normally distributed. The proof is similar to the proof of the corresponding statement in Fan and Li (1999) [where we use Lemma 3.2 in Hjellvik, Yao and Tjøstheim (1998) instead of their central limit theorem 2.1 for degenerate U-statistics] and is therefore omitted.

Theorem 2.1 *If assumptions (A1)-(A4) are satisfied for the process defined by (1.1) and the hypothesis H_0 is valid, then the statistics T_{1n}, T_{2n}, T_{3n} and T_n defined in (2.3) and (2.5) satisfy*

$$T_{2n} = o_p(n^{-1}h^{-d/2}), \tag{2.11}$$

$$T_{3n} = o_p(n^{-1}h^{-d/2}), \tag{2.12}$$

$$nh^{d/2}T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,V), \quad nh^{d/2}T_{1n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,V), \tag{2.13}$$

where the asymptotic variance is given by

$$V = 2\int K^2(u)\,du \int [\sigma^2(x)]^2\pi^2(x)\,dx. \tag{2.14}$$

We now discuss the asymptotic properties of the statistic T_n under fixed alternatives. In this case the terms T_{4n}, T_{5n}, T_{6n} in the decomposition (2.4) do not vanish. We will show that under a fixed alternative these random variables are of order $O_p(1/\sqrt{n})$ and determine the asymptotic behaviour of the statistic T_n, while Theorem 2.1 shows that the random varibles T_{1n}, T_{2n} and T_{3n} are asymptotically negligible, that is

$$T_{jn} = O_p(n^{-\frac{1}{2}}h^{-\frac{d}{2}}) = o_p(n^{-\frac{1}{2}}), \quad j=1,2,3$$

and

$$T_n = 2\,T_{4n} - 2\,T_{5n} + T_{6n} + o_p(n^{-\frac{1}{2}}). \tag{2.15}$$

The following lemma gives a simpler representation for the dominating term on the right hand side of (2.15). The proof is deferred to the appendix.

Lemma 2.2 *If assumptions (A1)-(A4) are satisfied and the hypothesis of linearity in (2.1) does not hold, then*

$$T_n = U_n + o_p(n^{-\frac{1}{2}}), \tag{2.16}$$

where the statistic U_n is defined by

$$U_n = 2\,T_{4n} - 2\,\tilde{T}_{5n} + T_{6n} \tag{2.17}$$

and

$$\tilde{T}_{5n} = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{l=1}^{p} g_l(X_i)(\Delta(X_i) + \varepsilon_i)\Delta(X_j) g_l(X_j)\pi(X_j). \tag{2.18}$$

Note that Lemma 2.2 shows that the asymptotic properties of the statistic T_n can be studied by considering the U-statistic

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} h_n(V_i, V_j), \tag{2.19}$$

where $V_i = (X_i, \varepsilon_i)$ and h_n denotes a symmetric kernel (depending on n) defined by

$$\begin{aligned} h_n(V_i, V_j) \;&=\; h^{-d} K\Big(\frac{X_i - X_j}{h}\Big)\{\varepsilon_i \Delta_j + \Delta_i \varepsilon_j + \Delta_i \Delta_j\} \\ &\quad - \sum_{l=1}^{p} \Big\{ g_l(X_i)(\Delta_i + \varepsilon_i)\Delta_j g_l(X_j)\pi(X_j) + g_l(X_i)\pi(X_i)\Delta_i g_l(X_j)(\Delta_j + \varepsilon_j) \Big\}. \end{aligned} \tag{2.20}$$

Note that the kernel of U_n depends on the sample size n and that this type of U-statistic has recently been investigated by Fan and Li (1999), who proved asymptotic normality for a standardized version of the U-statistic in the case of a degenerate kernel h_n, that is

$$E[h_n(v, V_j)] = 0 \quad \text{a.e.} \tag{2.21}$$

Unfortunately, if the null hypothesis of linearity is not satisfied, then the kernel defined by (2.20) is nondegenerate in the aforementioned way [see Lemma 2.4 below] and the central limit theorem of Fan and Li (1999) is not applicable for studying the asymptotic properties of the statistic U_n under fixed alternatives.

In order to establish a corresponding result in the nondegenerate case we will introduce a Hoeffding decomposition

$$U_n = \gamma_n + 2H_n^{(1)} + H_n^{(2)}, \tag{2.22}$$

where $\gamma_n = E^{\otimes} h_n(V_i, V_j)$, $E^{\otimes}$ denotes the expectation with respect to the product measure $P^{V_i} \otimes P^{V_j}$ and

$$\begin{aligned} H_n^{(1)} &= \frac{1}{n} \sum_{i=1}^{n} h_n^{(1)}(V_i), \\ H_n^{(2)} &= \frac{1}{n(n-1)} \sum_{i \neq j} h_n^{(2)}(V_i, V_j), \end{aligned} \tag{2.23}$$

where the kernels in the statistics $H_n^{(1)}$ and $H_n^{(2)}$ are defined by

$$\begin{aligned} h_n^{(1)}(v) &= E(h_n(v, V_j)) - \gamma_n, \\ h_n^{(2)}(v, w) &= h_n(v, w) - E(h_n(v, V_j)) - E(h_n(V_i, w)) + \gamma_n. \end{aligned} \tag{2.24}$$

The proof of the following theorem, which gives a central limit theorem for U-statistics with nondegenerate n-dependent kernel, is deferred to the appendix.

