Neumann, Michael H.; Paparoditis, Efstathios

Working Paper

A nonparametric test for the stationary density

SFB 373 Discussion Paper, No. 1998,58

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

A nonparametric test for the stationary density

Michael H. Neumann
Sonderforschungsbereich 373
Humboldt-Universität zu Berlin
Spandauer Straße 1
D – 10178 Berlin
Germany

Efstathios Paparoditis
Department of Mathematics and Statistics
University of Cyprus
P.O. Box 537, Nicosia
Cyprus

1991 Mathematics Subject Classification. Primary 62M07; secondary 62G09, 62G10.
Keywords and Phrases. Bootstrap, stationary density, test, weak dependence.
Short title. Test for the stationary density.

We thank Rolf Tschernig for helpful comments on this paper. The research has been partially carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

Abstract. We propose a nonparametric test for checking parametric hypotheses about the stationary density of weakly dependent observations. The test statistic is based on the L_2-distance between a nonparametric and a smoothed version of a parametric estimate of the stationary density. It can be shown that this statistic behaves asymptotically as in the case of independent observations. Accordingly, we propose an i.i.d.-type bootstrap to determine the critical value for the test.

1. Introduction

Especially in the context of data from time series, statisticians very often fit certain parametric or semiparametric models. Parametric restrictions can be imposed for the dependence mechanism between subsequent observations and/or their marginal distribution. For example, people often assume normality – either directly for the observed random variables or for the unobserved innovations in structural time series models. For some of these models it is known that normality of the innovations also implies normality for the observed random variables. The adequacy of such strong assumptions is almost always debatable and some guidelines for assessing their appropriateness are of interest. In the present paper we develop a test which can be used to check certain parametric or semiparametric assumptions on the marginal distribution.

There already exists a lot of theory for tests in the context of independent, identically distributed observations. Classical approaches are based on a comparison of the assumed cumulative distribution function with its empirical counterpart and include well-known tests such as the Kolmogorov-Smirnov and the Cramér-von Mises test. More recently people also developed tests based on a comparison of the assumed density with a nonparametric estimate. In the context of i.i.d. observations, Bickel and Rosenblatt (1973) proposed a test based on the L_2-distance between a nonparametric density estimate and a parametric fit. Although methods based on the cumulative distribution function such as the Kolmogorov-Smirnov and the Cramér-von Mises test mentioned above are perhaps more popular among applied statisticians, both approaches have their relative advantages and disadvantages. The relative merits of smoothing-based tests based on local characteristics like densities versus non-smoothing tests based on cumulative characteristics are discussed by Rosenblatt (1975) and Ghosh and Huang (1991), in a different context. The essential message is that non-smoothing tests look primarily at global deviations, and are therefore well suited for detecting classical Pitman-alternatives of the form $f = f_0 + n^{-1/2} g$. On the other hand, smoothing-based tests focus on more localized deviations, and are consequently more powerful for detecting alternatives of the form $f = f_0 + n^{-\delta} g(./n^{-\gamma})$ for suitable $\delta, \gamma > 0$.

In the context of dependent data, the development of practicable tests becomes usually more difficult than in the independent case, since even the limit distribution of a potential test statistic depends on the dependence mechanism within the observations. In this respect, smoothing-based methods have another, perhaps unexpected advantage since it turns out that certain test statistics have the same limit distribution as in the case of i.i.d. observations. Whereas this effect is well-known for the *pointwise* behaviour of nonparametric estimators (see, e.g., Robinson (1983)), it seems to be much less known for statistics that depend through some nonparametric estimator on the whole sample. Takahata and Yoshihara (1987) showed for the special case of m-dependent observations that the integrated squared error of a nonparametric estimate of the stationary density has the same limit distribution as in the case of i.i.d. data. We will actually make use of the methodology developed in that paper for proving a central limit theorem in our slightly different situation. Inspired be the work of Härdle and Mammen (1993), we will focus on the L_2-distance between a nonparametric estimate and a smoothed version of a parametric estimate rather than the parametric estimate itself. Moreover, we will also relax the assumptions of Takahata and Yoshihara (1987) which in particular allows us to include the interesting case of testing the joint distribution of $(X_i, X_{i-l_1} \dots, X_{i-l_{d-1}})'$. There exists some related work on nonparametric tests which is also based on the possibility to neglect weak dependence. Theory for L_2-tests is developed in Paparoditis (1997) for the spectral density and in Kreiss, Neumann and Yao (1998) for the autoregression function. The case of supremum-type statistics that are needed for the construction of simultaneous confidence bands and L_∞-tests is investigated in Neumann and Kreiss (1997) in the context of nonparametric autoregressive models, and in Neumann (1996, 1997) in the more general framework of weakly dependent processes without any additional structural assumptions.

Although one could choose the critical value according to the limit distribution of the test statistic, we propose to use the bootstrap for its determination. According to our asymptotic theory, we employ Efron's (1979) bootstrap which was originally designed for i.i.d. observations. Some experience in related cases (e.g., simulations reported in Härdle and Mammen (1993)) let us expect that some suitable bootstrap method improves the accuracy of approximation provided by the limiting normal distribution. Although we do not have a rigorous proof for the superiority of the bootstrap over a first-order asymptotic approximation, some simulations reported in Section 3 of this paper seem to corroborate this conjecture.

2. Test statistics and their limit distributions

Throughout the whole paper we assume that we have observations from a stationary process $\{X_i, -\infty < i < \infty\}$. We do not impose any kind of *structural* conditions on the dependence mechanism such as, for example, some finite-order autoregressive structure. All we need is some appropriate kind of mixing condition and some assumption on the joint densities. We impose in particular the following conditions:

Assumption 1

Let, for $j \le k$, $\mathcal{F}_j^k = \sigma(X_j, X_{j+1}, \ldots, X_k)$. The coefficient of absolute regularity (β-mixing coefficient) is defined as

$$\beta(k) = E\left\{ \sup_{V \in \mathcal{F}_{i+k}^{\infty}} \left\{ |P(V \mid \mathcal{F}_{-\infty}^{i}) - P(V)| \right\} \right\}.$$

We suppose that the $\beta(k)$ decay with an exponential rate, that is

$$\beta(k) \le C \exp(-Ck).$$

Let f be the stationary density of the process and $f_{X_{i_1},\ldots,X_{i_m}}$ be the joint density of $(X_{i_1}, \ldots, X_{i_m})$.

Assumption 2

(i) f is continuous,

(ii) $\sup_{x_1,\ldots,x_m} \left\{ f_{X_{i_1},\ldots,X_{i_m}}(x_1, \ldots, x_m) \right\} < \infty$ for all m and $i_1 < \ldots < i_m$.

We study either the case of d-dimensional random variables X_i or the case of one-dimensional random variables X_i where we are interested in testing hypotheses on the joint density of $(X_i, X_{i-l_1}, \ldots, X_{i-l_{d-1}})'$. To unify our notation, we introduce random variables Y_i, where $Y_i = X_i$ in the first and $Y_i = (X_i, X_{i-l_1}, \ldots, X_{i-l_{d-1}})'$ in the latter case.

Tests for parametric or semiparametric hypotheses can be derived at different levels concerning the cardinality of the null hypothesis. All essential mathematical features can already be studied in the simplest case of a single null hypothesis, which is the object of the following subsection. Then we will briefly discuss some issues related to the practically more important case of a composite hypothesis.

