Density testing in a contaminated sample

Hajo Holzmann,[1] Nicolai Bissantz and Axel Munk
Institute for Mathematical Stochastics
Georg-August-University Göttingen
37073 Göttingen, Germany

Abstract

We study non-parametric tests for checking parametric hypotheses about a multivariate density f of independent identically distributed random vectors $Z_1, Z_2, \ldots$ which are observed under additional noise with density ψ. The tests we propose are an extension of the test due to Bickel and Rosenblatt [*Ann. Statist.* 1 (1973), 1071–1095] and are based on a comparison of a nonparametric deconvolution estimator and the smoothed version of a parametric fit of the density f of the variables of interest Z_i. In an example the loss of efficiency is highlighted when the test is based on the convolved (but observable) density $g = f * \psi$ instead on the initial density of interest f.

AMS subject classification: Primary 62G07; Secondary 62G10, 62G20.

Keywords: asymptotic normality; deconvolution; goodness of fit; integrated square error; multivariate nonparametric density estimation

1 Introduction

The goodness of fit problem for testing whether i.i.d. random vectors $Z_1, \ldots, Z_n$ with values in $\mathbb{R}^p$, $p \geq 1$, are distributed according to a density f (or some parametric family) has been well-studied in the literature, and a variety of methods have been suggested. In particular, Bickel and Rosenblatt [1] proposed a test based on the L_2-distance between a non-parametric kernel density estimator and a smoothed version of a parametric fit. Their method was extended by Fan and Ullah [11] and Neumann and Paparoditis [19] to testing parametric hypotheses about the marginal distribution of a stationary process. Similar in spirit is the work of Härdle and Mammen [17] and Paparoditis [21] for the regression function and for the spectral density, respectively.
In this paper we consider the case when the Z_i can only be observed with an additional noise term, i.e. instead of Z_i one observes X_i, where

$$X_i = Z_i + \epsilon_i,$$

[1]corresponding author, Institut für Mathematische Stochastik, Universität Göttingen, Maschmühlenweg 8–10, 37073 Göttingen, Germany, Fon: +49/551/3913514, Fax: +49/551/39-13505
E-mail address: holzmann@math.uni-goettingen.de (H. Holzmann), bissantz@math.uni-goettingen.de (N. Bissantz), munk@math.uni-goettingen.de (A. Munk),

and the ϵ_i are i.i.d. with known density ψ and independent of the Z_i. Hence the X_i have density

$$g = f * \psi, \tag{1}$$

where $f * \psi$ denotes the convolution of f and ψ. Recovering f from observations $X_1, \ldots, X_n$ is called the deconvolution problem and has been treated extensively in the statistics literature (see e.g. [8, 9, 10, 25]). It is well known that the minimax rate of convergence of estimators of f depends sensitively on the tail behaviour of the characteristic function Φ_ψ of the errors ϵ_i (cf. [9]). Suppose that $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}^p$. If $|\Phi_\psi(t)|$ is of polynomial order $|t|^{-\beta}$ for some $\beta > 0$ as $|t| \to \infty$, we speak of the ordinary smooth case, and if $|\Phi_\psi(t)|$ is of exponential order $|t|^{\beta_0} e^{-|t|^\beta/\gamma}$, $\beta, \gamma > 0$, of the super smooth case. Here both the Euclidean norm on $\mathbb{R}^p$ and the absolute value on $\mathbb{R}$ are denoted by $|\cdot|$. Important examples for the ordinary smooth case are exponential, Laplace and Gamma deconvolution, and for the super smooth case normal and Cauchy deconvolution. A class of examples for ordinary smooth multivariate distributions is given in [14]. There are also cases of interest in which $\Phi_\psi(t) = 0$ for some t, e.g. if the ϵ_i are uniformly distributed (cf. [15]), and if the density ψ is a band-limited function (cf. [18]). For dealing with the deconvolution problem in the context of general statistical inverse problems see van Rooij, Ruymgaart and van Zwet [28] .

In this paper we are concerned with the problem of testing the goodness of fit of f to a parametric model $\mathcal{M} = \{f(\cdot, \theta)\}_{\theta \in \Theta}$, where $\Theta \subseteq \mathbb{R}^k$. There are two possible approaches to this problem. For the first, observe that parametric assumptions on the original density f can be expressed uniquely in terms of parametric assumptions on g. This is due to the fact that convolution with ψ is injective since by assumption, $\Phi_\psi(t) \neq 0$ for all t. Since we observe data distributed according to g, in principle all direct testing procedures (e.g. the test suggested in [1], or classical tests based on the cumulative distribution function G of g such as the Kolmogorov-Smirnov or the Cramer-von-Mises tests) could be applied to test such equivalent parametric assumptions on g as well. However, it turns out that this procedure is not appropriate, in general, for certain alternatives given in terms of f. The reason is that the deconvolution problem is ill-posed, i.e. the inverse of the convolution operator is unbounded. Thus it can happen that the true f_0 is at an arbitrarily large L_2-distance to the parametric model in the domain of f, whereas, the corresponding $g_0 = f_0 * \psi$ is very close to the parametric model in the domain of g. Hence direct application of tests to the (observable) data $X_1, \ldots, X_n$ will result in an inefficient procedure for those alternatives which can hardly be distinguished from the null in terms of g. Therefore, we suggest to take a different approach by constructing tests which are based on an inverse estimator $\hat{f}$ of f and hence deal directly with the original density f. To our knowledge this has never been treated so far in the literature.

More specifically, in this paper we develop a version of the Bickel-Rosenblatt (BR) test, based on a kernel deconvolution estimator of f (cf. [9]), for testing parametric assumptions on the density f in an ordinary smooth deconvolution problem. In Section 2 we introduce the test statistic and determine its asymptotic behaviour for a simple hypothesis $f_0 = f$. In Section 3 we discuss advantages of this test as compared to the direct testing procedures as mentioned above. To this end we study the behaviour of the test statistics under local alternatives as well as under fixed alternatives. In a particular example it is shown that non-linear local alternatives, which converge to f_0 at a slower rate than $1/\sqrt{n}$, cannot be detected by the direct BR test, since in the domain of g they converge arbitrarily fast to g_0. Furthermore, it turns out that a $\sqrt{n}$ rate (cf. [5]) under fixed alternatives is only valid under additional smoothness assumptions on f. In this case we are able to prove a limit theorem with a $\sqrt{n}$ rate. Our approach allows e.g. to construct confidence intervals for $\|f - f_0\|$, which is not possible by any direct method dealing with the density g. In addition, our findings concerning the superior power of our indirect testing procedure are supported by a small simulation experiment. Finally, in Section 4 we consider the case of testing a composite hypothesis, and discuss the related problem of testing assumptions on the derivatives of a density.

Technically, our work is related to [22] where asymptotic normality for the weighted integrated squared error of a kernel deconvolution density estimator is proved. However, this result does not apply to the unweighted L_2-distance as required for our purposes. Furthermore, their result only applies in the one-dimensional case. Consequently, our proofs are completely different and rely on Fourier methods together with a result of Hall [16].

In order to keep the paper more readable all proofs are deferred to an appendix.