Theorem 2.3 *Let $(V_i)_{i \in \mathbb{Z}}$ denote a strictly stationary, absolutely regular $\mathbb{R}^{d+1}$-valued process with mixing coefficients satisfying*

$$\sum_{k=1}^{\infty} k\, \beta^{\frac{\varepsilon}{1+\varepsilon}}(k) < \infty \tag{2.25}$$

for some $\varepsilon > 0$ and assume that

$$\begin{aligned} K_n &= \max\{ \sup_{\substack{i \neq j, k \neq l, \\ j \neq l}} E|h_n(V_i, V_j) h_n(V_k, V_l)|^{1+\varepsilon}, \sup_{\substack{i \neq j, k \neq l, \\ j \neq l}} E^{1\otimes}|h_n(V_i, V_j) h_n(V_k, V_l)|^{1+\varepsilon}, \\ &\quad \sup_{\substack{i \neq j, k \neq l, \\ j \neq l}} E^{3\otimes}|h_n(V_i, V_j) h_n(V_k, V_l)|^{1+\varepsilon}, \sup_{\substack{i \neq j, i \neq l, \\ j \neq l}} E^{2\otimes}|h_n^{(2)}(V_i, V_j) h_n^{(2)}(V_i, V_l)|^{1+\varepsilon}\} \\ &= O(h^{-d(1+\varepsilon)}), \end{aligned} \tag{2.26}$$

where $h \to 0$, $nh^d \to \infty$ and for $i_1 < i_2 < i_3 < i_4$ $E^{1\otimes}$, $E^{2\otimes}$ and $E^{3\otimes}$ denote the expectations with respect to the measures $P^{V_{i_1}} \otimes P^{(V_{i_2}, V_{i_3}, V_{i_4})}$, $P^{(V_{i_1}, V_{i_2})} \otimes P^{V_{i_3}}$ and $P^{(V_{i_1}, V_{i_2}, V_{i_3})} \otimes P^{V_{i_4}}$, respectively. In addition, if for some $q > 2$ the estimates

$$E|h_n^{(1)}(V_i))|^q = O(1) \quad \text{and} \quad E|h_n^{(2)}(V_i, V_j)|^2 = O(h^{-d}) \tag{2.27}$$

hold and the U-statistic U_n defined by (2.22) and (2.23) is nondegenerate, i.e. Var$(h_n^{(1)}(V_i)) \neq 0$, then the statistic U_n is asymptotically normally distributed, i.e.

$$\frac{U_n - E^{\otimes} U_n}{\sqrt{\operatorname{Var} U_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{2.28}$$

Moreover, the variance of U_n is given by

$$\begin{aligned} \operatorname{Var}(U_n) &= 4\operatorname{Var}(H_n^{(1)}) + o(\frac{1}{n}) \\ &= \frac{4}{n} \Big\{ \operatorname{Var} h_n^{(1)}(V_1) + 2 \sum_{t=1}^{n-1} \operatorname{Cov}(h_n^{(1)}(V_1), h_n^{(1)}(V_{t+1})) \Big\} + o(\frac{1}{n}). \end{aligned} \tag{2.29}$$

In order to apply this central limit theorem to the U-statistic U_n defined in (2.17) we use the following representation for the kernel $h_n^{(1)}$ in the U-statistic $H_n^{(1)}$.

Lemma 2.4 *If assumptions (A1)-(A4) are satisfied and the hypothesis of linearity in (2.1) does not hold, then the kernel $h_n^{(1)}$ can be expanded as*

$$h_n^{(1)}(v) = (\Delta(x) + e)[(\Delta\pi)(x) - \mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(x)] - \gamma_n + O(h^r)$$

uniformly with respect to $v = (x, e)$, where $\mathbb{P}_{\{g_1,\dots,g_p\}} f$ denotes the orthogonal projection of the function f onto span $\{g_1, \dots, g_p\}$ with respect to the inner product $\langle f_1, f_2\rangle = E[f_1(X_1)f_2(X_1)]$.

Note that the kernel $h_n^{(1)}$ in Lemma 2.4 is obviously not identically zero and therefore the kernel h_n defined by (2.20) is nondegenerate. Therefore we can use Theorem 2.3 and Lemma 2.4 to establish asymptotic normality of the statistic U_n in (2.17) under fixed alternatives.

Theorem 2.5 *If assumptions (A1)-(A4) are satisfied and the hypothesis of linearity in (2.1) does not hold, then the statistic T_n defined in (2.3) is asymptotically normally distributed, i.e.*

$$\sqrt{n}(T_n - B) \xrightarrow{\mathcal{D}} \mathcal{N}(0, W),$$

where the asymptotic bias is given by

$$B = \int\int K(v)\,(\Delta\pi)(x)\,(\Delta\pi)(x - hv)\;dx\,dv,$$

the asymptotic variance is

$$\begin{aligned} W \;=\; & 4\Big[\int \sigma^2(x)\Big\{(\Delta\pi)(x) - \mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(x)\Big\}^2 \pi(x)dx \\ & + \text{Var}\Big(\Delta(X_1)[(\Delta\pi)(X_1) - \mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]\Big) \\ & + 2\sum_{t=1}^{\infty} \text{Cov}\Big([\Delta(X_1) + \varepsilon_1]\,[(\Delta\pi)(X_1) - \mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)], \\ & \Delta(X_{1+t})\,[(\Delta\pi)(X_{1+t}) - \mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_{1+t})]\Big)\Big] \end{aligned}$$

and $\mathbb{P}_{\{g_1,\dots,g_p\}} f$ denotes the orthogonal projection of the function f onto span $\{g_1, \dots, g_p\}$ with respect to the inner product $\langle f_1, f_2\rangle = E[f_1(X_1)f_2(X_1)]$.