2.1. Testing of single hypotheses. In order to present the essential mathematical ideas in an as clear as possible manner, we consider first the basic case of testing a single hypothesis, that is of

$$H_0 : \quad f = f_0.$$

Let

$$\widehat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - Y_i}{h}\right) \tag{2.1}$$

be a usual kernel estimator of f, where $h = h(n)$ denotes a bandwidth tending to 0 as $n \to \infty$. Our test statistic relates $\widehat{f}_n$ with the hypothetical density f_0. To avoid any kind of bias problems, we compare $\widehat{f}_n$ with a smoothed version of f_0. This leads to

$$T_n = nh^{d/2} \int \left[\widehat{f}_n(x) - (K_h * f_0)(x)\right]^2 dx, \tag{2.2}$$

where the smoothing operator K_h is defined by

$$(K_h * g)(\cdot) = \int h^{-d} K\left(\frac{\cdot - z}{h}\right) g(z)\, dz. \tag{2.3}$$

Before we state a theorem about the limit distribution of T_n, we introduce two more assumptions.

Assumption 3
K is bounded and compactly supported.

Assumption 4
(i) $h = o([\log(n)]^{-3})$,
(ii) $h^{-d} = o(n)$.

The asymptotic behaviour of statistics similar to T_n was already investigated by Takahata and Yoshihara (1987). They found in the special case of m-dependent observations that $T_n - ET_n$ converges to a normal distribution with the same variance as if the Y_i were independent. The following theorem provides a similar result under a different set of assumptions.

Theorem 2.1. *Suppose that Assumptions 1 to 4 are fulfilled. Then*

$$T_n \;-\; h^{-d/2} \int K^2(u)\, du \stackrel{d}{\longrightarrow} N(0, \sigma^2),$$

where

$$\sigma^2 \;=\; 2 \int f^2(x)\, dx \;\times \int \left[\int K(u) K(u+v)\, du \right]^2 dv.$$

The proof of this assertion is based on a central limit theorem for sums of dependent random variables due to Dvoretzky (1972) and follows in large parts the pattern of the proof of a similar assertion in Takahata and Yoshihara (1987). In order to provide a self-contained version of this paper, and since our technical assumptions are different from those in Takahata and Yoshihara (1987), we give a full proof in Section 4.

2.2. Testing of composite hypotheses. In this subsection we consider the perhaps more important case of testing composite hypotheses. Instead of a single null hypothesis, $f = f_0$, we have now

$$H_0: \quad f \in \mathcal{F},$$

where $\mathcal{F}$ is some parametric or even semiparametric class of density functions. It will turn out that, under suitable regularity conditions on the class $\mathcal{F}$, the problem can be reduced to the case of a single hypothesis investigated in the previous subsection. Practitioners are probably most interested in testing (finite-dimensional) parametric hypotheses, that is $\mathcal{F} = \mathcal{F}^\Theta = \{f_\theta, \theta \in \Theta\}$, where $\Theta \subseteq \mathbb{R}^d$. We will study this case in some detail, and will discuss the semiparametric problem of testing independence of certain components of Y_i briefly at the end of this section.
In the case of $f \in \mathcal{F}^\Theta$, let $\theta_0 \in \Theta$ be such that $f_{\theta_0} = f$. Our test will be based on the L_2-distance between our nonparametric estimate $\widehat{f}_n$ and a smoothed version

of a parametric fit, $f_{\widehat{\theta}}$, namely

$$T_{n,\widehat{\theta}} = nh^{d/2} \int \left[\widehat{f}_n(x) - (K_h * f_{\widehat{\theta}})(x) \right]^2 dx, \tag{2.4}$$

where K_h is the smoothing operator defined by (2.3). By looking at

$$\begin{aligned} T_{n,\widehat{\theta}} - T_n &= 2nh^{d/2} \int \left[\widehat{f}_n(x) - (K_h * f_{\theta_0})(x) \right] \left[(K_h * f_{\theta_0})(x) - (K_h * f_{\widehat{\theta}})(x) \right] dx \\ &\quad + nh^{d/2} \int \left[(K_h * (f_{\widehat{\theta}} - f_{\theta_0}))(x) \right]^2 dx \end{aligned}$$

it is easy to find sufficient conditions for the asymptotic equivalence of $T_{n,\widehat{\theta}}$ and T_n. To formulate such a set of conditions, we write f_θ in the form

$$f_\theta(x) = f_{\theta_0}(x) + (\theta - \theta_0) f'_{\theta_0}(x) + R(\theta, \theta_0, x).$$

In the following we will assume:

Assumption 5

(i) $\int \left[(K_h * f_{\widehat{\theta}})(x) - (K_h * f_{\theta_0})(x) \right]^2 dx = o_P(n^{-1} h^{-d/2})$,

(ii) $\widehat{\theta} - \theta_0 = o_P(n^{-1/2} h^{-d/2})$,

(iii) $\sup_x \{ |f'_{\theta_0}(x)| \} < \infty$,

(iv) $\int R^2(\widehat{\theta}, \theta_0, x)\, dx = o_P(n^{-1})$.

It is easy to see that Assumptions 1 and 3 and (ii)-(iv) of Assumption 5 imply that

$$\begin{aligned} &\left| \int \left[\widehat{f}_n(x) - (K_h * f_{\theta_0})(x) \right] \left[(K_h * f_{\theta_0})(x) - (K_h * f_{\widehat{\theta}})(x) \right] dx \right| \\ &= O_P\left(\frac{1}{nh^d} (\widehat{\theta} - \theta_0) \sqrt{ \operatorname{var}\left(\sum_{i=1}^n \int \left[K(\frac{x - Y_i}{h}) - EK(\frac{x - Y_1}{h}) \right] (K_h * f'_{\theta_0})(x)\, dx \right) } \right) \\ &\quad + O\left(\sqrt{ \int \left[\widehat{f}_n(x) - (K_h * f_{\theta_0})(x) \right]^2 dx } \sqrt{ \int R^2(\widehat{\theta}, \theta_0, x)\, dx } \right) \\ &= o_P(n^{-1} h^{-d/2}). \end{aligned} \tag{2.5}$$

This leads immediately to the following theorem:

Theorem 2.2. *Suppose that Assumptions 1 to 5 are fulfilled. Then*

$$T_{n,\widehat{\theta}} - h^{-d/2} \int K^2(u)\, du \xrightarrow{d} N(0, \sigma^2).$$

Remark 1. In cases where Assumption 5 is not satisfied, one may still construct a conservative test based on theory developed for T_n. In this case we may consider

$$T_{n,inf} = \inf_{\theta\in\Theta}\left\{\int\left[\widehat{f}_n(x) - (K_h * f_\theta)(x)\right]^2 dx\right\}. \tag{2.6}$$

If $f \in \mathcal{F}^\Theta$, it follows immediately that

$$P\left(T_{n,inf} \geq t\right) \leq P\left(T_n \geq t\right).$$

Hence, we can apply the asymptotic theory given in Theoren 2.1 or the bootstrap approximation proposed in the next section to construct a test which has an asymptotic error of the first kind not larger than α.

Remark 2. It seems also possible to develop a test of independence of certain components of Y_i in complete analogy to a proposal of Rosenblatt (1975) in the independent case. To be more specific, for testing independence of the two components Y_{i1} and Y_{i2} of $Y_i = (Y_{i1}', Y_{i2}')'$, one might use the statistic $\int[\widehat{f}_n(x) - \widehat{f}_{1n}(x^{(1)})\widehat{f}_{2n}(x^{(2)})]^2 dx$, where $\widehat{f}_{1n}$ and $\widehat{f}_{2n}$ are kernel estimators of f_1 and f_2, respectively, and $f(x) = f_1(x^{(1)})f_2(x^{(2)})$ under H_0. We expect that this test statistic has the same limit distribution as given in Theorem 2 of Rosenblatt (1975) in the independent case.