2 The Bickel-Rosenblatt test in deconvolution models

In this section we describe the asymptotic behaviour of the BR-test in an ordinary smooth deconvolution problem for a simple hypothesis. To fix the notation, the Fourier transform of f is given by

$$\mathcal{F}(f)(t) = \Phi_f(t) = \int_{\mathbb{R}^p} f(x) e^{it\cdot x}\, dx,$$

where $t \cdot x$ denotes the inner product on $\mathbb{R}^p$. Under the assumptions that $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}^p$ and that the Fourier transform Φ_K of the kernel K has compact support, the kernel deconvolution density estimator

$$\hat{f}_n(x) = \frac{1}{(2\pi)^p} \int_{\mathbb{R}^p} e^{-it\cdot x} \Phi_K(ht) \frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)}\, dt \tag{2}$$

is well-defined. Here $h > 0$ is a smoothing parameter called bandwidth and $\hat{\Phi}_n(t) = 1/n \sum e^{it\cdot X_k}$ is the empirical characteristic function of $X_1, \ldots, X_n$. Local properties of this estimator were studied by Fan [9, 10] and van Es and Uh [27], mean integrated squared error (MISE) properties by Diggle and Hall [8] and Zhang [29], and rates of convergence of quadratic functionals of $\hat{f}_n$ were investigated by Delaigle and Gijbels [2]. Practical suggestions of how to choose the bandwidth in order to minimize the mean integrated squared error are given in Delaigle and Gijbels [3]. Here we use the kernel deconvolution density estimator in (2) to construct a deconvolution version of the Bickel-Rosenblatt test.

Suppose first that we want to test the simple hypothesis $f = f_0$. Then the BR test statistic is

$$T_n = \int_{\mathbb{R}^p} \left(\hat{f}_n - K_h * f_0\right)^2 (x)\, dx, \tag{3}$$

where $K_h(x) = K(x/h)/h^p$. We study this statistic for the ordinary smooth case. Thus we impose the following assumptions:

Assumption A: The characteristic function Φ_ψ of the error variable ϵ satisfies $\Phi_\psi(t) \neq 0$ $\forall\, t \in \mathbb{R}^p$, and $|\Phi_\psi(t)| \approx |t|^{-\beta}$ for a $\beta \geq 0$ (i.e. there are $c, C > 0$ such that $c|t|^{-\beta} \leq |\Phi_\psi(t)| \leq$

$C|t|^{-\beta}$ for sufficiently large $|t|$).

Examples for univariate and multivariate densities satisfying assumption A are given in Example 1. We will also need the following smoothness assumption on the density f.

Assumption B: The Fourier transform Φ_f of the density f satisfies $|\Phi_f(t)| = O(|t|^{-r})$ for some $r > p$ as $|t| \to \infty$.

Finally, we need the following regularity assumption for the Fourier transform of the kernel K.

Assumption C: The Fourier transform Φ_K of K is symmetric ($\Phi_K(t) = \Phi_K(-t)$) and compactly supported, where we assume w.l.o.g. that the support is contained in $[-1, 1]^p$, and $|\Phi_K(t)| \leq 1$.

Our first result concerns the asymptotic distribution of T_n under the null hypothesis $H_0: \ f = f_0$.

Theorem 1. *Suppose that assumptions A, B and C are satisfied, and that the hypothesis $H_0: \ f = f_0$ holds. If $h \to 0$ and $nh^p \to \infty$, then*

$$n/C_{V,h}^{1/2}\Big(T_n - C_{M,h}/\big((2\pi)^p n\big)\Big) \xrightarrow{\mathcal{L}} N\big(0, 2/(2\pi)^{2p}\big), \tag{4}$$

where

$$C_{M,h} = \int_{\mathbb{R}^p} |\Phi_K(ht)|^2/|\Phi_\psi(t)|^2 \, dt, \quad C_{M,h} \approx h^{-(2\beta+p)},$$

and

$$C_{V,h} = \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \frac{\big|\Phi_K(th)\big|^2\big|\Phi_K(sh)\big|^2}{|\Phi_\psi(t)|^2|\Phi_\psi(s)|^2} \, |\Phi_g(t+s)|^2 \, ds \, dt, \quad C_{V,h} \approx h^{-(4\beta+p)}.$$

A related result on the asymptotic distribution of the integrated squared error for a deconvolution density estimator is given by Piterbarg and Penskaya [22]. However, their result only applies in the one-dimensional case and if in addition an integrable weight function is used in (3). Further, their proof follows the rather sophisticated strong approximation arguments in [1]. In contrast, our proof is based on a simple limit theorem for U-statistics with variable kernels due to Hall [16]. However, since our arguments rely heavily on Fourier transformation methods, they only apply (at least not without strong modifications) to the case where no weight function is used in (3). Therefore our results are in the one-dimensional case complementary to those in [22].

Example 1. A class of characteristic functions satisfying assumption A in any dimension is given by

$$\Phi_\psi(t) = 1/(1 + c|t|^\alpha)^\gamma, \quad t \in \mathbb{R}^p, \tag{5}$$

where $c > 0, \ 0 < \alpha \leq 2, \ \gamma \geq 0$ (cf. [14], where such functions occur as correlation functions of stationary Gaussian processes). Notice that this class consists of radial functions. For the density corresponding to (5), in general there is no closed form expression available. In the univariate case, for $\alpha = 2$, $2\gamma = c + 1$, it is McKay's Bessel function density (cf. [13]),

$$\psi(x) = \frac{\nu^{1/2}}{\pi^{1/2}\Gamma((\nu+1)/2)} \Big(\frac{\nu^{1/2}|x|}{2}\Big)^{\nu/2} \, K_{\nu/2}(\nu^{1/2}|x|),$$

where $\nu = 1/c$ and K is the modified Bessel function.
Now suppose that there exist $C_\psi, \delta > 0$ such that

$$\left|C_\psi |t|^{2\beta} |\Phi_\psi(t)|^2 - 1\right| = O\big(|t|^{-(p/2+\delta)}\big). \tag{6}$$

Then under the assumptions of Theorem 1 we get explicit formulas for the asymptotic expectation and variance, namely

$$nh^{2\beta+p/2}\Big(T_n - \frac{C_\psi C_{M,K}}{(2\pi)^p\, nh^{2\beta+p}}\Big) \xrightarrow{\mathcal{L}} N\Big(0, \frac{C_\psi^2 C_{V,K} \|g\|^2}{2^{p-1}\pi}\Big), \tag{7}$$

where $\|g\|$ denotes the L_2 norm of g, and

$$C_{M,K} = \int_{\mathbb{R}^p} |t|^{2\beta} |\Phi_K(t)|^2\, dt, \quad C_{V,K} = \int_{\mathbb{R}^p} |t|^{4\beta} |\Phi_K(t)|^4\, dt. \tag{8}$$

Condition (6), for example, is satisfied by the univariate exponential distribution and the Laplace distribution (for both $C_\psi = 1$). More generally, the characteristic function of form (5) satisfies (6) if $\gamma \geq 1$ and $2\alpha > p$. In this case $\beta = \alpha\gamma$ and $C_\phi = c^{2\gamma}$.
The asymptotic result (7) now allows to construct an asymptotic level-α test for the simple hypothesis $H_0 : f = f_0$ as follows. Reject H_0 if

$$T_n > \frac{C_\psi C_{V,K}^{1/2}\, \|g_0\|\, q_{1-\alpha}}{nh^{2\beta+p/2}\,(2^{p-1}\pi)^{1/2}} + \frac{C_\psi C_{M,K}}{(2\pi)^p\, nh^{2\beta+p}},$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the normal distribution and $g_0 = f_0 * \psi$. Note that the constants C_ψ, $C_{M,K}$ and $C_{V,K}$ are explicitly available.

3 Comparison with direct testing

In this section we compare the indirect testing procedure based on the density f of the variables of interest with the BR test based on the density g of the observations.