Proof of Theorem 2.5
By the discussion in Lemma 2.2 the asymptotic normality of the statistic T_n under fixed alternatives can be established by proving asymptotic normality for the U-statistic U_n defined in (2.19). To this end it is sufficient to prove that the assumptions of Theorem 2.3 are satisfied for the kernel h_n given in (2.20). Conditions (2.25) and (2.27) are obviously satisfied [because

of assumption (A2) and Lemma 2.4] and an application of Hölder's inequality [$\frac{1}{\xi}+\frac{1}{\eta}=1$, $\eta<\frac{2}{1+\varepsilon}$] yields for $j\neq l$ for the first two terms in $h_n(V_i,V_j)$

$$E\Big|h^{-d}K\Big(\frac{X_i-X_j}{h}\Big)\Big\{\Delta_i\varepsilon_j+\varepsilon_i\Delta_j+\Delta_i\Delta_j\Big\}h^{-d}K\Big(\frac{X_k-X_l}{h}\Big)\Big\{\Delta_k\varepsilon_l+\varepsilon_k\Delta_l+\Delta_k\Delta_l\Big\}\Big|^{1+\varepsilon}$$

$$= \Big[E\Big|\{\Delta_i\varepsilon_j+\varepsilon_i\Delta_j+\Delta_i\Delta_j\}\{\Delta_k\varepsilon_l+\varepsilon_k\Delta_l+\Delta_k\Delta_l\}\Big|^{\xi(1+\varepsilon)}\Big]^{\frac{1}{\xi}}$$

$$\times\Big[E\Big|h^{-d}K\Big(\frac{X_i-X_j}{h}\Big)h^{-d}K\Big(\frac{X_k-X_l}{h}\Big)\Big|^{\eta(1+\varepsilon)}\Big]^{\frac{1}{\eta}}$$

$$= O(1)\times O(h^{-2d(1+\varepsilon)}h^{\frac{2d}{\eta}})=O(h^{-d(1+\varepsilon)}).$$

The third term of $h_n(V_i,V_j)$ is treated by the same arguments and it follows that

$$E|h_n(V_i,V_j)h_n(V_k,V_l)|^{1+\varepsilon}=O(h^{-d(1+\varepsilon)}).$$

Similar calculations for the other cases yield for the quantitiy K_n defined in (2.26)

$$K_n=O(h^{-d(1+\varepsilon)}),$$

which establishes the assumption (2.26). From Theorem 2.3 it therefore follows that

$$\frac{U_n-E^{\otimes}U_n}{\sqrt{\mathrm{Var}\,U_n}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1)$$

and it remains to calculate the expectation and the variance. The calculation of $E^{\otimes}U_n$ is straightforward and follows by similar arguments as given in Dette (1999), that is

$$E^{\otimes}U_n=\int\int K(v)\,\Delta(x)\,\Delta(x-hv)\,\pi(x)\,\pi(x-hv)\,dx\,dv.$$

For the determination of the variance we apply the second part of Theorem 2.3 and Lemma 2.4.

$$n\mathrm{Var}(U_n)=\Big\{\mathrm{Var}\,\tilde h_n^{(1)}(V_1)+2\sum_{t=1}^{n-1}\mathrm{Cov}\,(\tilde h_n^{(1)}(V_1),\tilde h_n^{(1)}(V_{1+t}))\Big\}+o(1)$$

where

$$\tilde h_n^{(1)}(V_1)=(\Delta(X_1)+\varepsilon_1)[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]$$

and

$$\begin{aligned}\mathrm{Var}\,\tilde h_n^{(1)}(V_1) &= \mathrm{Var}\Big\{\Delta(X_1)[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]\Big\}\\ &+ \mathrm{Var}\Big\{\varepsilon_1[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]\Big\}\\ &+ 2\mathrm{Cov}\Big\{\Delta(X_1)[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)],\varepsilon_1[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]\Big\}\\ &= \mathrm{Var}\Big\{\Delta(X_1)[(\Delta\pi)(X_1)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(X_1)]\Big\}\\ &+ \int\sigma^2(x)\Big\{(\Delta\pi)(x)-\mathbb{P}_{\{g_1,\dots,g_p\}}(\Delta\pi)(x)\Big\}^2\pi(x)\,dx,\end{aligned}$$

$$
\begin{aligned}
\mathrm{Cov}\,(\tilde{h}_n^{(1)}(V_1), \tilde{h}_n^{(1)}(V_{1+t})) \;=\; & \mathrm{Cov}\Big\{[\Delta(X_1)+\varepsilon_1]\,[(\Delta\pi)(X_1) - \mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi)(X_1)], \\
& [\Delta(X_{1+t})+\varepsilon_{1+t}]\,[(\Delta\pi)(X_{1+t}) - \mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi)(X_{1+t})]\Big\} \\
\;=\; & \mathrm{Cov}\Big\{[\Delta(X_1)+\varepsilon_1]\,[(\Delta\pi)(X_1) - \mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi)(X_1)], \\
& \Delta(X_{1+t})\,[(\Delta\pi)(X_{1+t}) - \mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi)(X_{1+t})]\Big\}.
\end{aligned}
$$

From the proof of Theorem 2.3 it finally follows that

$$
\lim_{n\to\infty} \sum_{t=1}^{n-1} \mathrm{Cov}\,(\tilde{h}_n^{(1)}(V_1), \tilde{h}_n^{(1)}(V_{1+t}))
$$

exists, which completes the proof of Theorem 2.5.

3 Further discussion

3.1 Asymptotic equivalence

In a recent paper Neumann and Kreiss (1998) derived a strong approximation of a local polynomial estimator in the nonparametric autoregression model

$$
X_t = m(X_{t-1}, \ldots, X_{t-p}) + \nu(X_{t-1}, \ldots, X_{t-q})\varepsilon_t \tag{3.1}
$$

by a local polynomial estimator in a corresponding regression model. These results generally suggest the application of typical tools used in regression analysis for statistical inference in nonparametric autoregressive models. While this principle was applied successfully by Kreiss, Neumann and Yao (1998) to derive the asymptotic distribution of a test statistic for testing linearity of the function m under the null hypothesis (2.1), our results indicate that the asymptotic equivalence is not applicable to obtain a complete asymptotic analysis of goodness-of-fit tests in the model (3.1). To be precise, note at first, that the model in this paper contains the nonparametric autoregressive model as a special case. The asymptotic distribution of the statistic T_n in the nonparametric regression model with a fixed design under the hypothesis of linearity was derived by Dette (1999) and coincides with the asymptotic normal distribution in Theorem 2.1. This indicates that under the null hypothesis the principle suggested by Neumann and Kreiss (1998) is again applicable for the statistic T_n considered in this paper. However, under fixed alternatives the results of Dette (1999) for the corresponding fixed and random design nonparametric regression model yield a substantially different limiting variance of the statistic T_n, which demonstrates that the asymptotic principle proposed by Neumann and Kreiss (1998) is not applicable under fixed alternatives of linearity. In this case a more sophisticated asympotic analysis, as described in section 2 and the appendix, is necessary to obtain the distributional properties of the test statistic T_n.