3. Bootstrapping the test statistic

The theoretical results of the previous section motivate the use of bootstrap methods similar to that designed for the i.i.d. case in order to approximate the distribution of both test statistics considered. In fact Theorem 2.1 and Theorem 2.2 suggest that in order to get an asymptotically correct estimator of the distributions of these statistics it is not necessary to reproduce the whole (and probably very complicated) dependence structure of the stochastic process generating the observations. We stress here the fact that the theorems obtained are based on asymptotic considerations, i.e., we expect that for finite sample sizes n such a simple bootstrap procedure which neglects the dependence in the data will lead to valuable approximations only if the smoothing bandwidth h is small enough and the dependence of the data weak enough. Since we focus our considerations primarily to the error probability of the first type, it suffices to provide a consistent estimator of the distribution of the test statistics under the null hypothesis. On the other hand, since one is of course interested in a good power performance, we should also approximate (one of the) distributions corresponding to the null if the true distribution does not correspond to the hypothesis. Hence, we should not use resampling with replacement from the observations $Y_1, Y_2, \ldots, Y_n$. Rather, we generate independent bootstrap resamples $Y_1^*, Y_2^*, \ldots, Y_n^*$ according to the density $f_{\widehat{\theta}}$.

3.1. Bootstrap approximations. Consider first the case of testing a composite hypothesis, i.e., the case where $f = f_{\theta_0}$. In order to ensure that certain random integrals convergence in probability to the correct limits as $\widehat{\theta} \to \theta_0$, the following additional assumptions are imposed on the parametric density estimate $f_{\widehat{\theta}}$.

Assumption 6

(i) $\sup_x \{f_{\widehat{\theta}}(x)\} = O_P(1)$,

(ii) $\int \left[f_{\widehat{\theta}}(x) - f_{\theta_0}(x)\right]^2 dx = o_P(1)$.

The bootstrap procedure proposed in this case can then be described as follows. Let Y_i^*, $i = 1, 2, \ldots, n$, be a random sample from $f_{\widehat{\theta}}$ and $\hat{f}_n^*(x)$ be a kernel estimator of $f_{\widehat{\theta}}$ defined by

$$\hat{f}_n^*(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - Y_i^*}{h}\right). \tag{3.1}$$

In view of the equivalence of $T_{n,\widehat{\theta}}$ and T_n, it suffices to imitate the statistic T_n, i.e., we consider the bootstrap statistic

$$T_n^* = nh^{d/2} \int \left[\hat{f}_n^*(x) - (K_h * f_{\widehat{\theta}})(x)\right]^2 dx. \tag{3.2}$$

The following theorem justifies theoretically the use of the statistic T_n^* in order to approximate the distribution of T_n and, therefore, also of $T_{n,\widehat{\theta}}$. It enables us to use the quantiles of this distribution in order to carry out the test procedure.

Theorem 3.1. *Suppose that Assumptions 1, 3, 4 and 6 are fulfilled. Then we have conditionally on* $Y_1, Y_2, \ldots, Y_n$

$$T_n^* - h^{-d/2} \int K^2(u)\, du \xrightarrow{d} N(0, \sigma^2) \quad \textit{in probability.}$$

One could of course also directly approximate the distribution of $T_{n,\widehat{\theta}}$ by the distribution of the bootstrap statistic $T^*_{n,\widehat{\theta}^*}$ where the latter is defined by

$$T^*_{n,\widehat{\theta}^*} = nh^{d/2} \int \left[\hat{f}_n^*(x) - (K_h * f_{\widehat{\theta}^*})(x)\right]^2 dx. \tag{3.3}$$

In the above expression $f_{\widehat{\theta}^*}$ denotes the estimated parametric fit obtained using the bootstrap sample $Y_1^*, Y_2^*, \ldots, Y_n^*$.
The validity of this method follows from Theorem 2.2 and Theorem 3.1 if the difference between T_n^* and $T^*_{n,\widehat{\theta}^*}$ is asymptotically negligible. To be more specific, we need the fact that with an increasing probability the bootstrap distributions of T_n^* and $T^*_{n,\widehat{\theta}^*}$ are close to each other, i.e., for arbitrary $\varepsilon > 0$ we would like to have

$$E\left[P^*\left(|T_n^* - T^*_{n,\widehat{\theta}^*}| > \varepsilon \middle| Y_1, \ldots, Y_n\right)\right] = o(1).$$

This is conveniently expressed by the following assumption on the *unconditional* probability:

$$P\left(|T_n^* - T^*_{n,\widehat{\theta}^*}| > \varepsilon\right) = o(1). \tag{3.4}$$

In analogy to Assumption 5, this is ensured by the following assumption:

Assumption 7

$$\text{(i)} \quad \int \Big[(K_h * f_{\widehat{\theta}^*})(x) - (K_h * f_{\widehat{\theta}})(x)\Big]^2 dx = o_P(n^{-1}h^{-d/2})$$
$$\text{(ii)} \quad \widehat{\theta}^* - \widehat{\theta} = o_P(n^{-1/2}h^{-d/2}),$$
$$\text{(iii)} \quad \sup_x \Big\{|f'_{\widehat{\theta}}(x)|\Big\} = O_P(1),$$
$$\text{(iv)} \quad \int R^2(\widehat{\theta}^*, \widehat{\theta}, x)dx = o_P(n^{-1}),$$

where o_P and O_P refer here to the *joint* distribution of $(Y_1, \ldots, Y_n)$ and $(Y_1^*, \ldots, Y_n^*)$.

As an immediate corollary to Theorem 3.1 we get

Corollary 3.1. *Suppose that Assumptions 1, 3, 4, 6 and 7 are fulfilled. Then we have conditionally on* $Y_1, Y_2, \ldots, Y_n$

$$T^*_{n,\widehat{\theta}^*} - h^{-d/2} \int K^2(u)\, du \xrightarrow{d} N(0, \sigma^2) \quad \textit{in probability.}$$

Consider next the case of testing a single hypothesis, i.e., the case $f = f_0$. Since in this case the distribution of Y_i is completely known, the appropriate bootstrap statistic is given by

$$T_n^* = nh^{d/2} \int \Big[\hat{f}_n^*(x) - (K_h * f_0)(x)\Big]^2 dx \tag{3.5}$$

where $\hat{f}_n^*(x)$ is defined as in (3.1) and the Y_i^*'s are now i.i.d. samples from the known density f_0. The following theorem can then be established. Its proof follows exactly the same lines as the proof of Theorem 2.1.

Theorem 3.2. *Suppose that Assumptions 2(i), 3 and 4 are fulfilled. Then conditionally on* $Y_1, Y_2, \ldots, Y_n$

$$T_n^* - h^{-d/2} \int K^2(u)\, du \xrightarrow{d} N(0, \sigma^2).$$

3.2. Simulated examples. The theory of the previous section justifies asymptotically the use of the proposed bootstrap procedure in order to approximate the distribution the test statistic considered. In this section we study the finite sample performance of the bootstrap by means of a small simulation experiment. For this realizations of length $n = 200$ have been generated from the first order autoregressive process $X_t = \phi X_{t-1} + \varepsilon_t$, where ε_t is an i.i.d. sequence with $\varepsilon_t \sim N(0, \sqrt{1-\phi^2})$ and the autoregressive parameter ϕ takes its values in the set $\{0, \pm 0.4, 0.8\}$. Note that for $\phi = 0$ we are in the i.i.d. setting, i.e., our test is identical to the test proposed by Bickel and Rosenblatt (1973). The case $\phi = \pm 0.4$ corresponds to a 'rather moderate' dependence while $\phi = 0.8$ to a 'rather strong' dependence in the data. The null hypothesis is that of Gaussian distribution with unknown mean and variance. The test statistic $T_{n,\hat{\theta}}$ has been calculated using the kernel $K(x) = (2\sqrt{3})^{-1} I(-\sqrt{3} \le x \le \sqrt{3})$ for which some optimality properties has been derived in the testing context considered here; cf. Ghosh and Huang (1991). The smoothing bandwidth has been set equal to $h = 0.03$. To estimate the exact distribution of $T_{n,\hat{\theta}}$, 1000 replications of the model considered have been used while the bootstrap approximations are based on 1000 samples.