3.1 Alternative testing procedures. Observe that under the assumption $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}^p$, the convolution operator $Conv_\psi$ given in (1) is injective. Therefore any hypothesis

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

formulated in terms of the density f of the Z_i can be equivalently expressed as

$$H_0' : \quad g = g_0, \quad g_0 = f_0 * \psi, \quad g \in \mathrm{Im}(Conv_\psi).$$

Since under the hypothesis H_0' we have observations $X_1, \ldots, X_n$ distributed according to g_0, in principle the hypothesis H_0' could be tested as well by any direct testing procedure. In order to illustrate the lack of efficiency of this procedure we will restrict ourselves to the classical Bickel-Rosenblatt test statistic without weight function

$$T_n^{BR} = \int_{\mathbb{R}^p} \big(\hat{g}_n - L_h * g\big)^2(x)\, dx,$$

where $\hat{g}_n(x) = 1/n \sum_{k=1}^n L_h(x - X_k)$ is a multivariate kernel density estimator with compactly supported kernel L which satisfies some additional regularity conditions (cf. [16] for details). Under the null hypothesis H_0', one has the limit behaviour

$$nh^{p/2}\Big(T_n^{BR} - \frac{1}{nh^p}\int_{\mathbb{R}^p} L^2(x)\,dx\Big) \xrightarrow{\mathcal{L}} N(0, S_{BR}^2),$$

where the variance S_{BR}^2 is given by

$$S_{BR}^2 = 2\,\|g_0\|^2 \int_{\mathbb{R}^p} (L * L)^2(x)\,dx.$$

3.2 Local alternatives. For simplicity, and since the phenomenon depends on the ill-posedness of the deconvolution problem rather than on multivariate densities, in this section we will restrict the presentation to the univariate case. Usually one studies local alternatives of the following linear form (cf. [1], or [4], for the context of regression)

$$f_n(t) = f_0(t) + \delta_n l(t), \tag{9}$$

where $\delta_n \to 0$ at a certain rate. This implies that $\|f_n - f_0\|$ is of order δ_n, but due to the assumed linearity we will see that these types of alternatives are not suitable to highlight the additional difficulty encountered with deconvolution problems. Nevertheless, for the moment assume that f_0 satisfies assumption B and that l is bounded and square integrable, and let $\delta_n = (2C_{V,h}\pi^2)^{1/4}/n^{1/2}$, which is of order $1/(h^{\beta+1/4}n^{1/2})$. Define T_n and $C_{V,h}$ in terms of f_0. An inspection of the proof of Theorem 1 shows that if $h \to 0$ and $nh^{4\beta+1/2} \to \infty$,

$$n/(2C_{V,h}\pi^2)^{1/2}\big(T_n - C_{M,h}/(2\pi n)\big) \xrightarrow{\mathcal{L}_{f_n}} N(\|l\|^2, 1), \tag{10}$$

where f_n is given in (9), and $\|l\|^2$ will be called the shift parameter of the test statistic T_n. Thus, roughly speaking, the BR test based on f can detect alternatives which converge to f_0 at any rate slower than $n^{-1/2}$. On the other hand, for the classical BR test with local alternatives of the linear form

$$g_n(x) = g_0(x) + \delta_n l(x), \tag{11}$$

where $\delta_n = 1/\sqrt{nh^{1/2}}$ and $h = O(n^{-\gamma})$ for $0 < \gamma < 1/4$, one can show that

$$nh^{1/2}\big(T_n^{BR} - \frac{1}{nh}\int_{\mathbb{R}} L^2(x)\,dx\big) \xrightarrow{\mathcal{L}_{g_n}} N(\|l\|^2, S_{BR}^2). \tag{12}$$

Thus the test based on T_n^{BR} can also detect *linear* alternatives of the type (11) which converge to g_0 at any rate slower than $n^{-1/2}$.

In the literature it has been observed that while tests based on the cumulative distribution function are better at detecting linear local alternatives, smoothing based tests such as the BR test outperform these tests at detecting non-linear alternatives, cf. Rosenblatt [24] or Ghosh and Huang [12]. Therefore, we will compare our test with the BR-test on the basis of certain non-linear alternatives. First notice that the linear alternatives (9) are mapped under the convolution operator to alternatives of the form (11), and the rate is preserved (albeit the norm $\|l\|$ may change drastically). However, this does not hold if one considers alternatives f_n which converge to some f_0 in a non-linear fashion, as shown in the subsequent Example 2. The reason is that the inverse of the operator $Conv_\psi$ (defined on a subspace of L_2) is unbounded in general, but bounded on finite-dimensional linear subspaces.

Example 2. Let $f_0(x) = 1/\pi(1+x^2)$ be the Cauchy density and consider the sequence

$$f_n(x) = f_0(x) + \delta_n l_n(x),$$

where

$$l_n(x) = 2\cos(a_n x)\Big(\frac{\sin x}{x}\Big)^k,$$

$k \geq 2$ and $a_n \to \infty$. Let us first show that for small δ_n and large a_n, f_n is indeed a density. Firstly, if δ_n is small enough, $f_n > 0$. Let Φ denote the Fourier transform of $(\sin x/x)^k$, and notice that Φ has compact support. We have

$$\Phi_{l_n}(t) = \Phi(t - a_n) + \Phi(t + a_n),$$

therefore for large a_n, $\int l_n = \Phi_{l_n}(0) = 0$. Hence for suitable δ_n and a_n, f_n is indeed a density. Moreover, for large a_n, $\|\Phi_{l_n}\|^2 = 2\|\Phi\|^2$. Thus

$$\|f_n - f_0\|^2 = \delta_n^2 \|l_n\|^2 = \delta_n^2 \|\Phi\|^2/\pi,$$

and f_n converges with rate δ_n to f_0 in L_2. Now let ψ be the Laplace density which has Fourier transform $1/(1+x^2)$, and let $g_n = f_n * \psi$. Then

$$\|g_n - g_0\|^2 = \delta_n^2 \|l_n * \psi\|^2 = O(\delta_n^2/a_n^4).$$

Thus, with a proper choice of $a_n \to \infty$, g_n converges at an arbitrarily fast rate to g_0 in L_2, and the shift parameter in (12) will converge to zero. In contrast, one can show that a statement analogous to (10), even for $a_n \to \infty$, with non-zero shift parameter is still valid.

In summary, certain non-linear local alternatives can be detected by the indirect test T_n but not by the direct BR-test T_n^{BR} and also not by the classical $\sqrt{n}$-consistent tests such as the Kolmogorov-Smirnov test, and in this respect our test appears to be more efficient. Let us point out that there is a wide variety of efficiency concepts for goodness-of-fit tests in the literature, see e.g. [20], which we will not pursue any further. However, each efficiency concept appears to bear its own difficulty and challenge in interpretation, therefore, we will support our theoretical findings by a simulation experiment. This is conducted in Section 3.4 and shows that at least in the scenario considered there, the indirect test T_n outperforms the BR-test as well as classical goodness-of-fit tests.

3.3 Fixed alternatives. Finally we consider the asymptotic behaviour of the test statistics under fixed alternatives. First consider the classical BR test. Suppose that for the true density g we have that $\|g - g_0\| > 0$. Under the assumptions of Theorem 3 (cf. Section 4.2), we have that

$$n^{1/2}\Big(T_n^{BR} - \|g - g_0\|^2\Big) \xrightarrow{\mathcal{L}} N(0, \tau_{BR}^2),$$

where

$$\tau_{BR}^2 = \mathrm{Var}\big((g - g_0)(X_1)\big),$$

see also Dette and Bachmann [5] for a related result in the univariate case. This result can be used to test hypotheses of the form

$$H_0' : \|g - g_0\| > \pi \quad \text{against} \quad H_1' : \|g - g_0\| \leq \pi, \tag{13}$$

for some $\pi > 0$. For further details on such testing problems in a regression context cf. Dette and Munk [6]. Observe that the variance of T_n^{BR} under fixed alternatives is of order n^{-1}. Surprisingly, in the indirect case this is no longer true, the variance of T_n under a fixed alternative may be larger than n^{-1} in general. For the following considerations it will be convenient to use the following stronger assumption on the kernel K, which allows a simple estimation of relevant bias terms.