3.2 Related goodness-of-fit tests

It is worthwhile to mention that several related tests have been discussed in the literature in the case of independent observations [see Gonzalez-Manteiga and Cao-Abad (1993), Azzalini and Bowman (1993), Härdle and Mammen (1993), Zheng (1996), Dette (1999)]. Some of these methods have been generalized to the situation of stationary time series [see Fan and Li (1999), Hjellvik, Yao and Tjøstheim (1998), Kreiss, Neumann and Yao (1998)]. While the asymptotic distribution of the corresponding test statistics under the null hypothesis of linearity can be found in these papers, related results for fixed alternatives are either not available or incorrectly stated in the cited references. We consider as an example the statistics proposed by Dette (1999) and Hjellvik, Tjøstheim and Yao (1998). The first author proposed using a difference of variance estimators for a goodness-of-fit test of linearity in the model (1.1), i.e.

$$T_n^* = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\vartheta}_n^{\mathsf{T}} g(X_i))^2 - \frac{1}{f}\sum_{i=1}^{n}(Y_i - \hat{m}_h(X_i))^2, \tag{3.2}$$

where

$$\hat{m}_h(X_i) = \sum_{j=1}^{n} w_{ij} Y_j$$

denotes the common Nadaraya-Watson estimator (evaluated at X_i) with bandwidth h and kernel K and the constant f depends on the sample size and is defined by

$$f = n - 2\sum_{i=1}^{n} w_{ii} + \sum_{i,k=1}^{n} w_{ik}^2.$$

Note that the first term in (3.2) is the variance estimator in the linear model, while the second term is a variance estimator in the nonparametric model [see Hall and Marron (1990)] and therefore the statistic in (3.2) is the natural generalization of the classical F-test for hypotheses in linear models. Under similar assumptions as stated in section 2, it can be shown that under the null hypothesis of linearity we have

$$nh^{\frac{d}{2}}(T_n^* + \frac{c_3}{nh^d}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_*^2),$$

where

$$c_3 = \Big(2K(0) - \int K^2(t)\,dt\Big)\Big(\int \sigma^2(t)\pi(t)\,dt - \int \sigma^2(t)\,dt\Big),$$

and the asymptotic variance is given by

$$\mu_*^2 = 2\int \Big(2K(u) - K * K(u)\Big)^2 du \int \sigma^4(u)\,du.$$

Under the alternative asymptotic normality is still valid but with a different rate of convergence, i.e.

$$\sqrt{n}(T_n^* - E[\Delta^2(X_1)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_*^2),$$

where the asymptotic variance is given by

$$\lambda_*^2 = 4E[\Delta^2(X_1)\sigma^2(X_1)] + \text{Var}[\Delta^2(X_1)] + 2\sum_{t=1}^{\infty}\text{Cov}[(\Delta(X_1)+\varepsilon_1)\Delta(X_1), \Delta^2(X_{1+t})]. \quad (3.3)$$

A similar statistic was proposed by Hjellvik, Yao and Tjøstheim (1998) [see also Gonzalez-Manteiga and Cao-Abad (1993) or Kreiss, Neumann and Yao (1998)]

$$T_n^{**} = \frac{1}{n}\sum_{i=1}^{n}(\hat{m}_h(X_i) - \hat{\vartheta}_n^{\mathsf{T}} g(X_i))^2, \quad (3.4)$$

where $\hat{m}_h$ is a local polynomial estimator of order $T \geq 0$ ($T = 0$ corresponds to the Nadaraya-Watson estimator). Note that the statistic (3.4) is more general than the statistic considered by Hjellvik, Yao and Tjøstheim (1998) and contains the situation considered by these authors as a special case. Under the null hypothesis of linearity we have

$$nh^{\frac{d}{2}}(T_n^{**} - \frac{A_3}{nh^d}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu_{**}^2),$$

where the constant A_3 depends on the smoothing method [see Hjellvik, Yao and Tjøstheim (1998) or Kreiss, Neumann and Yao (1998)]. The asymptotic variance is given by

$$\mu_{**}^2 = 2\int(\tilde{K} * \tilde{K})^2(u)\,du \int \sigma^4(u)\,du$$

where $\tilde{K}$ is the asymptotic equivalent kernel for the local polynomial estimator under consideration [see Wand and Jones (1995)]. Under the alternative it follows by similar arguments as given in the appendix

$$\sqrt{n}(T_n^{**} - E[\Delta^2(X_1)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \lambda_{**}^2),$$

where the asymptotic variance is given by

$$\lambda_{**}^2 = 4\Big\{E[\Delta^2(X_1)\sigma^2(X_1)] + \text{Var}[\Delta^2(X_1)] + 2\sum_{t=1}^{\infty}\text{Cov}[(\Delta(X_1)+\varepsilon_1)\Delta(X_1), \Delta^2(X_{1+t})]\Big\}. \quad (3.5)$$

Note that in the case $p = 1$ this statement corrects a conjecture in Theorem 3.3 (i) of Hjellvik, Yao and Tjøstheim (1998). The general case $p > 1$ and the two other statements in Theorem 3.3 of this reference have to be corrected similarly.