Please insert Figure 1 and Figure 2 about here

Figure 1 and Figure 2 show the simulated exact densities and three bootstrap estimates of these densities based on different original time series. In each case the estimated exact density of $T_{n,\hat{\theta}}$ as well as the densities of the corresponding bootstrap approximations shown in these exhibits have been obtained using the Gaussian smoothing kernel and a bandwidth selection according to Silverman's rule. Finally, to make some comparisons with the asymptotic Gaussian approximation, we have plotted in these figures also the corresponding Gaussian densities. As these exhibits show the asymptotic Gaussian distribution is a poor approximation to the (estimated) exact one. Furthermore, for small and moderate dependence the bootstrap approximations are more satisfactorily improving upon the Gaussian approximation and reproducing more closely the overall behavior and the skewness of the (estimated) exact density. Only in the case $\phi = 0.8$ with a rather strong positive dependence in the data the bootstrap approximations become worse. Clearly, we expect that in this case other bootstrap approaches like the block bootstrap which explicitly takes into account the dependence structure of the data, will lead to better results.

4. Proofs

Proof of Theorem 2.1. According to a well-known theorem of Brown (1971), one can derive a central limit theorem for statistics that can be written as a sum of an increasing number of martingale differences that satisfy an asymptotic negligibility condition. Dvoretzky (1972) extended this result to statistics that form such a scheme only approximately, which is of particular importance in the context of weakly dependent random variables. Before we begin with checking the conditions of Dvoretzky's theorem, we decompose T_n in such a way that the leading term satisfies just these conditions while the remaining terms are of negligible order. Our proof follows essentially the same pattern as a proof of a similar assertion in Takahata and Yoshihara (1987).
First we write T_n in the form

(4.1)
$$T_n = \sum_{1 \leq i < j \leq n} H_n(Y_i, Y_j) + \left[\frac{1}{2} \sum_{i=1}^{n} H_n(Y_i, Y_i) - h^{-d/2} \int K^2(u)\, du \right],$$

where

$$\begin{aligned} H_n(x, y) &= \frac{2}{nh^{3d/2}} \int \left[K\left(\frac{u - x}{h}\right) - EK\left(\frac{u - Y_1}{h}\right) \right] \\ &\qquad \times \left[K\left(\frac{u - y}{h}\right) - EK\left(\frac{u - Y_1}{h}\right) \right] du. \end{aligned} \tag{4.2}$$

The proof of the desired central limit theorem for T_n is facilitated by using a decomposition of $Y_1, \ldots, Y_n$ into an alternating sequence of large and small blocks. The gaps between the large blocks are of length $\rho_n = [C_1 \log(n)]$, where an appropriate choice of C_1 becomes clear from the calculations below. The length of the large blocks is denoted by l_n, where the only requirement is that $\rho_n \ll l_n \ll n$. In accordance with this, the k-th large block is formed by $Y_{a_k}, Y_{a_k+1}, \ldots, Y_{b_k}$, where $a_k = (k-1)(l_n+\rho_n)+1$ and $b_k = [(k-1)(l_n+\rho_n)+l_n] \wedge n$, while the k-th small block is given by $Y_{b_k+1}, \ldots, Y_{a_k-1}$.
Now we approximate T_n by

$$U_n = \sum_k S_k, \tag{4.3}$$

where

$$S_k = \sum_{i=1}^{b_{k-1}} \sum_{j=a_k}^{b_k} H_n(Y_i, Y_j). \tag{4.4}$$

(i) *Central limit theorem for U_n*

Let $\mathcal{G}_k = \sigma(Y_1, \ldots, Y_{b_k})$. In what follows we will show that

$$\sum_k E(S_k \mid \mathcal{G}_{k-1}) \xrightarrow{P} 0, \tag{4.5}$$

$$\sum_k E(S_k^2 \mid \mathcal{G}_{k-1}) \xrightarrow{P} \sigma^2, \tag{4.6}$$

and, for each $\epsilon > 0$,

$$\sum_k E[S_k^2 I(|S_k| > \epsilon)] \xrightarrow{P} 0. \tag{4.7}$$

Then we obtain, according to Theorem 2 of Dvoretzky (1972), that

$$U_n \xrightarrow{d} N(0, \sigma^2). \tag{4.8}$$

Now we turn to the proofs of (4.5) to (4.7).

(i.a) *Proof of (4.5)*

Let $i \leq b_{k-1}$ and $a_k \leq j \leq b_k$. We obtain by Theorem 1.2.3 in Yoshihara (1994) [This is the correct formulation of Lemma 2 in Takahata and Yoshihara (1987), which contained some typos.] that

$$E\left| E\left(H_n(Y_i,Y_j) \mid \mathcal{G}_{k-1}\right) - \int\int H_n(y_i,y_j)\, dP^{Y_i}(y_i)\, dP^{Y_j}(y_j) \right| \leq 2 \sup_{x,y}\{|H_n(x,y)|\} \beta(\rho_n+1).$$

Since $EH_n(x, Y_1) = 0$, this yields

$$E\left| \sum_k E(S_k \mid \mathcal{G}_{k-1}) \right| = o(1),$$

provided we choose C_1 in the definition of ρ_n large enough. This implies (4.5).

(i.b) *Proof of (4.6)*

Define

$$G_k(x,y) \;=\; \sum_{j,j'=a_k}^{b_k} E^{Y_j,Y_{j'}} H_n(x,Y_j)H_n(y,Y_{j'}). \tag{4.9}$$

Let $\widetilde{Y}_1,\ldots,\widetilde{Y}_n$ be independent random variables with common density f. To derive an approximation for the conditional variances, we split up

$$\begin{aligned}
&E\left|E(S_k^2 \mid \mathcal{G}_{k-1}) \;-\; \sum_{i=1}^{b_{k-1}}\sum_{j=a_k}^{b_k} E[H_n(\widetilde{Y}_i,\widetilde{Y}_j)]^2\right| \\
&\quad\le\; \sum_{i,i'=1}^{b_{k-1}} E\left|\sum_{j,j'=a_k}^{b_k} E\left(H_n(Y_i,Y_j)H_n(Y_{i'},Y_{j'}) \mid \mathcal{G}_{k-1}\right) \;-\; G_k(Y_i,Y_{i'})\right| \\
&\qquad +\, E\left|\sum_{i,i'=1}^{b_{k-1}} [G_k(Y_i,Y_{i'}) \;-\; EG_k(Y_i,Y_{i'})]\right| \\
&\qquad +\left|\sum_{i,i'=1}^{b_{k-1}} EG_k(Y_i,Y_{i'}) \;-\; \sum_{i=1}^{b_{k-1}}\sum_{j=a_k}^{b_k} E[H_n(\widetilde{Y}_i,\widetilde{Y}_j)]^2\right| \\
&\quad=\; R_1 \,+\, R_2 \,+\, R_3.
\end{aligned} \tag{4.10}$$