Assumption C*: (*Flat top kernel*) The Fourier transform Φ_K of K is symmetric with support in $[-1,1]^p$, $|\Phi_K(t)| \leq 1$, and $\Phi_K(t) = 1$ for $t \in [-\delta, \delta]^p$ for some $\delta > 0$.

Flat top kernels were previously used in multivariate direct density estimation by Politis and Romano [23]. For a fixed density f let r_0 denote the maximal r such that assumption B is satisfied for both Φ_f and Φ_{f_0} ($r_0 = \infty$ if assumption B is satisfied for all r). The next proposition gives upper bounds for the variance of T_n under fixed alternatives. It turns out, that these bounds depend on the index r_0, i.e. on the tail behaviour of Φ_f and Φ_{f_0}.

Proposition 1. *Suppose that assumptions A, B and C* hold and let r_0 be as above. Under the alternative $f \neq f_0$, if $h \to 0$ and $nh^{2\beta+p} \to \infty$,*

$$ET_n = \|f - f_0\|^2 + O\big(h^{2r_0-p}\big) + O\big(1/(nh^{2\beta+p})\big) \tag{14}$$

and if $nh^{4\min(r_0,\beta)} \to 0$,

$$Var\ T_n = \begin{cases} O\big(1/(nh^{2(\beta-r_0)+p}\big) & : \ \beta \geq r_0 \\ O\big(1/(nh^{p-2(r_0-\beta)}\big) & : \ p/2 > r_0 - \beta > 0 \\ O\big(\log(1/h)/n\big) & : \ p/2 = r_0 - \beta \\ O\big(1/n\big) & : \ p/2 < r_0 - \beta \end{cases}. \tag{15}$$

In general, it appears to be difficult to determine the exact order of the variance of T_n under fixed alternatives, and hence to derive a corresponding limit law. However, if the densities are assumed to be sufficiently smooth, this is still possible, as shown in the following theorem.

Theorem 2. *Suppose that $f \neq f_0$, that assumption C* holds and that both f and f_0 satisfy assumption B for some $r > \beta + p$. If $h \to 0$ is such that $nh^{4\beta+2p} \to \infty$ and $nh^{4r-2p} \to 0$, then*

$$n^{1/2}\big(T_n - \|f - f_0\|^2\big) \xrightarrow{\mathcal{L}} N(0, \tau^2), \tag{16}$$

where

$$\tau^2 = Var\Big(\mathcal{F}^{-1}\big(\frac{\Phi_f - \Phi_{f_0}}{\Phi_\psi}\big)(X_1)\Big). \tag{17}$$

Under the additional smoothness assumption in Theorem 2 we are in the position to test the hypothesis

$$H_0 : \|f - f_0\| > \pi \quad \text{against} \quad H_1 : \|f - f_0\| \leq \pi, \tag{18}$$

formulated in terms of f. In addition, confidence intervals for $\|f - f_0\|^2$ can be derived from (16) as well. Notice that (18) and (13) are not equivalent, since the inverse of the convolution operator $Conv_\psi$ is unbounded, thus (18) could not be tested by a direct testing procedure. Hence, hypotheses of the type (18) and (13) nicely express the additional difficulty due to the ill-posed character of the problem, which is not captured by the classical hypothesis $H_0 : f = f_0$. However, it might not be easy to apply Theorem 2 in practice, since the rather

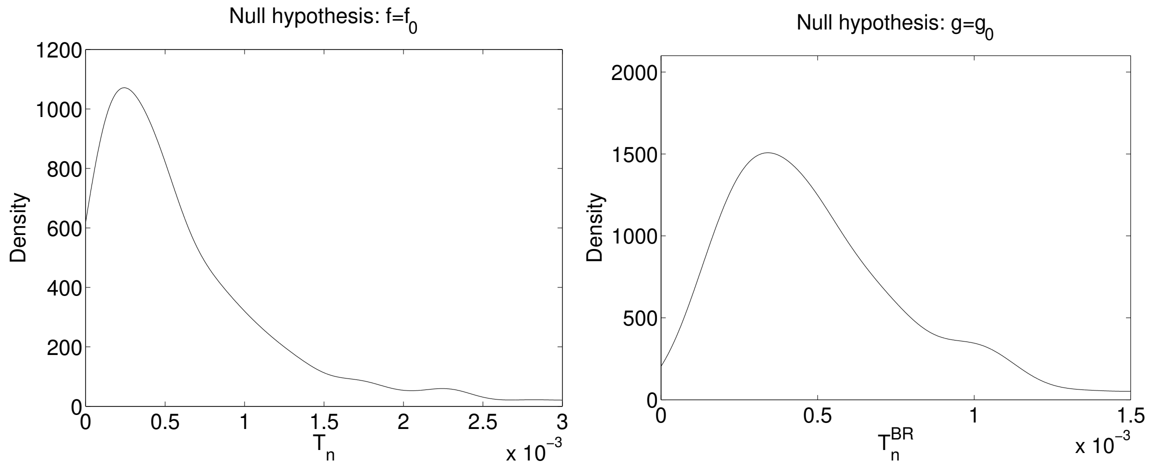

Figure 1: Empirical densities of T_n and T_n^{BR} from 500 simulations under the null hypothesis $f = f_0$ (indirect test, left panel), respectively $g = g_0$, where $g_0 := f_0 * \psi$ (BR test, right panel).

complicated expression (17) for the variance has to be estimated, and bootstrapping T_n may become necessary.

3.4 Simulation study

In this section we present the results of a small simulation study of our testing procedure. To this end we generate data from a centered normal density f_0 with variance $\sigma^2 = 0.3$, contaminated with an additional noise term ε with characteristic function

$$\Phi_\psi(t) = \frac{1}{(1+t^2)^2}.$$

A density ψ with this characteristic function is obtained by convolution of a Laplace density with parameter 1 with itself, and satisfies assumption A for $\beta = 4$. The sample size in the subsequent simulations is $n = 500$, the sinc kernel $K(x) = \sin x/(\pi x)$ is used, and the bandwidth is chosen to maximize the power of the test for the null hypothesis $H_0 : f = f_0$.
Fig. 1 shows the empirical density of the test statistic T_n under the null hypothesis. Note, that the actual distribution is left skewed and still for 500 samples poorly approximated by the normal limit law in Theorem 1, whereas the moments match rather well. This phenomenon is well known for quadratic statistics such as T_n and occurs in various situations (cf. [17], [4]). Improvement can be achieved by matching a scaled, noncentral χ^2-distribution with estimated number of freedoms (cf. [7]) or by bootstrap variants of the test (cf. [17]). The estimated mean and variance of T_n are $3.4 \cdot 10^{-3}$ and $8.4 \cdot 10^{-6}$, respectively, whereas the asymptotic values of eq. (4) in Theorem 1 are $3.3 \cdot 10^{-3}$ and $8.3 \cdot 10^{-6}$, respectively.
In a second part we simulate the power of the test for the null hypothesis $H_0 : f = f_0$ under the same assumptions as used so far. We use the quantiles from the empirical distributions of T_n to determine the critical values for T_n, and as alternative we consider the mixture density

$$f_1 = 0.7 \cdot f_0 + 0.3 \cdot \psi.$$

In Table 3 we compare the power of our indirect testing procedure with the power of the BR test based on the density g of the observations, where the sample size is again $n = 500$, and

Level	BR test	Indirect test	KS test
0.05	0.49	0.52	0.31
0.10	0.56	0.72	0.50
0.20	0.73	0.86	0.71

Table 1: Simulated power from 500 replications of the BR test and the indirect testing procedure. For both tests the bandwidth is selected such that the power is maximized. Its values are $h = 0.89$ and $h = 0.95$ for the BR test and the indirect test, respectively. Moreover, we simulate the Kolmogorov-Smirnov (KS) test for the same hypothesis $g = g_0$, with $g_1 := f_1 * \psi$.

the bandwidth has been selected such that the power of the (direct) BR test is maximized. Note that, because we use critical values from the empirical distributions of T_n and T_n^{BR}, respectively, both our indirect and the BR test maintain the desired level.
The results from the power simulations show that the indirect testing procedure outperforms the direct BR test as well as the Kolmogorov-Smirnov test by some amount. For other settings we found a similar behaviour, the inverse BR-test outperforms the BR-test as well as the KS test in most cases slightly. The magnitude of the difference in power depends on the behaviour of Φ_ψ and Φ_f near the frequency cutoff given by the support of $\Phi_K(h\cdot)$. Usage of the inverse test implies that we first aim to recover the characteristic function of f. In general, in the spectral window defined by the support of $\Phi_K(h\cdot)$, Φ_ψ attains its smallest absolute values near the boundaries. Therefore, the inverse test will outperform the BR-test most if Φ_{f_0} and Φ_{f_1} mainly differ in high frequencies which are still in the support of $\Phi_K(h\cdot)$.