Finally we note that the representations for the asymptotic variances under the alternative indicate that in many cases a test based on T_n^* should be more powerful than a test based on T_n^{**}. To be precise, assume that the innovations in the model (1.1) $(\varepsilon_t)_{t\in\mathbb{Z}}$ are independent of $(X_t)_{t\in\mathbb{Z}}$, then it follows that

$$\nu^2 = \text{Var}[\Delta^2(X_1)] + 2\sum_{t=1}^{\infty}\text{Cov}[(\Delta^2(X_1), \Delta^2(X_{1+t})] = \lim_{n\to\infty}\frac{1}{n}\text{Var}\Big(\sum_{t=1}^{n}\Delta^2(X_t)\Big)$$

is nonnegative. Consequently the term ν^2 appears with different factors in the asymptotic variances (3.3) and (3.5) and we obtain

$$\lambda_*^2 < \lambda_{**}^2.$$

A straightforward analysis shows that the leading term in the asymptotic expansion for the probability of rejection is given by

$$\Phi\Big(\sqrt{n}\,\frac{E[\Delta^2(X_1)]}{\lambda}\Big), \qquad \lambda^2 = \lambda_{**}^2, \lambda_*^2$$

and consequently we expect a test based on T_n^* to have more power in the case of an independent process of innovations. We finally note that such a comparison is not possible for the test based on the statistic T_n defined in (2.3) because of its more complicated representation of the asymptotic variance in Theorem 2.5.

A Proof of technical results

We will start with a tool, which allows us to estimate the expectations of functions of strictly stationary processes. Let $(V_i)_{i\in\mathbb{Z}}$ be an absolutely regular process and let $t_1 < \ldots < t_k \in \mathbb{Z}$. We are interested in comparing the expectations

$$Eh_n(V_{t_1},\ldots,V_{t_k}) = \int h_n\, dP^{(V_{t_1},\ldots,V_{t_k})} \tag{A.1}$$

and

$$E^{j\otimes}h_n(V_{t_1},\ldots,V_{t_k}) = \int h_n\, dP^{(V_{t_1},\ldots,V_{t_j})} \otimes dP^{(V_{t_{j+1}},\ldots,V_{t_k})} \tag{A.2}$$

for an appropriately defined function h_n. The following result shows that the difference of these two expectations can be estimated by the mixing coefficient $\beta(t_{j+1}-t_j)$. The proof is similar to the proof of Lemma 2 in Takahata and Yoshihara (1987) and is therefore omitted [see also Lee (1990), Hjellvik, Yao and Tjøstheim (1998) and Fan and Li (1999) for similar results].

Lemma A.1 *If $(V_i)_{i\in\mathbb{Z}}$ is an absolutely regular process with mixing coefficients $\beta(j)$ and if for some $\delta > 0$ and $1 \le j < k$*

$$M_n = \max\{E|h_n(V_{t_1},\ldots,V_{t_k})|^{1+\delta}, E^{j\otimes}|h_n(V_{t_1},\ldots,V_{t_k})|^{1+\delta}\} < \infty,$$

then

$$|Eh_n(V_{t_1},\ldots,V_{t_k}) - E^{j\otimes}h_n(V_{t_1},\ldots,V_{t_k})| \le 3M_n^{\frac{1}{1+\delta}}\beta^{\frac{\delta}{1+\delta}}(t_{j+1}-t_j). \tag{A.3}$$

Proof of Theorem 2.3

The proof consists of three steps:

(1) We show that the statistic $H_n^{(2)}$ in the decomposition (2.22) is asymptotically negligible, that is

$$H_n^{(2)} = o_p(n^{-1/2}). \tag{A.4}$$

(2) We prove the representation of the variance of U_n, especially $\mathrm{Var}(U_n) = O(\frac{1}{n})$, $\mathrm{Var}(H_n^{(1)}) = O(\frac{1}{n})$.

(3) We establish asymptotic normality of

$$\frac{U_n - \gamma_n}{\sqrt{\mathrm{Var}\, U_n}}.$$

(1) The estimate (A.4) can be obtained by calculating the second moment of $H_n^{(2)}$ with the aid of Lemma A.1,

$$E(H_n^{(2)})^2 = \Big(\frac{2}{n(n-1)}\Big)^2 \sum_{i<j,k<l} E\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\}.$$

Observing the assumptions (2.26), (2.27) and (A4) it follows by a straightforward application of Lemma A.1 that

$$E(H_n^{(2)})^2 = \frac{4}{n^2(n-1)} \sum_{i<j,k<l}{}^{*} E\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\} + o(\frac{1}{n}), \tag{A.5}$$

where $\sum^*$ denotes the summation over pairwise different indices. For the treatment of the leading term on the right hand side of (A.5) we consider as an example the summation over the indices with $i < j < k < l$. All remaining cases are treated similarly. Note that the kernel $h_n^{(2)}$ is degenerate, i.e.

$$Eh_n^{(2)}(v,V_j) = 0 \quad \text{a.e.},$$

which implies, by Fubini's theorem

$$E^{3\otimes}\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\} = \int h_n^{(2)}(v_i,v_j)h_n^{(2)}(v_k,v_l)\, dP^{(V_i,V_j,V_k)} \otimes dP^{V_l} = 0$$

and similarly

$$E^{1\otimes}\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\} = 0.$$

Consequently, we obtain from Lemma A.1 and assumption (2.26)

$$|E\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\}| \le 3K_n^{\frac{1}{1+\varepsilon}} \min\{\beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i), \beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k)\}$$

which shows

$$\begin{aligned} A_n &= \frac{2}{n^2(n-1)^2} \sum_{i<j<k<l} |E\{h_n^{(2)}(V_i,V_j)h_n^{(2)}(V_k,V_l)\}| \\ &\le \frac{2}{n^2(n-1)^2}\, 3K_n^{\frac{1}{1+\varepsilon}} \sum_{i<j<k<l} \min\{\beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i), \beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k)\} \\ &\le \frac{2}{n^2(n-1)^2}\, 3K_n^{\frac{1}{1+\varepsilon}} \Big[\sum_{\substack{i<j<k<l \\ l\le j-i+k}} \beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i) + \sum_{\substack{i<j<k<l \\ i>j+k-l}} \beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k) \Big]. \end{aligned} \tag{A.6}$$