Now we have, according to Theorem 1.2.3 in Yoshihara (1994),

$$\begin{aligned}
&\sup_{1\le i,i'\le b_{k-1}} \left\{E\left|\sum_{j,j'=a_k}^{b_k} E\left(H_n(Y_i,Y_j)H_n(Y_{i'},Y_{j'}) \mid \mathcal{G}_{k-1}\right) \;-\; G_k(Y_i,Y_{i'})\right|\right\} \\
&\qquad\qquad =\; 2(b_k-a_k+1)^2 \sup_{x,y}\{|H_n(x,y)|^2\}\beta(\rho_n+1),
\end{aligned}$$

which implies in particular

$$R_1 \;=\; o(l_n/n), \tag{4.11}$$

provided C_1 is chosen sufficiently large.
Before we turn to an estimate of R_2, we derive some useful estimates for $G_k(x,y)$. Since f is bounded, we get $\sup_u\{EK((u-Y_1)/h) = O(h^d)\}$, which implies

$$\begin{aligned}
&E\left\{\int \left[K\left(\frac{u-x}{h}\right) \,-\, EK\left(\frac{u-Y_1}{h}\right)\right]\left[K\left(\frac{u-Y_j}{h}\right) \,-\, EK\left(\frac{u-Y_1}{h}\right)\right] du\right. \\
&\qquad \left.\times \int\left[K\left(\frac{u'-y}{h}\right) \,-\, EK\left(\frac{u'-Y_1}{h}\right)\right]\left[K\left(\frac{u'-Y_{j'}}{h}\right) \,-\, EK\left(\frac{u'-Y_1}{h}\right)\right] du'\right\} \\
&\quad=\; E\left\{\int K\left(\frac{u-x}{h}\right)K\left(\frac{u-Y_j}{h}\right) du \times \int K\left(\frac{u'-y}{h}\right)K\left(\frac{u'-Y_{j'}}{h}\right) du'\right\} \,+\, O(h^{4d}).
\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
&|G_k(x,y)| \\
&\le \frac{4}{n^2h^{3d}} \sum_{j=a_k}^{b_k} \Big[E\Big| \int K\Big(\frac{u-x}{h}\Big) K\Big(\frac{u-Y_j}{h}\Big)\,du \\
&\qquad\qquad \times \int K\Big(\frac{u'-y}{h}\Big) K\Big(\frac{u'-Y_j}{h}\Big)\,du'\Big| + O(h^{4d})\Big] \\
&\quad + \frac{4}{n^2h^{3d}} \sum_{\substack{a_k\le j,j'\le b_k\\ 1\le |j-j'|<C_2\log(n)}} \Big[E\Big| \int K\Big(\frac{u-x}{h}\Big) K\Big(\frac{u-Y_j}{h}\Big)\,du \\
&\qquad\qquad \times \int K\Big(\frac{u'-y}{h}\Big) K\Big(\frac{u'-Y_{j'}}{h}\Big)\,du'\Big| + O(h^{4d})\Big] \\
&\quad + \sum_{\substack{a_k\le j,j'\le b_k\\ |j-j'|\ge C_2\log(n)}} |EH_n(x,Y_j)H_n(y,Y_{j'})| \\
&= O\Big(\frac{1}{n^2h^{3d}} l_n h^{3d}\Big) \\
&\quad + O\Big(\frac{1}{n^2h^{3d}} l_n \log(n) h^{3d+1}\Big) \\
&\quad + O\Big(l_n^2 \sup_{x,y}\{|H_n(x,y)|^2\}\beta(C_2\log(n))\Big)
\end{aligned}
$$

$$
= O(n^{-2}l_n), \tag{4.12}
$$

if C_2 is sufficiently large. [The upper bound of order $O(n^{-2}l_n\log(n)h)$ follows by the fact that $\sup_{x,y}\{P(\|Y_j-x\|\le Ch, \|Y_{j'}-y\|\le Ch)\} = O(h^{d+1})$ holds for $j\ne j'$, which is a consequence of Assumption 2(ii).]
By analogous considerations as in (4.12) we get, for $i\ne i'$,

$$
\begin{aligned}
&E\,|G_k(Y_i,Y_{i'})| \\
&\le \frac{4}{n^2h^{3d}} \sum_{j=a_k}^{b_k} E\Big| \int \Big[\int K\Big(\frac{u-Y_i}{h}\Big) K\Big(\frac{u-z}{h}\Big)\,du\Big] \\
&\qquad\qquad \times \Big[\int K\Big(\frac{u'-Y_{i'}}{h}\Big) K\Big(\frac{u'-z}{h}\Big)\,du'\Big] f^{Y_j}(z)\,dz\Big| \\
&\quad + \frac{4}{n^2h^{3d}} \sum_{\substack{a_k\le j,j'\le b_k\\ 1\le |j-j'|<C_2\log(n)}} E\Big| \int\int \Big[\int K\Big(\frac{u-Y_i}{h}\Big) K\Big(\frac{u-z}{h}\Big)\,du\Big] \\
&\qquad\qquad \times \Big[\int K\Big(\frac{u'-Y_{i'}}{h}\Big) K\Big(\frac{u'-z'}{h}\Big)\,du'\Big] f^{Y_j,Y_{j'}}(z,z')\,dz\,dz'\Big| \\
&\quad + O\Big(\frac{1}{n^2h^{3d}} l_n \log(n) h^{4d}\Big) + O\Big(l_n^2 \sup_{x,y}\{|H_n(x,y)|^2\}\beta(C_2\log(n))\Big)
\end{aligned}
$$

$$
= O\Big(n^{-2}l_n\log(n)h\Big). \tag{4.13}
$$

We consider the index sets

$$\mathcal{J}_1 \;=\; \left\{(i_1,\dots,i_4)\in\{1,\dots,b_{k-1}\}^4 \,\middle|\; i_j\neq i_k \text{ for some } j,k;\quad \max_{k=1,\dots,4}\min_{j\neq k}\{|i_k-i_j|\} \,<\, C_3\log(n)\right\},$$

and

$$\mathcal{J}_2 \;=\; \left\{(i_1,\dots,i_4)\in\{1,\dots,b_{k-1}\}^4 \,\middle|\; \max_{k=1,\dots,4}\min_{j\neq k}\{|i_k-i_j|\} \,\geq\, C_3\log(n)\right\}.$$

For $(i_1,\dots,i_4)\in\mathcal{J}_1$ (W.l.o.g., we assume that $i_1\neq i_2$.), then we can find the following estimate:

$$|EG_k(Y_{i_1},Y_{i_2})G_k(Y_{i_3},Y_{i_4})| \,\leq\, \sup_{x,y}\{|G_k(x,y)|\}\,E|G_k(Y_{i_1},Y_{i_2})| \,\leq\, O\left(n^{-4}l_n^2\log(n)h\right).$$