4 Further extensions and modifications

4.1 Composite hypotheses. More important from a practical point of view than testing a simple hypothesis $H_0 : f = f_0$ is to test whether f belongs to some finite-dimensional parametric family, i.e.

$$H_0 \;:\; f \in \mathcal{M}.$$

Here, we compare the nonparametric fit to a smoothed version of the parametric fit, i.e. we consider

$$T_{n,\hat{\theta}} = \int_{\mathbb{R}^p} \big(\hat{f}_n(x) - K_h * f_{\hat{\theta}}\big)^2(x)\, dx,$$

where $\hat{\theta}$ is a consistent parametric estimate to be specified later. As in Neumann and Paparoditis [19], we derive the limit distribution of $T_{n,\hat{\theta}}$ by estimating the difference between T_n and $T_{n,\hat{\theta}}$. Denote the Fourier transform of f_θ by Φ_θ and the true parameter by θ_0. We have, using Parseval's formula,

$$\begin{aligned} 2\pi\big(T_{n,\hat{\theta}} - T_n\big) &= 2\int_{\mathbb{R}^p} \Phi_K(ht)\Big(\frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)} - \Phi_{\theta_0}(t)\Big)\overline{\Phi_K(ht)\big(\Phi_{\theta_0}(t) - \Phi_{\hat{\theta}}(t)\big)}\, dt \\ &\quad + \int_{\mathbb{R}^p} \big|\Phi_K(ht)\big|^2\big|\Phi_{\theta_0}(t) - \Phi_{\hat{\theta}}(t)\big|^2\, dt. \end{aligned} \tag{19}$$

Write

$$\Phi_{\hat{\theta}}(t) - \Phi_{\theta_0}(t) = (\hat{\theta} - \theta_0)\Phi'_{\theta_0}(t) + R(\hat{\theta}, \theta_0, t),$$

where $\Phi'_{\theta_0} = \nabla_{\theta_0}\Phi_{\theta_0}$ is the gradient of Φ_θ w.r.t. θ at θ_0.

We will need the following assumptions.

Assumption D

1. $(\hat\theta - \theta_0) = o_P\big(n^{-1/2}h^{-\beta}\big)$,
2. $\sup_{t\in\mathbb{R}} |\Phi'_\theta(t)| < \infty$, $\theta \in \Theta$,
3. $\int_{\mathbb{R}^p} R^2(\hat\theta, \theta_0, t)\, dt = o_P(n^{-1})$,
4. $\int_{\mathbb{R}^p} \big|\Phi_{\theta_0}(t) - \Phi_{\hat\theta}(t)\big|^2\, dt = o_P\big(n^{-1}h^{-2(\beta+p/2)}\big)$.

Under assumptions A–D, one shows similarly as in Neumann and Pararoditis [19], using (19) and an estimate for the order of the variance $Var\, T_n$ similar to (15) that

$$T_{n,\hat\theta} - T_n = o_P\big(n^{-1}h^{-2\beta-p/2}\big).$$

Therefore the assertion of Theorem 1 remains unchanged for $T_{n,\hat\theta}$, provided assumption D holds.

4.2 Testing the derivative of a density. It has been observed in the literature (cf. van Es and Kok, [26]) that estimates of the derivative of a density behave similarly as estimates in an ordinary smooth deconvolution problem. In this section let $p = 1$ and consider the kernel estimator of the k-th derivative ($k \geq 0$) of g,

$$\begin{aligned} \hat g_n^{(k)}(x) &= \frac{1}{nh^{k+1}} \sum_{j=1}^n K^{(k)}\big((x - X_j)/h\big) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx}\, \Phi_K(th)\, (-it)^k\, \hat\Phi_n(t)\, dt, \end{aligned} \tag{20}$$

where we assume that $X_1, \ldots, X_n$ are i.i.d. with density g and that K satisfies assumption C. Recall that the Fourier transform Φ_K of K has compact support, therefore $K \in C^\infty$. Notice that taking the k-th derivative is an injective operation on the k-times differentiable densities on the real line. The BR test statistic is in this case given by

$$T_n = \int_{\mathbb{R}} \big(\hat g_n^{(k)}(x) - g_{n,0}^{(k)}(x)\big)^2\, dx,$$

where $g_{n,0}^{(k)}$ is as in (20) but replacing the empirical characteristic function by the characteristic function Φ_{g_0}. The proof of the following theorem is similar (in fact simpler) to the proofs of Theorems 1 and 2 and therefore omitted.

Theorem 3. *Under the hypothesis $g = g_0$, if g satisfies assumption B with some $r > k + 1$ and $h \to 0$, $nh \to \infty$, then*

$$nh^{2k+1/2}\big(T_n - \frac{C_{M,K}}{2\pi nh^{2k+1}}\big) \xrightarrow{\mathcal{L}} N\big(0, \frac{C_{V,K}\|g\|^2}{\pi}\big),$$

where $C_{M,K}$ and $C_{V,K}$ are given in (8) for $\beta = k$ and $p = 1$. Under a fixed alternative $g \neq g_0$, if assumption C holds and if both g and g_0 satisfy assumption B for some $r > k + 1$ and if $nh^{4k+2} \to \infty$ and $nh^{4r-2} \to 0$, then*

$$\sqrt{n}\big(T_n - \|g^{(k)} - g_0^{(k)}\|^2\big) \xrightarrow{\mathcal{L}} N(0, \tau_D^2),$$

where

$$\tau_D^2 = Var\Big(\mathcal{F}^{-1}\big(s^k(\Phi_g(s) - \Phi_{g_0}(s))\big)(X_1)\Big).$$

Remark 1. It is possible to combine the estimators (2) and (20) in order to estimate the derivative of a density of a random variable which is observed under noise (cf. [9, 2]). In the univariate case it would then be possible to combine Theorems 1 and 3.

Acknowledgements: This research was supported by the Deutsche Forschungsgemeinschaft, Grant MU 1230/8-1. We would like to thank T. Gneiting for pointing out some references, and the associate editor as well as two referees for detailed comments, which led to an improved presentation.