We will now calculate the two sums on the right hand side of (A.6) separately. For the first term we have from (2.25)

$$\sum_{\substack{i<j<k<l \\ l\le j-i+k}} \beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i) = \sum_{k=3}^{n-1}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1}\sum_{l=k+1}^{j-i+k} \beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i)$$

$$= \sum_{k=3}^{n-1}\sum_{j=2}^{k-1}\sum_{i=1}^{j-1}(j-i)\beta^{\frac{\varepsilon}{1+\varepsilon}}(j-i) = O(n^2),$$

and similarly we obtain for the second term using assumption (2.25)

$$\sum_{\substack{i<j<k<l \\ i>j+k-l}} \beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k) = \sum_{j=2}^{n-2}\sum_{k=j+1}^{n-1}\sum_{l=k+1}^{n}\sum_{i=j+k-l+1}^{j-1} \beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k)$$

$$= \sum_{j=2}^{n-2}\sum_{k=j+1}^{n-1}\sum_{l=k+1}^{n}(l-k-1)\beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k)$$

$$= \sum_{j=2}^{n-2}\sum_{k'=1}^{n-j-1}\sum_{l=k'+j+1}^{n}(l-k'-j-1)\beta^{\frac{\varepsilon}{1+\varepsilon}}(l-k'-j) = O(n^2).$$

Combining these estimates with (2.26) shows that the term on the left hand side of (A.6) can be estimated as

$$A_n = O\Big(\frac{K_n^{\frac{1}{1+\varepsilon}}}{n^2}\Big) = O\Big(\frac{1}{n^2h^d}\Big) = o(\frac{1}{n}),$$

where we used $nh^d \to \infty$ for the last estimate. A similar argument for the remaining cases proves

$$|E(H_n^{(2)})^2| = o(\frac{1}{n}), \tag{A.7}$$

and the estimate (A.4) follows from Markov's inequality.

(2) For the calculation of the variance we have by standard calculation

$$\text{Var}\, H_n^{(1)} = \text{Var}(\frac{1}{n}\sum h_n^{(1)}(V_i)) = \frac{1}{n}\text{Var}\, h_n^{(1)}(V_1) + \frac{2}{n^2}\sum_{t=1}^{n-1}(n-t)\text{Cov}(h_n^{(1)}(V_1), h_n^{(1)}(V_{1+t})) \tag{A.8}$$

where the statistic $H_n^{(1)}$ is defined in (2.23). Moreover, we have $Eh_n^{(1)}(V_1) = 0$ and $E|h_n^{(1)}(V_i)h_n^{(1)}(V_j)|^{1+\varepsilon} < \infty$ by assumption (2.27), which, using Lemma A.1, yields

$$\Big|\text{Cov}(h_n^{(1)}(V_1), h_n^{(1)}(V_{1+t}))\Big| = \Big|E(h_n^{(1)}(V_1)h_n^{(1)}(V_{1+t}))\Big| \le \text{const}\, \beta^{\frac{\varepsilon}{1+\varepsilon}}(t).$$

Since $\sum t\beta^{\frac{\varepsilon}{1+\varepsilon}}(t) < \infty$ [by assumption (2.25)] we obtain

$$\frac{1}{n^2}\sum_{t=1}^{n-1} t\,\text{Cov}(h_n^{(1)}(V_1), h_n^{(1)}(V_{1+t})) = o(\frac{1}{n})$$

and from (A.8) the representation

$$\operatorname{Var} H_n^{(1)} = \frac{1}{n}\Big\{\operatorname{Var} h_n^{(1)}(V_1) + 2\sum_{t=1}^{n} \operatorname{Cov}(h_n^{(1)}(V_1), h_n^{(1)}(V_{t+1}))\Big\} + o(\frac{1}{n}).$$

The assertion (2.29) now follows from the decomposition (2.22), (A.7) and Cauchy's inequality.

(3) Following on from the discussion in the previous two steps we can prove asymptotic normality by establishing

$$W_n = \frac{H_n^{(1)}}{\sqrt{\operatorname{Var} H_n^{(1)}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{A.9}$$

To this end we will use a central limit theorem for α-mixing arrays which was recently obtained by Liebscher (1996). To be precise define $Y_{ni} = n^{-1/2} h_n^{(1)}(V_1)/\sqrt{\operatorname{Var} H_n^{(1)}}$, then the statistic W_n can be rewritten as

$$T_n = \sum_{i=1}^{n} Y_{ni}.$$

Note that the random variables Y_{ni} are α-mixing [Lee (1990), p.146] with mixing coefficients satisfying $\alpha(k) \leq \frac{1}{2}\beta(k)$ [Doukhan (1994), p.4]. We will now check the assumptions of Theorem 2.1 (i) in Liebscher (1996). From (2.25) we have

$$\sum_{k=1}^{\infty} \alpha^{1-\frac{2}{p}}(k) < \infty$$

for $p = 2 + 2\varepsilon$ and (2.27) shows for $q > p \quad E|Y_{ni}|^q = O(n^{-\frac{q}{2}})$, $E|Y_{ni}|^p = O(n^{-\frac{p}{2}})$,

$$\sum_{i=1}^{n} (E|Y_{ni}|^q)^{2/p} = O(n^{1-\frac{q}{p}}) = o(1),$$

$$\sum_{i=1}^{n} (E|Y_{ni}|^p)^{2/p} = O(1),$$

so that all assumptions of Liebscher's (1996) Theorem 2.1 (i) are satisfied (note that $EW_n^2 = \operatorname{Var}(W_n) = 1$). The central limit theorem in this reference proves asymptotic normality of W_n, which completes the proof of Theorem 2.3. □