If $(i_1,\dots,i_4)\in\mathcal{J}_2$, then we obtain by $EG_k(x,Y_1)=0$ the estimate

$$|EG_k(Y_{i_1},Y_{i_2})G_k(Y_{i_3},Y_{i_4})| \,\leq\, 2\sup_{x,y}\left\{|G_k(x,y)|^2\right\}\beta(C_3\log(n)) \;=\; O(n^{-7}),$$

say, if C_3 is sufficiently large. These two estimates imply

$$\begin{aligned}
\operatorname{var}\left(\sum_{i,i'=1}^{b_{k-1}} G_k(Y_i,Y_{i'})\right) \;&\leq\; E\left|\sum_{i,i'=1}^{b_{k-1}} G_k(Y_i,Y_{i'})\right|^2 \\
&=\; \sum_{i=1}^{b_{k-1}} E|G_k(Y_i,Y_i)|^2 \\
&\quad + \sum_{(i_1,\dots,i_4)\in\mathcal{J}_1} EG_k(Y_{i_1},Y_{i_2})G_k(Y_{i_3},Y_{i_4}) \\
&\quad + \sum_{(i_1,\dots,i_4)\in\mathcal{J}_2} EG_k(Y_{i_1},Y_{i_2})G_k(Y_{i_3},Y_{i_4}) \\
&=\; O(n^{-3}l_n^2) \,+\, O\left(\#\mathcal{J}_1 n^{-4}l_n^2\log(n)h\right) \,+\, O\left(\#\mathcal{J}_2 n^{-7}\right) \\
&=\; O\left(n^{-3}l_n^2 \,+\, n^{-2}l_n^2(\log(n))^3h\right). \qquad (4.14)
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
R_2 \;&=\; O\left(\sqrt{\operatorname{var}\left(\sum_{i,i'=1}^{b_{k-1}} G_k(Y_i,Y_{i'})\right)}\right) \\
&=\; O\left(n^{-3/2}l_n \,+\, n^{-1}l_n(\log(n))^{3/2}h^{1/2}\right). \qquad (4.15)
\end{aligned}$$

According to (4.12) we obtain

$$\begin{aligned}
EG_k(Y_i,Y_i) \;&=\; \sum_{j=a_k}^{b_k} E[H_n(Y_i,Y_j)]^2 \\
&\quad + O\left(\frac{1}{n^2h^{3d}}l_n\log(n)h^{3d+1} \,+\, l_n^2\sup_{x,y}\{|H_n(x,y)|^2\}\beta(C_2\log(n))\right) \\
&=\; \sum_{j=a_k}^{b_k} E[H_n(\widetilde{Y}_i,\widetilde{Y}_j)]^2 \,+\, O\left(n^{-2}l_n\log(n)h\right). \qquad (4.16)
\end{aligned}$$

This implies, in conjunction with (4.12) and (4.13), that

$$\begin{aligned} R_3 \;&\leq\; \sum_{i=1}^{b_{k-1}} \left| EG_k(Y_i, Y_i) \;-\; \sum_{j=a_k}^{b_k} E[H_n(\widetilde{Y}_i, \widetilde{Y}_j)]^2 \right| \\ &\quad + \sum_{1 \leq |i-i'| < C_4 \log(n)} E|G_k(Y_i, Y_{i'})| \\ &\quad + \sum_{|i-i'| \geq C_4 \log(n)} E|G_k(\widetilde{Y}_i, \widetilde{Y}_{i'})| \;+\; 2 \sup_{x,y}\{|G_k(x,y)|\} \beta(C_4 \log(n)) \\ &= O\left(n^{-1} l_n \log(n) h\right) \\ &\quad + O\left(n^{-1} l_n (\log(n))^2 h\right) \\ &\quad + O\left(n^{-1}\right) \\ &= O\left(n^{-1} l_n (\log(n))^2 h\right). \end{aligned} \tag{4.17}$$

Since f is continuous and bounded, we easily obtain that

$$\begin{aligned} &E[H_n(\widetilde{Y}_i, \widetilde{Y}_j)]^2 \\ &= \frac{4}{n^2 h^{3d}} \int \left[\int K(u) K(u+v)\, du \right]^2 dv \times \int f^2(x)\, dx \;+\; o\left(\frac{1}{n^2 h^{3d}} \right). \end{aligned} \tag{4.18}$$

Moreover, because of

$$\sum_k \#\left\{(i,j) \mid 1 \leq i \leq b_{k-1}, \quad a_k \leq j \leq b_k\right\} \;=\; n^2/2 \;+\; o(n^2), \tag{4.19}$$

we obtain by (4.10), (4.11), (4.15), (4.17) and (4.18) that

$$E \left| \sum_k E(S_k \mid \mathcal{G}_{k-1}) \;-\; \sigma^2 \right| \;=\; o(1),$$

which implies (4.6).

(i.c) *Proof of (4.7)*

Using Berbee's reconstruction lemma and (4.14) we obtain

$$\begin{aligned} ES_k^4 \;&=\; E \left| \sum_{i,i'=1}^{b_{k-1}} G_k(Y_i, Y_{i'}) \right|^2 \;+\; O\left((l_n h^{-d/2})^4 \beta(\rho_n + 1) \right) \\ &=\; O\left(n^{-3} l_n \;+\; n^{-2} l_n^2 (\log(n))^3 h \right), \end{aligned} \tag{4.20}$$

which implies (4.7) because of

$$\sum_k E[S_k^2 I(|S_k| > \epsilon)] \;\leq\; \epsilon^{-4} \sum_k ES_k^4 \;=\; O\left(n^{-2} \;+\; n^{-1} l_n (\log(n))^3 h \right).$$

(ii) *Difference between T_n and U_n*

We have

$$
\begin{aligned}
&\left[T_n \,-\, h^{-d/2} \int K^2(u)\,du\right] \,-\, U_n \\
&= \;\frac{1}{2}\sum_{i=1}^{n} H_n(Y_i, Y_i) \;-\; \frac{1}{nh^{d/2}} \int K^2(u)\,du \\
&\quad + \sum_k \sum_{i=1}^{b_{k-1}-\rho_n} \sum_{j=b_{k-1}+1}^{a_k-1} H_n(Y_i, Y_j) \\
&\quad + \sum_k \sum_{j=b_{k-1}+1}^{a_k-1} \sum_{i=b_{k-1}-\rho_n+1}^{j-1} H_n(Y_i, Y_j) \\
&\quad + \sum_k \sum_{j=a_k}^{b_k} \sum_{i=b_{k-1}+1}^{j-1} H_n(Y_i, Y_j) \\
&= \; R_4 \,+\, \ldots \,+\, R_7. \qquad (4.21)
\end{aligned}
$$

It is easy to see that

$$
EH_n(Y_i, Y_i) \;=\; \frac{2}{nh^{d/2}} \int K^2(u)\,du. \qquad (4.22)
$$

Since $\sup_x\{|H_n(x,x)|\} \,=\, O(n^{-1}h^{-d/2})$, we obtain by Lemma 1.2.2.3 in Doukhan (1994) that

$$
\operatorname{var}\left(\sum_{i=1}^{n} H_n(Y_i, Y_i)\right) \;=\; O\left(n^{-1}h^{-d}\right), \qquad (4.23)
$$

which implies, in conjunction with (4.22), that

$$
R_4 \;=\; o_P(1). \qquad (4.24)
$$

Next, observe that R_5 has the same structure as U_n – with the only difference that the large blocks $Y_{a_k}, \ldots, Y_{b_k}$ are replaced by small blocks $Y_{b_{k-1}+1}, \ldots, Y_{a_k-1}$. Hence, it is clear that

$$
R_5 \;=\; o_P(1). \qquad (4.25)
$$

Notice that $R_6 + R_7$ can be written as

$$
R_6 \,+\, R_7 \;=\; \sum_{(i,j)\in\mathcal{K}} H_n(Y_i, Y_j),
$$

where $\mathcal{K}$ is an appropriate set of indices with $\mathcal{K} \subseteq \{(i,j) \mid |i-j| < l_n + \rho_n\}$. Furthermore, let

$$
\mathcal{K}_1 \;=\; \left\{(i_1, \ldots, i_4) \in \mathcal{K}\times\mathcal{K} \;\middle|\; \quad i_j \neq i_k \text{ for some } j,k; \quad \max_{k=1,\ldots,4}\min_{j\neq k}\{|i_k - i_j|\} \;<\; C_4\log(n)\right\},
$$

and

$$
\mathcal{K}_2 \;=\; \left\{(i_1, \ldots, i_4) \in \mathcal{K}\times\mathcal{K} \;\middle|\; \quad \max_{k=1,\ldots,4}\min_{j\neq k}\{|i_k - i_j|\} \;\geq\; C_4\log(n)\right\}.
$$