Appendix

Proof of Theorem 1. Note that $\Phi_{\hat{f}_n}(t) = \Phi_K(th)\frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)}$ and let $\Phi_n(t) = E\Phi_{\hat{f}_n}(t) = \Phi_K(th)\Phi_f(t)$, $\Phi_{n,0}(t) = \Phi_K(th)\Phi_{f_0}(t)$. From Parseval's equation,

$$\begin{aligned}
&(2\pi)^p\, T_n = \int_{\mathbb{R}^p} \big|\Phi_{\hat{f}_n}(t) - \Phi_{n,0}(t)\big|^2\, dt \\
= \;& \int_{\mathbb{R}^p} \big|\Phi_{n,0}(t) - \Phi_n(t)\big|^2\, dt + \int_{\mathbb{R}} \big|\Phi_{\hat{f}_n}(t) - \Phi_n(t)\big|^2\, dt \\
&+2\Re\Big(\int_{\mathbb{R}^p} \big(\Phi_{\hat{f}_n}(t) - \Phi_n(t)\big)\overline{\big(\Phi_n(t) - \Phi_{n,0}(t)\big)}\, dt\Big) \\
= \;& \int_{\mathbb{R}^p} \big|\Phi_K(th)\big|^2\big|\Phi_{f_0} - \Phi_f\big|^2(t)\, dt + n^{-2}\sum_{j=1}^n \int_{\mathbb{R}^p} \frac{\big|\Phi_K(th)\big|^2}{|\Phi_\psi(t)|^2}\big|e^{it\cdot X_j} - \Phi_g(t)\big|^2\, dt \\
&+ 2n^{-2}\sum_{1\le j<k\le n} \Re\Big(\int_{\mathbb{R}^p} \frac{\big|\Phi_K(th)\big|^2}{|\Phi_\psi(t)|^2}\big(e^{it\cdot X_j} - \Phi_g(t)\big)\overline{\big(e^{it\cdot X_k} - \Phi_g(t)\big)}\, dt\Big) \qquad (21)\\
&+ 2n^{-1}\sum_{k=1}^n \Re\Big(\int_{\mathbb{R}^p} \big|\Phi_K(th)\big|^2\big(e^{it\cdot X_k}/\Phi_\psi(t) - \Phi_f(t)\big)\overline{\big(\Phi_f(t) - \Phi_{f_0}(t)\big)}\, dt\Big) \\
= \;& C_{N,h} + T_A + T_B + T_C.
\end{aligned}$$

Observe that all integrals in (21) are in fact real valued, so that we can skip the $\Re()$ in the following. For example, for the last term,

$$\begin{aligned}
&\int_{\mathbb{R}^p} \big|\Phi_K(th)\big|^2\overline{\big(e^{it\cdot X_k}/\Phi_\psi(t) - \Phi_f(t)\big)}\big(\Phi_f(t) - \Phi_{f_0}(t)\big)\, dt \\
= \;& \int_{\mathbb{R}^p} \big|\Phi_K(-th)\big|^2\big(e^{-it\cdot X_k}/\Phi_\psi(-t) - \Phi_f(-t)\big)\big(\Phi_f(t) - \Phi_{f_0}(t)\big)\, dt,
\end{aligned}$$

since $\left|\Phi_K(th)\right|^2$ is symmetric and since $\overline{\phi_f(t)} = \phi_f(-t)$ for real-valued f. Substituting $s = -t$ shows that the term is invariant under complex conjugation and hence is real. The other term is dealt with similarly.

Now let us consider the diagonal term T_A.

Lemma A.1. *Under the assumptions A and B, we have that $C_{M,h} \approx h^{-(2\beta+p)}$ and that*

$$T_A = C_{M,h}/n + O(1/n) + O_P\Big(1/(n^{3/2}h^{2\beta+p})\Big). \tag{22}$$

Proof. To compute the order of $C_{M,h}$, notice that

$$C_{M,h} \approx \int_{\mathbb{R}^p} |\Phi_K(ht)|^2 |t|^{2\beta}\, dt = h^{-(2\beta+p)} \int_{\mathbb{R}^p} |\Phi_K(u)|^2 |u|^{2\beta}\, du.$$

Let us compute the expectation of T_A. Since

$$\begin{aligned} E\left|e^{it\cdot X_j} - \Phi_g(t)\right|^2 &= E\big(1 + |\Phi_g(t)|^2 - e^{it\cdot X_j}\overline{\Phi_g(t)} - e^{-it\cdot X_j}\Phi_g(t)\big) \\ &= 1 - |\Phi_g(t)|^2 = 1 - |\Phi_\psi(t)|^2 |\Phi_f(t)|^2, \end{aligned}$$

we get

$$ET_A = \frac{C_{M,h}}{n} - \frac{1}{n}\int_{\mathbb{R}^p} \left|\Phi_K(th)\right|^2 |\Phi_f(t)|^2\, dt.$$

From assumption B, $\Phi_f \in L_2$, therefore the second term is $O(1/n)$. Moreover,

$$\begin{aligned} \operatorname{Var} T_A &= 1/n^3 \operatorname{Var}\Big[\int_{\mathbb{R}^p} \frac{\left|\Phi_K(th)\right|^2}{|\Phi_\psi(t)|^2}\left|e^{it\cdot X_1} - \Phi_g(t)\right|^2 dt\Big] \\ &\le 1/n^3 E\Big|\int_{[-1/h,1/h]^p} \frac{1}{|\Phi_\psi(t)|^2}\big(1 + |\Phi_g(t)|^2 - e^{it\cdot X_1}\overline{\Phi_g(t)} - e^{-it\cdot X_j}\Phi_g(t)\big)\, dt\Big|^2 \\ &\le 1/n^3 \Big(\int_{[-1/h,1/h]^p} \frac{1}{|\Phi_\psi(t)|^2}\big(1 + |\Phi_g(t)|^2 + 2|\Phi_g(t)|\big)\, dt\Big)^2 \\ &\le 1/n^3 \Big(O(h^{-(2\beta+p)}) + 2\int_{[-1/h,1/h]^p} \frac{|\Phi_f(t)|}{|\Phi_\psi(t)|}\, dt + \int_{[-1/h,1/h]^p} |\Phi_f(t)|^2\, dt\Big)^2 \\ &\le 1/n^3 O(h^{-(4\beta+2p)}). \end{aligned}$$

□

Next we give the asymptotic variance of T_B.

Lemma A.2. *Under the assumptions A and B, we have that $C_{V,h} \approx h^{-(4\beta+p)}$ and that*

$$ET_B^2 = 2/n^2\Big[C_{V,h} + O\big(1/h^{2\max(\beta+p-r,0)}\big)\Big]. \tag{23}$$

Proof. Set

$$H_n(x,y) = \int_{\mathbb{R}^p} \frac{\left|\Phi_K(th)\right|^2}{|\Phi_\psi(t)|^2}\big(e^{it\cdot x} - \Phi_g(t)\big)\overline{\big(e^{it\cdot y} - \Phi_g(t)\big)}\, dt. \tag{24}$$

Then

$$E\big(H_n(X_1,X_2)\big)^2 = E\Big[\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \frac{|\Phi_K(th)|^2}{|\Phi_\psi(t)|^2}\frac{|\Phi_K(sh)|^2}{|\Phi_\psi(s)|^2}\big(e^{it\cdot X_1}-\Phi_g(t)\big)\overline{\big(e^{it\cdot X_2}-\Phi_g(t)\big)} \big(e^{is\cdot X_1}-\Phi_g(s)\big)\overline{\big(e^{is\cdot X_2}-\Phi_g(s)\big)}\,ds\,dt\Big]. \tag{25}$$

Expanding this expression, we compute the different terms separately. Let us start with the variance-dominating term

$$E\Big(\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{\big|\Phi_K(th)\big|^2}{|\Phi_\psi(t)|^2}\frac{\big|\Phi_K(sh)\big|^2}{|\Phi_\psi(s)|^2}\cos\big(t\cdot(X_1-X_2)\big)\cos\big(s\cdot(X_1-X_2)\big)\,ds\,dt\Big).$$

From addition theorems for the cosine,

$$E\cos\big(t\cdot(X_1-X_2)\big)\cos\big(s\cdot(X_1-X_2)\big) = 1/2\big(|\Phi_g(t+s)|^2+|\Phi_g(t-s)|^2\big)$$