Proof of Lemma 2.2

From the law of the iterated logarithm for absolutely regular processes [see Oodaira und Yoshihara (1971)] it is easy to see that

$$T_{5n} = \frac{1}{n(n-1)} \sum_{i,j} \sum_{l=1}^{p} g_l(X_i)(\Delta_i + \varepsilon_i)\Delta_j \frac{1}{n} \sum_{\substack{k=1 \\ k\neq j}}^{n} g_l(X_k) \frac{1}{h^d} K\Big(\frac{X_k - X_j}{h}\Big) + o_p(n^{-1/2}), \tag{A.10}$$

where we have changed the index of summation in the definiton of T_{5n}. If $\bar{T}_{5n}$ denotes the right hand side of equation (A.10), then the assertion of the lemma can be established by proving

$$\tilde{T}_{5n} - \bar{T}_{5n} = o_p(n^{-1/2}). \qquad \text{(A.11)}$$

To this end we show L^2-convergence, i.e. $E[(\tilde{T}_{5n} - \bar{T}_{5n})^2] = o(n^{-1})$, observing the representation

$$E(\bar{T}_{5n} - \tilde{T}_{5n})^2 = \frac{1}{n^4(n-1)^2} \sum_{\substack{i_1,j_1,k_1,i_2,j_2,k_2 \\ i_1\neq j_1, k_1\neq j_1, i_2\neq j_2, k_2\neq j_2}} E(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2}) \;+\; o(n^{-1}), \qquad \text{(A.12)}$$

where

$$\begin{aligned} \mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2} \;&=\; \sum_{l_1,l_2=1}^{p} g_{l_1}(X_{i_1})(\Delta_{i_1} + \varepsilon_{i_1})\Delta_{j_1} g_{l_2}(X_{i_2})(\Delta_{i_2} + \varepsilon_{i_2})\Delta_{j_2} \\ &\times\; \Big[g_{l_1}(X_{k_1})\frac{1}{h^d}K\Big(\frac{X_{k_1} - X_{j_1}}{h}\Big) - g_{l_1}(X_{j_1})\pi(X_{j_1})\Big] \\ &\times\; \Big[g_{l_2}(X_{k_2})\frac{1}{h^d}K\Big(\frac{X_{k_2} - X_{j_2}}{h}\Big) - g_{l_2}(X_{j_2})\pi(X_{j_2})\Big]. \end{aligned}$$

The expectations in (A.12) can now be estimated with the aid of Lemma A.1. To this end several cases for the structure of the tuple $(i_1, i_2, j_1, j_2, k_1, k_2)$ have to be distinguished and we discuss exemplarily two situations. As a first example we consider the summation over $i_1 < k_1 < j_2 < i_2 < j_1 < k_2$. Note that

$$E(g_l(X_i)[\Delta_i + \varepsilon_i]) = \int \Delta(x) g_l(x)\pi(x)\, dx = 0, \quad l = 1, \ldots, p$$

and therefore

$$E^{1\otimes}(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2}) = E^{34\otimes}(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2}) = 0,$$

where $E^{1\otimes}$ and $E^{34\otimes}$ denote the expectation with respect to the measures $P^{V_{i_1}} \otimes P^{(V_{k_1}, V_{j_2}, V_{i_2}, V_{j_1}, V_{k_2})}$ and $P^{(V_{i_1}, V_{k_1}, V_{j_2})} \otimes P^{V_{i_2}} \otimes P^{(V_{j_1}, V_{k_2})}$, respectively. A straightforward but tedious calculation shows

$$S_n := E|\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2}|^{1+\varepsilon} = O(h^{-2d(1+\varepsilon)} h^{\frac{2d}{\eta}}) + O(h^{-d(1+\varepsilon)} h^{\frac{d}{\eta}}) + O(1) = o(h^{-d(1+\varepsilon)})$$

for any $\eta < \frac{2}{1+\varepsilon}$, and Lemma A.1 yields the estimates

$$|E(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2})| \leq S_n^{\frac{1}{1+\varepsilon}} \beta^{\frac{\varepsilon}{1+\varepsilon}}(k_1 - i_1)$$

and

$$|E(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2}) - E^{34\otimes}(\mathcal{S}_{n,i_1 i_2 j_1 j_2 k_1 k_2})| \leq S_n^{\frac{1}{1+\varepsilon}} \Big[\beta^{\frac{\varepsilon}{1+\varepsilon}}(i_2 - j_2) + \beta^{\frac{\varepsilon}{1+\varepsilon}}(j_1 - i_2)\Big].$$

Now a similar argument as given in the proof of Theorem 2.3 shows

$$\sum_{i_1<k_1<j_2<i_2<j_1<k_2} \min\Big\{\beta^{\frac{\varepsilon}{1+\varepsilon}}(k_1-i_1), \beta^{\frac{\varepsilon}{1+\varepsilon}}(i_2-j_2)+\beta^{\frac{\varepsilon}{1+\varepsilon}}(j_1-i_2)\Big\} = O(n^4),$$

which gives for the corresponding sum in (A.12) the estimate

$$\frac{1}{n^4(n-1)^2}\sum_{i_1<k_1<j_2<i_2<j_1<k_2} E(\mathcal{S}_{n,i_1i_2j_1j_2k_1k_2}) = O(n^{-6}h^{-d}n^4) = o(n^{-1}).$$

Terms with one pair of equal indices can be treated in the same way, exept in the case of $i_1 = i_2$, for which we have

$$\begin{aligned}
& E^{\otimes}\Big[\frac{1}{n}\sum_{k_1} g_{l_1}(X_{k_1})\frac{1}{h^d}K\Big(\frac{X_{k_1}-X_{j_1}}{h}\Big) - g_{l_1}(X_{j_1})\pi(X_{j_1})\Big] \\
\times & \Big[\frac{1}{n}\sum_{k_2} g_{l_2}(X_{k_2})\frac{1}{h^d}K\Big(\frac{X_{k_2}-X_{j_1}}{h}\Big) - g_{l_2}(X_{j_1})\pi(X_{j_1})\Big] \\
= & \Big[E^{\otimes}\Big(\frac{1}{n}\sum_{k_1} g_{l_1}(X_{k_1})\frac{1}{h^d}K\Big(\frac{X_{k_1}-X_{j_1}}{h}\Big) - g_{l_1}(X_{j_1})\pi(X_{j_1})\Big)\Big]^2 \\
= & O(h^{2r})
\end{aligned}$$