If $(i_1, \ldots, i_4) \in \mathcal{K}_1$, then

$$
EH_n(Y_{i_1}, Y_{i_2})H_n(Y_{i_3}, Y_{i_4}) \;=\; O(n^{-2}h). \qquad (4.26)
$$

Moreover, if $(i_1, \ldots, i_4) \in \mathcal{K}_2$, then

(4.27)
$$E H_n(Y_{i_1}, Y_{i_2}) H_n(Y_{i_3}, Y_{i_4}) \;=\; O\left(\sup_{x,y}\{|H_n(x,y)|\}\beta(C_4 \log(n))\right) \;=\; O(n^{-5}),$$

say, if C_4 is large enough. Hence, we obtain

$$\begin{aligned} & E|R_6 \,+\, R_7|^2 \\ & \quad = \sum_{(i,j)\in\mathcal{K}} E[H_n(Y_i, Y_j)]^2 \\ & \qquad + \sum_{(i_1,\ldots,i_4)\in\mathcal{K}_1} E H_n(Y_{i_1}, Y_{i_2}) H_n(Y_{i_3}, Y_{i_4}) \\ & \qquad + \sum_{(i_1,\ldots,i_4)\in\mathcal{K}_2} E H_n(Y_{i_1}, Y_{i_2}) H_n(Y_{i_3}, Y_{i_4}) \\ & \quad = O(n l_n\, n^{-2}) \,+\, O(\#\mathcal{K}_1\, n^{-2} h) \,+\, O(\#\mathcal{K}_2\, n^{-5}) \\ & \quad = O(l_n/n \,+\, (\log(n))^2 h), \end{aligned}$$

which implies

$$R_6 \,+\, R_7 \;=\; o_P(1). \tag{4.28}$$

From (4.21), (4.24), (4.25), and (4.28) we see that the difference between T_n and U_n is of negligible order, which completes the proof. □

Proof of Theorem 3.1. We write T_n^* in the form

$$T_n^* \;=\; \sum_{1\le i<j\le n} a_{ij,n} W_n(Y_i^*, Y_j^*) \;+\; \frac{1}{nh^{3d/2}} \sum_{i=1}^{n} \widetilde{H}_n(Y_i^*, Y_i^*) \tag{4.29}$$

where

$$a_{ij,n} = \frac{2}{n}\left\{h^{-3d} E^*[\widetilde{H}_n(Y_i^*, Y_j^*)]^2\right\}^{1/2}, \tag{4.30}$$

$$W_n(Y_i^*, Y_j^*) = \frac{\widetilde{H}_n(Y_i^*, Y_j^*)}{\left\{E^*[\widetilde{H}_n(Y_i^*, Y_j^*)]^2\right\}^{1/2}}, \tag{4.31}$$

(4.32)
$$\widetilde{H}_n(x,y) = \int \left[K\left(\frac{u-x}{h}\right) - E^* K\left(\frac{u-Y_1^*}{h}\right)\right]\left[K\left(\frac{u-y}{h}\right) - E^* K\left(\frac{u-Y_1^*}{h}\right)\right] du$$

and E^* denotes expectation with respect to the bootstrap distribution.

Since $Y_i^* \sim f_{\widehat{\theta}}$ we get using Assumption 6(i)

$$\begin{aligned}
&\frac{1}{nh^{3d/2}} \sum_{i=1}^n E^* \widetilde{H}_n(Y_i^*, Y_i^*) \quad - \quad h^{-d/2} \int K^2(u) du \\
&= \; h^{-3d/2} \int\int \left[K\left(\frac{u-z}{h}\right) - E^* K\left(\frac{u-Y_1^*}{h}\right) \right]^2 f_{\widehat{\theta}}(z)\, du\, dz \quad - \quad h^{-d/2} \int K^2(u)\, du \\
&= \; h^{-d/2} \int K^2(u)\, du \int f_{\widehat{\theta}}(z)\, dz \,+\, O_P(h^{d/2}) \quad - \quad h^{-d/2} \int K^2(u)\, du \\
&= \; O_P(h^{d/2}).
\end{aligned} \tag{4.33}$$

Furthermore, by the independence of the Y_i^*'s and because of $\sup_x\{\widetilde{H}(x,x)\} = O(h)$ we get by straightforward calculations

(4.34)

$$E^* \left[\frac{1}{nh^{3d/2}} \sum_{i=1}^n \left(\widetilde{H}_n(Y_i^*, Y_i^*) \,-\, E^* \widetilde{H}_n(Y_i^*, Y_i^*) \right) \right]^2 \;=\; O_P(n^{-1}h^{-d}).$$

From this and (4.33) we conclude that

$$\left| \frac{1}{nh^{3d/2}} \sum_{i=1}^n \widetilde{H}_n(Y_i^*, Y_i^*) - h^{-d/2} \int K^2(u)\, du \right| \to 0$$

in probability.
Consider next the quadratic form $\sum_{1 \le i < j \le n} a_{ij,n} W_n(Y_i^*, Y_j^*)$ and note that $E^* W_n(Y_i^*, y) = E^* W_n(y, Y_j^*) = 0$ and $E^* W_n^2(Y_i^*, Y_j^*) = 1$.
First, we have by the independence of the Y_i^*'s that

$$\begin{aligned}
&Var^*\Big[\sum_{1 \le i < j \le n} a_{ij,n} W_n(Y_i^*, Y_j^*) \Big] \\
&\asymp \; 2\, h^{-3d} \int \left\{ \int \left[K\left(\frac{u_1 - z}{h}\right) - E^* K\left(\frac{u_1 - Y_1^*}{h}\right) \right] \right. \\
&\qquad \times \left. \left[K\left(\frac{u_2 - z}{h}\right) - E^* K\left(\frac{u_2 - Y_1^*}{h}\right) \right] f_{\widehat{\theta}}(z)\, dz \right\}^2 du_1\, du_2
\end{aligned}$$

$$\to \; 2 \int f_{\theta_0}^2(x)\, dx \int [\int K(u)K(u+v)du]^2 dv \;=\; \sigma^2 \tag{4.35}$$

in probability.
To prove that $\sum_{1 \le i < j \le n} a_{ij,n} W_n(Y_i^*, Y_j^*) \to N(0, \sigma^2)$ it suffices therefore, by Theorem 5.3 of de Jong (1987) to show that the following three conditions are satisfied:

(4.36)

$$c(n)^2 \left\{ Var^*\Big[\sum_{1 \le i < j \le n} a_{ij,n} W_n(Y_i^*, Y_j^*) \Big] \right\}^{-1} \max_{1 \le i \le n} \sum_{j=1}^n a_{ij,n}^2 \;\to\; 0,$$

(4.37)
$$E^*\Big[W_n^2(Y_i^*,Y_j^*)I(|W_n(Y_i^*,Y_j^*)|>c(n))\Big]\;\to\;0$$

and

(4.38)
$$E^*\Big[W_n(Y_1^*,Y_2^*)W_n^*(Y_1^*,Y_3^*)W_n(Y_4^*,Y_2^*)W_n(Y_4^*,Y_3^*)\Big]\;\to\;0,$$

for a sequence of real numbers $c(n)$ such that $c(n)\to\infty$ as $n\to\infty$.