Since Φ_K and $|\Phi_\psi|^2$ are symmetric, it follows that

$$E\Big(\int_{\mathbb{R}^p}\frac{\big|\Phi_K(th)\big|^2}{|\Phi_\psi(t)|^2}\cos\big(t\cdot(X_1-X_2)\big)\,dt\Big)^2 = C_{V,h}.$$

Let us compute the order of $C_{V,h}$. From assumption A, this order is the same as that of

$$\begin{aligned}
&\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{|\Phi_K(ht)|^2|\Phi_K(hs)|^2|t|^{2\beta}|s|^{2\beta}}{(1+(s+t)^2)^\beta}|\Phi_f(s+t)|^2\,ds\,dt\\
= {} & h^{-(2p+4\beta)}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{|\Phi_K(x)|^2|\Phi_K(y)|^2|x|^{2\beta}|y|^{2\beta}}{(1+((x+y)/h)^2)^\beta}|\Phi_f((x+y)/h)|^2\,dx\,dy\\
= {} & h^{-(p+4\beta)}\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}|\Phi_K(u)|^2|\Phi_K(hw-u)|^2|u|^{2\beta}|hw-u|^{2\beta}\frac{|\Phi_f(w)|^2}{(1+w^2)^\beta}\,dw\,du\\
\sim {} & h^{-(p+4\beta)}\int_{\mathbb{R}^p}|\Phi_K(u)|^4|u|^{4\beta}\,du\int_{\mathbb{R}^p}\frac{|\Phi_f(w)|^2}{(1+w^2)^\beta}\,dw,
\end{aligned} \tag{26}$$

where $f(h)\sim g(h)$ if and only if $\lim_{h\to 0} f(h)/g(h) = 1$. Now let us consider another term from the expansion of (25). We have

$$\Big|E\Big(e^{it\cdot(X_1-X_2)}\Phi_g(s)e^{-is\cdot X_2}\Big)\Big| = |\Phi_g(s)||\Phi_g(t)||\Phi_g(t+s)| \le |\Phi_g(t)||\Phi_g(s)|.$$

Therefore

$$\begin{aligned}
&E\Big[\int_{\mathbb{R}^p}\int_{\mathbb{R}^p}\frac{|\Phi_K(ht)|^2}{|\Phi_\psi(t)|^2}\frac{|\Phi_K(hs)|^2}{|\Phi_\psi(s)|^2}e^{it\cdot(X_1-X_2)}\Phi_g(s)e^{-is\cdot X_2}\,ds\,dt\Big]\\
= {} & O\Big(\big(\int_{[-1/h,1/h]^p}\frac{|\Phi_f(t)|}{|\Phi_\psi(t)|}\,dt\big)^2\Big) = O\big(1/h^{2\max(\beta+p-r,0)}\big).
\end{aligned}$$

The other terms in (25) are dealt with similarly. Thus

$$EH_n^2(X_1,X_2) = C_{V,h} + O\big(1/h^{2\max(\beta+p-r,0)}\big), \tag{27}$$

and the lemma is proved. □

Proof of Theorem 1 continued. Under the hypothesis $f = f_0$ we have $T_C = C_{N,h} = 0$. From Lemmas A.1 and A.2,

$$(2\pi)^p \, T_n = C_{M,h}/n + O(1/n) + O_P\Big(1/(n^{3/2}h^{2\beta+p})\Big) + T_B,$$

and it suffices to show asymptotic normality of T_B. To this end we will apply Theorem 1 in [16], and thus have to check that

$$\frac{EH_n^4(X_1,X_2)}{n\big[EH_n^2(X_1,X_2)\big]^2} \to 0 \quad \text{and} \quad \frac{EG_n^2(X_1,X_2)}{\big[EH_n^2(X_1,X_2)\big]^2} \to 0, \tag{28}$$

where $H_n(x,y)$ is defined in (24) and

$$G_n(x,y) = E\big(H_n(X_1,x)H_n(X_1,y)\big).$$

Expanding $EH_n^4(X_1,X_2)$, all terms can be estimated separately. We give an example of this procedure. From addition theorems for the cosine, one shows that

$$\begin{aligned}
& E\cos\big(t_1\cdot(X_1-X_2)\big)\cos\big(t_2\cdot(X_1-X_2)\big)\cos\big(t_3\cdot(X_1-X_2)\big)\cos\big(t_4\cdot(X_1-X_2)\big)\\
= \; & 1/8\Big(|\Phi_g(t_1+t_2+t_3+t_4)|^2 + |\Phi_g(t_1-t_2+t_3+t_4)|^2 + |\Phi_g(t_1+t_2-t_3+t_4)|^2\\
& +|\Phi_g(t_1+t_2+t_3-t_4)|^2 + |\Phi_g(t_1-t_2-t_3+t_4)|^2 + |\Phi_g(t_1-t_2+t_3-t_4)|^2\\
& +|\Phi_g(t_1+t_2-t_3-t_4)|^2 + |\Phi_g(t_1-t_2-t_3-t_4)|^2\Big).
\end{aligned}$$

Thus

$$\begin{aligned}
& \int_{\mathbb{R}^p}\cdots\int_{\mathbb{R}^p} E\prod_{i=1}^4 \frac{|\Phi_K(ht_i)|^2}{|\Phi_\psi(t_i)|^2}\cos\big(t_i\cdot(X_1-X_2)\big)\,dt_1\ldots dt_4\\
= \; & \int_{\mathbb{R}^p}\cdots\int_{\mathbb{R}^p} \prod_{i=1}^4 \frac{|\Phi_K(ht_i)|^2}{|\Phi_\psi(t_i)|^2}|\Phi_\psi(t_1+t_2+t_3+t_4)|^2|\Phi_f(t_1+t_2+t_3+t_4)|^2\,dt_1\ldots dt_4.
\end{aligned}$$

A computation similar to that in (26) shows that this term is $O(1/h^{3p+8\beta})$. The other terms are dealt with similarly, and we get

$$EH_n^4(X_1,X_2) = O(1/h^{3p+8\beta}). \tag{29}$$

Now let us consider G_n. We have

$$EG_n^2(X_3,X_4) = E\big(H_n(X_1,X_3)H_n(X_1,X_4)H_n(X_2,X_3)H_n(X_2,X_4)\big).$$

This expression is again expanded and the terms are estimated separately. Consider for example

$$\begin{aligned}
& \Big|\int_{\mathbb{R}^p}\cdots\int_{\mathbb{R}^p} E\prod_{i=1}^4 \frac{|\Phi_K(ht_i)|^2}{|\Phi_\psi(t_i)|^2}\cos\big(t_1\cdot(X_1-X_3)\big)\cos\big(t_2\cdot(X_1-X_4)\big)\cos\big(t_3\cdot(X_2-X_3)\big)\\
& \cos\big(t_4\cdot(X_2-X_4)\big)dt_1\ldots dt_4\Big|\\
\leq \; & \int_{\mathbb{R}^p}\cdots\int_{\mathbb{R}^p} \prod_{i=1}^4 \frac{|\Phi_K(ht_i)|^2}{|\Phi_\psi(t_i)|^2}|\Phi_g(t_1+t_2)||\Phi_g(t_1+t_3)||\Phi_g(t_3+t_4)||\Phi_g(t_2+t_4)|\,dt_1\ldots dt_4.
\end{aligned}$$