and Lemma A.1 yields for the corresponding sum over $i_1 = i_2 < j_1 < j_2 < k_1 < k_2$

$$\begin{aligned}
& \Big|\frac{1}{n^4(n-1)^2}\sum E(\mathcal{S}_{n,i_1i_1j_1j_2k_1k_2})\Big| \\
\leq & \frac{1}{n^4(n-1)^2}\Big[\sum\Big|E^{\otimes}(\mathcal{S}_{n,i_1i_1j_1j_2k_1k_2})\Big| + \sum S_n^{\frac{1}{1+\varepsilon}}\Big|\beta^{\frac{\varepsilon}{1+\varepsilon}}(j_1-i_1)+\ldots+\beta^{\frac{\varepsilon}{1+\varepsilon}}(k_2-k_1)\Big|\Big] \\
= & O(n^{-1}h^{2r}) + o(n^{-2}h^{-d}) = o(n^{-1}).
\end{aligned}$$

All other terms on the right hand side of (A.12) are treated similarly, which completes the proof of Lemma 2.2.

Proof of Lemma 2.4

Recalling the definition of the kernel h_n in (2.20) we obtain for $v = (x, e)$

$$\begin{aligned}
E[h_n(v, V_j)] &= E\Big[h^{-d}K\Big(\frac{x-X_j}{h}\Big)\{e\Delta(X_j) + \Delta(x)\Delta(X_j)\} \\
&\quad - \sum_{l=1}^{p} g_l(x)(\Delta(x)+e)\Delta(X_j)g_l(X_j)\pi(X_j)\Big] \\
&= [e+\Delta(x)]\{(\Delta\pi)(x) - \mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi)(x)\} + O(h^r),
\end{aligned}$$

where the last line follows from a standard calculation observing the orthogonality of the regression functions $g_1, \ldots, g_p$ which implies

$$\mathbb{P}_{\{g_1,\ldots,g_p\}}(\Delta\pi) = \sum_{l=1}^{p}\langle\Delta\pi, g_l\rangle g_l.$$

Using the definiton $h_n^{(1)}(v) = E[h_n(v, V_j)] - \gamma_n$ the assertion of the lemma follows.

Acknowledgements. The authors would like to thank E. Liebscher for some constructive commens regarding Theorem 2.3. The work of H. Dette was supported by the SFB 475 (Komplexitätsreduktion in multivariaten Datenstrukturen).

References

Azzalini, A., Bowman, A. (1993). On the use of nonparametric regression for checking linear relationships. *J. R. Stat. Soc.*, Ser. B 55, No.2, 549-557.

Berger, J.O., Delampady, M. (1987). Testing precise hypotheses. *Statist. Sci.* 2, 317-352.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Stat. 27,* No. 3, 1012-1040.

Dette, H., Munk, A. (1998). Validation of linear regression models. *Ann. Stat. 26,* No.2, 778-800.

Doukhan, P. *Mixing. Properties and Examples.* Springer.

Fan, Y., Li, Q. (1999). Central limit theorem for degenerate U-Statistics of absolutely regular processes with applications to model specification testing. *Nonpar. Statistics* 10, 245-271.

Gonzalez-Manteiga, W., Cao-Abad, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test 2*, No.1-2, 161-188.

Hall, P., Marron, J.S. (1990). On variance estimation in nonparametric regression. *Biometrika* 77, No.2, 415-419.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Stat.* 21, No.4, 1926-1947.

Hjellvik, V., Tjøstheim, D. (1995). Nonparametric tests of linearity for time series. *Biometrika* 82, No.2, 351-368.

Hjellvik, V., Tjøstheim, D. (1996). Nonparametric statistics for testing of linearity and serial independence. *J. Nonparametric Stat.* 6, No.2-3, 223-251.

Hjellvik, V., Yao, Q., Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Stat. Plann. Inference* 68, No.2, 295-321.

Hong-zhi, A., Bing, C. (1991). A Kolmogorov-Smirnov Type statistic with application to test ofr nonlinearity in time series. *Int. Stat. Review* 59, 3, 287-307.

Koul, H., Stute, W. (1999). Nonparametric model checks for time series. *Ann. Stat.* 27, No.1, 204-236.

Kreiss, J. P., Neumann, M. H., Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. *Technical Report.* Universität Braunschweig.

Lee, A.J. (1990). *U-Statistics. Theory and Practice.* Marcel Dekker.

Liebscher, E. (1996). Central limit theorems for sums of α-mixing random variables. *Stochastics and Stochastic Reports* 59, 241-258.

Masry, E., Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series. *Econ. Theory* 11, 258-289.

Neil, J.W., Johnson, D.E. (1985), Testing linear regression function without replication. *Ann. Statist.* 13, 1482-1489.

Neumann, M., Kreiss, J. (1998). Regression-type inference in nonparametric autoregression. *Ann. Stat.* 26, No.4, 1570-1613.

Oodaira, H., Yoshihara, K. (1971). The law of the iterated logarithm for stationary processes satisfying mixing conditions. *Kodai Math. Sem. Rep.* 23, 311-334.

Robinson, P.M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* 4, 185-207.

Sellke, T., Bayarri, M.Y., Berger, J.O. (2001). Calibration of p values for testing precise hypotheses. *The Amer. Statist.* 50, 62-71.

Shillington, E.R. (1979). Testing lack of fit without replication. *Canad. J. Statist.* 7, 137-146.

Stone, C.J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* 13, 689-705.

Takahata, H., Yoshihara, K. (1987) Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences. *Yokohama Math. J.* 35, No.1/2, 95-111.

Tjøstheim, D. (1994). Nonlinear time series: A selective review. *Scand. J. Stat.* 21, No.2, 97-130.

Wand, M.P., Jones, M.C. (1995). *Kernel smoothing*. Chapman & Hall.

Zheng, J.X. (1996). A consistent test of a functional form via nonparametric estimation techniques. *J. of Econometrics* 75,263-289.