Now, let $c(n)=n^{1/2}h^{d/2}$ and verify that as in (4.35)

(4.39)
$$h^{-3d}E^*\Big[\widetilde{H}_n(Y_1^*,Y_2^*)\Big]^2\;\to\;\int f_{\theta_0}^2(x)\,dx\;\int[\int K(u)K(u+v)du]^2dv$$

in probability. Therefore,

$$\begin{aligned}
c(n)^2\max_{1\le i\le n}\sum_{j=1}^n a_{ij,n}^2 &= \frac{4c(n)^2}{nh^{3d}}E^*\Big[\widetilde{H}_n(Y_1^*,Y_2^*)\Big]^2\\
&= O_P(c(n)^2\,n^{-1})\\
&= O_P(h^d),
\end{aligned}$$

which together with (4.35) implies (4.36).

Consider next (4.37). By the above choice of $c(n)$ we have

$$\begin{aligned}
&E^*\Big[W_n^2(Y_i^*,Y_j^*)I(|W_n(Y_i^*,Y_j^*)|>c(n))\Big]\le\frac{1}{c^2(n)}E^*\Big[W_n^4(Y_1^*,Y_2^*)\Big]\\
&= O(n^{-1}\,h^{-7d})E^*\Big[\widetilde{H}_n(Y_1^*,Y_2^*)\Big]^4\Big\{h^{-3d}E^*[\widetilde{H}_n(Y_i^*,Y_j^*)]^2\Big\}^{-2}\\
&= O_P(n^{-1}\,h^{-2d})
\end{aligned}$$

where the last equation follows since

$$\begin{aligned}
&h^{-5d}E^*\Big[\widetilde{H}_n(Y_1^*,Y_2^*)\Big]^4\\
&= \int\Big\{\int\Big(K\Big(\frac{u_1-z}{h}\Big)E^*\,-\,K\Big(\frac{u_1-Y^*}{h}\Big)\Big)\\
&\quad\times\Big(K\Big(\frac{u_2-z}{h}\Big)E^*\,-\,K\Big(\frac{u_2-Y^*}{h}\Big)\Big)\Big(K\Big(\frac{u_3-z}{h}\Big)E^*\,-\,K\Big(\frac{u_3-Y^*}{h}\Big)\Big)\\
&\quad\times\Big(K\Big(\frac{u_4-z}{h}\Big)E^*\,-\,K\Big(\frac{u_4-Y^*}{h}\Big)\Big)\;f_{\widehat{\theta}}(z)\,dz\Big\}^2du_1\,du_2\,du_3\,du_4\\
&\to\int f_{\theta_0}^2(x)dx\int\Big[\int K(u)K(u+v_1)K(u+v_2)K(u+v_3)\,du\Big]^2dv_1\,dv_2\,dv_3
\end{aligned}$$

in probability.

To show (4.38) note first that

$$
\begin{aligned}
&h^{-6d}E^*\left[\widetilde{H}_n(Y_1^*,Y_2^*)\widetilde{H}_n(Y_1^*,Y_3^*)\widetilde{H}_n(Y_4^*,Y_2^*)\widetilde{H}_n(Y_4^*,Y_3^*)\right] \\
&= h^{-6d}\int \left(K\left(\frac{u_1-z_1}{h}\right)-E^*K\left(\frac{u_1-Y_1^*}{h}\right)\right) \\
&\quad\times\left(K\left(\frac{u_2-z_1}{h}\right)-E^*K\left(\frac{u_2-Y_1^*}{h}\right)\right)\left(K\left(\frac{u_1-z_2}{h}\right)-E^*K\left(\frac{u_1-Y_1^*}{h}\right)\right) \\
&\quad\times\left(K\left(\frac{u_3-z_2}{h}\right)-E^*K\left(\frac{u_2-Y_1^*}{h}\right)\right)\left(K\left(\frac{u_2-z_3}{h}\right)-E^*K\left(\frac{u_2-Y_1^*}{h}\right)\right) \\
&\quad\times\left(K\left(\frac{u_4-z_3}{h}\right)-E^*K\left(\frac{u_4-Y_1^*}{h}\right)\right)\left(K\left(\frac{u_4-z_4}{h}\right)-E^*K\left(\frac{u_4-Y_1^*}{h}\right)\right) \\
&\quad\times\left(K\left(\frac{u_3-z_4}{h}\right)-E^*K\left(\frac{u_3-Y_1^*}{h}\right)\right) f_{\widehat{\theta}}(z_1)\, f_{\widehat{\theta}}(z_2)\, f_{\widehat{\theta}}(z_3)\, f_{\widehat{\theta}}(z_4)\, dz_1\ldots dz_4\, du_1\ldots du_4 \\
&= h^{-2d}\int K(v_1)K\left(v_1+\frac{u_2-u_1}{h}\right)K(v_2)K\left(v_2+\frac{u_1-u_3}{h}\right)K(v_3)K\left(v_3+\frac{u_4-u_2}{h}\right) \\
&\quad\times K(v_4)K\left(v_4+\frac{u_3-u_4}{h}\right) f_{\widehat{\theta}}(u_1-v_1h)\ldots f_{\widehat{\theta}}(u_4-v_4h)\, dv_1\ldots dv_4\, du_1\ldots du_4 \\
&\quad+o_P(h^{-2d}) \\
&= O_P(h^d).
\end{aligned}
$$

From this and because of (4.39), we get using the definition of $W_n(Y_i^*,Y_j^*)$ that

$$
E^*\Big[W_n(Y_1^*,Y_2^*)W_n^*(Y_1^*,Y_3^*)W_n(Y_4^*,Y_2^*)W_n(Y_4^*,Y_3^*)\Big]=O_P(h^d).
$$

This completes the proof of the theorem. □

Acknowledgment. The work on this project has been initiated while the first author was visiting the University of Cyprus. He gratefully acknowledges financial support from this institution.

References

Bickel, P. and Rosenblatt, M. (1973). On some global measures of the derivation of density function estimators. *Ann. Statist.* **1**, 1071–1095

Brown, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59–66.

Doukhan, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**, Springer, New York.

Dvoretzky, A. (1972). Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* (ed. L. LeCam *et al.*) Los Angeles: University of California Press 2, 513–555.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.

Ghosh, B. K. and Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit. *Ann. Statist.* **19**, 999–1009.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–1947.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Rel. Fields* **75**, 261–277.

Kreiss, J.-P., Neumann, M. H. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. Preprint No. 98/07, TU Braunschweig.

Neumann, M. H. (1996). Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.*, to appear.

Neumann, M. H. (1997). On robustness of model-based bootstrap schemes in nonparametric time series analysis. Discussion Paper 88/97, SFB 373, Humboldt University, Berlin.

Neumann, M. H. and Kreiss, J.-P. (1997). Regression-type inference for the autoregression function. *Ann. Statist.*, to appear.

Paparoditis, E. (1997). Spectral density based goodness of fit tests for time series models. Technical Report 97/14, University of Cyprus.

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

Rosenblatt, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3**, 1–14.

Takahata, H. and Yoshihara, K. (1987). Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences. *Yokohama Math. J.* **35**, 95–111.

Yoshihara, K. (1994). *Weakly Dependent Stochastic Sequences and Their Applications. Vol. IV. Curve Estimation Based on Weakly Dependent Data.* Sanseido, Tokyo.