Computing as in (26) shows that this term is $O(1/h^{p+8\beta})$, and we get

$$EG_n^2(X_1, X_2) = O(1/h^{p+8\beta}) \tag{30}$$

From (27), (29) and (30) it follows that (28) holds. □

Proof of Proposition 1. The formula for the expectation follows from (21) and (22) and straightforward computations using the properties of Φ_K. From (22) and (23), the variances of T_A and T_B are of lower order than those given in (15). Thus it remains to estimate the variance of T_C. We have that

$$\begin{aligned}
& E\Big|\int_{\mathbb{R}^p} |\Phi_K(th)|^2 \big(e^{it\cdot X_k}/\Phi_\psi(t) - \Phi_f(t)\big) \overline{\big(\Phi_f(t) - \Phi_{f_0}(t)\big)}\, dt\Big|^2 \qquad (31)\\
= & -\Big|\int_{\mathbb{R}^p} |\Phi_K(th)|^2 \Phi_f(t) \overline{\big(\Phi_f(t) - \Phi_{f_0}(t)\big)}\, dt\Big|^2 \\
& + \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \frac{\Phi_f(t-s)\Phi_\psi(t-s)}{\Phi_\psi(t)\Phi_\psi(-s)} |\Phi_K(th)|^2 \overline{\big(\Phi_f(t) - \Phi_{f_0}(t)\big)}\, |\Phi_K(sh)|^2 \big(\Phi_f(s) - \Phi_{f_0}(s)\big)\, dt\, ds.
\end{aligned}$$

We consider the case $\beta \geq r_0$, the other cases are dealt with similarly. The first term on the right-hand side of (31) is bounded, therefore it suffices to estimate the second term, which is bounded by

$$\begin{aligned}
& C \int_{[-1/h,1/h]^p} \int_{[-1/h,1/h]^p} \frac{|t|^{\beta - r_0} |s|^{\beta - r_0}}{(1 + |s+t|^{\beta + r_0})} ds\, dt \\
\leq \; & C\, h^{-(p+2(\beta - r_0))} \int_{[-1,1]^p} \int_{[-1/h,1/h]^p + u} \frac{|u|^{\beta - r_0} |hw - u|^{\beta - r_0}}{1 + |w|^{\beta + r_0}}\, dw\, du \\
\sim \; & C\, h^{-(p+2(\beta - r_0))} \int_{[-1,1]^p} |u|^{2(\beta - r_0)}\, du \int_{\mathbb{R}^p} \frac{1}{1 + |w|^{\beta + r_0}}\, dw.
\end{aligned}$$

□

Proof of Theorem 2. We start by observing that the remainder terms in (14), when multiplied with $\sqrt{n}$, tend to 0 due to the assumptions of the theorem. The variance-dominating term in the decomposition (21) is again T_C. To compute the constant τ^2 in the variance, we have from the dominated convergence theorem

$$\begin{aligned}
& \int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \Phi_f(t-s)\Phi_\psi(t-s) |\Phi_K(sh)|^2 \frac{\Phi_f(s) - \Phi_{f_0}(s)}{\Phi_\psi(-s)}\, ds\; |\Phi_K(th)|^2 \frac{\overline{\Phi_f(t) - \Phi_{f_0}(t)}}{\Phi_\psi(t)}\; dt \\
\to \; & \int_{\mathbb{R}^p} (\Phi_f \Phi_\psi) * \Big(\frac{\Phi_f - \Phi_{f_0}}{\overline{\Phi_\psi}}\Big)(t)\; \frac{\overline{\Phi_f(t) - \Phi_{f_0}(t)}}{\Phi_\psi(t)}\; dt = (2\pi)^p \int_{\mathbb{R}^p} g(x) \Big|\mathcal{F}^{-1}\Big(\frac{\Phi_f - \Phi_{f_0}}{\overline{\Phi_\psi}}\Big)(x)\Big|^2\; dx,
\end{aligned}$$

using the Plancherel formula in the last equality. The value of τ^2 now follows from (31). To obtain asymptotic normality, we apply the Lyapounov central limit theorem to T_C. Indeed, Lyapounov's condition is satisfied since

$$\begin{aligned}
& E\Big|\int_{\mathbb{R}^p} |\Phi_K(th)|^2 \big(e^{itX_k}/\Phi_\psi(t) - \Phi_f(t)\big) \overline{\big(\Phi_f(t) - \Phi_{f_0}(t)\big)}\, dt\Big|^4 \\
\leq \; & \Big(\int_{\mathbb{R}^p} \big(1/|\Phi_\psi(t)| + |\Phi_f(t)|\big)\big(|\Phi_f(t)| + |\Phi_{f_0}(t)|\big)\, dt\Big)^4 < \infty
\end{aligned}$$

is bounded. This completes the proof of the theorem. □

References

[1] Bickel, P. J. and Rosenblatt, M. (1973), On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.

[2] Delaigle, A. and Gijbels, I. (2002), Estimation of integrated squared density derivatives from a contaminated sample. *J. R. Statist. Soc. B* **64**, 869–886.

[3] Delaigle, A. and Gijbels, I. (2004), Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45**, 249–267.

[4] Dette, H. (1999), A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012–1040.

[5] Dette, H. and Bachmann, D. (2004), A note on the Bickel-Rosenblatt test in autoregressive time series. Technical Report, Ruhr-Universität Bochum.

[6] Dette, H. and Munk, A. (1998), Validation of linear regression models. *Ann. Statist.* **26**, 778–800.

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

[8] Diggle, P. and Hall, P. (1993), A Fourier Approach to nonparametric deconvolution of a density estimate. *J. R. Stat. Soc. Ser. B* **55**, 523–531.

[9] Fan, J. (1991a), On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.

[10] Fan, J. (1991b), Asymptotic normality for deconvolution kernel density estimators. *Sankhya Ser. A* **53**, 97–110.

[11] Fan, Y. and Ullah, A. (1999), On goodness-of-fit tests for weakly dependent processes using kernel method. *J. Nonparametr. Statist.* **11**, 337–360.

[12] 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.

[13] Gneiting, T. (1997), Normal scale mixtures and dual probability densities. *J. Statist. Comput. Simul.* **59**, 375–384.

[14] Gneiting, T. and Schlather, M. (2004), Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Rev.* **46**, 269–282.

[15] Groeneboom, P. and Jongbloed, G. (2003), Density estimation in the uniform deconvolution model. *Statist. Neerlandica* **57**, 136–157.

[16] Hall, P. (1984), Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14**, 1–16.

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

[18] Ibragimov, I. A. and Khas'minskij, R. Z. (1982), Estimation of distribution density belonging to a class of entire functions. *Theory Probab. Appl.* **27**, 551–562.

[19] Neumann, M. H. and Paparoditis, E. (2000), On bootstrapping L_2-type statistics in density testing. *Statist. Probab. Lett.* **50**, 137–147.

[20] Nikitin, Y. (1995), *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, Cambridge.

[21] Paparoditis, E. (2000), Spectral density based goodness of fit tests for time series models. *Scand. J. Statist.* **27**, 145–176.

[22] Piterbarg, V. I. and Penskaya, M. Y. (1993), On asymptotic distribution of integrated square error of an estimate of a component of a convolution. *Math. Meth. Statist.* **2**, 30–41.

[23] Politis, D. N. and Romano, J. P. (1999), Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal.* **68**, 1–25.

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

[25] Stefanski, L. and Carroll, R. J. (1990), Deconvoluting kernel density estimators. *Statistics* **21**, 169–184.

[26] van Es, A. J. and Kok, A. R. (1998), Simple kernel estimators for certain nonparametric deconvolution problems. *Statist. Probab. Lett.* **39**, 151–160.

[27] van Es, A. J. and Uh, H.-W. (2004), Asymptotic normality for kernel type deconvolution density estimators: Crossing the Cauchy boundary. *J. Nonparametr. Statist.* **16**, 261–277.

[28] van Rooij, A. C. M., Ruymgaart F. H. and van Zwet, W. R. (1999), Asymptotic efficiency of inverse estimators. *Theory Probab. Appl.* , **44**, 722–737.

[29] Zhang, C. H. (1990), Fourier methods for estimating mixing densities and distributions. *Ann Statist.* **18**, 806–831.