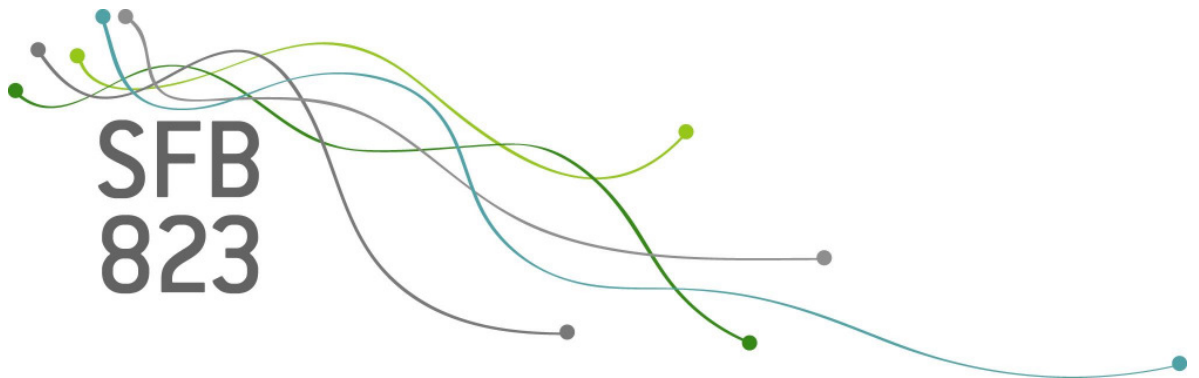


SFB
823

Model checks in inverse regression models with convolution-type operators

Nicolai Bissantz, Holger Dette,
Katharina Proksch

Nr. 29/2009



Discussion Paper

Model checks in inverse regression models with convolution-type operators

Nicolai Bissantz¹, Holger Dette¹ and Katharina Proksch¹

¹Fakultät für Mathematik

Ruhr-Universität Bochum, Germany

October 16, 2009

Abstract

We consider the problem of testing parametric assumptions in an inverse regression model with a convolution-type operator. An L_2 -type goodness-of-fit test is proposed which compares the distance between a parametric and a nonparametric estimate of the regression function. Asymptotic normality of the corresponding test statistic is shown under the null hypothesis and under a general nonparametric alternative with different rates of convergence in both cases. The feasibility of the proposed test is demonstrated by means of a small simulation study. In particular, the power of the test against certain types of alternative is investigated.

Keywords: Inverse problems, Model selection, Goodness-of-fit tests, Limit theorems for quadratic forms.

1 Introduction

In this paper we consider the model

$$Y_{i,n} = K_\psi f(x_{i,n}) + \varepsilon_{i,n}, \quad i = -n, \dots, n \quad (1)$$

where $x_{i,n} = i/(na_n)$ are fixed design-points with $a_n \rightarrow 0$, $na_n \rightarrow \infty$ for $n \rightarrow \infty$, and $\varepsilon_{-n,n}, \dots, \varepsilon_{n,n}$ are real-valued independent identically distributed errors with $E[\varepsilon_{i,n}] = 0$, $\sigma^2 := E[\varepsilon_{i,n}^2] < \infty$ and existing fourth moments. In (1) K_ψ denotes the convolution operator defined by

$$g(x) := K_\psi f(x) = f * \psi(x) = \int_{\mathbb{R}} \psi(x-t)f(t)dt, \quad (2)$$

¹Address for correspondence: Katharina Proksch, Ruhr-Universität Bochum, Fakultät für Mathematik, NA 3/75, Universitätsstr. 150, D-44780 Bochum, Germany, email: katharina.proksch@rub.de, Fon: +49/234/32-23287, Fax: +49/551/32-14559

where ψ is a given function, which will be specified below. The reconstruction of the function f from $g = K_\psi f$ at any location x on the real line requires (at least asymptotically) information about g on the full real line. This is achieved by using a design which includes an additional sequence $a_n \rightarrow 0$ to ensure that some of the design points $x_{k,n}$ will converge to infinity. Recovering the function f from the observations $(x_{-n,n}, Y_{-n,n}), \dots, (x_{n,n}, Y_{n,n})$ is a statistical inverse problem (see for example Bissantz et al. (2007) or Mair and Ruymgaart (1996)). Estimation of the regression function f in model (1) and associated confidence bands have been discussed for Fourier based estimators in Bissantz and Birke (2009) and Birke et al. (2009).

In inverse regression models with convolution operator it is often assumed that the signal f is periodic on a compact interval, say $[0, 1]$ (see e.g. Cavalier and Tsybakov (2002)). In this case K_ψ is a convolution operator on $[0, 1]$ with periodic kernel ψ . However, in many examples such as the reconstruction of astronomical and biological images from telescopic and microscopic imaging devices this assumption is not realistic, because the signal in general is not periodic.

Some applications of model (1) can be found in Lauer et al. (2005), who used a parametric model to investigate the surface brightness data from Hubble Space Telescope observations of the centers of early-type galaxies, and in Claxton et al. (2005) and Spring and Inoué (1997), where the PSF is modelled by a theoretically predicted Airy pattern, which can be tested by pilot observations e.g. of a spherical bead based on model (1). Because any statistical analysis depends sensitively on the parametric form of the function f many authors point out that it is important to check these assumptions by means of a statistical test.

The problem of checking (parametric) model assumptions regarding the function g has found considerable interest in direct regression models of the form $Y_{i,n} = g(x_{i,n}) + \varepsilon_{i,n}$. The (weighted) L_2 -distance of a nonparametric kernel estimator of the function and a smoothed version of a parametric estimate was used by Härdle and Mammen (1993) to test the validity of a parametric model, and the (weighted) L_2 -norm of the function and its derivatives was estimated in Huang and Fan (1999) by integrating the corresponding coefficients of a local polynomial estimator. These results were derived in a random design setting. Likewise, in a regression model with fixed design on a compact interval, a test statistic can be based on the difference of a nonparametric kernel-based estimator and a parametric estimator of the variance (see Dette (1999), where the asymptotic distribution of the test statistic under both the hypothesis of a linear model and under fixed alternatives is derived). Bissantz et al. (2005) constructed L_2 -based tests for parametric assumptions on time- or band-limited functions (i.e. functions where either the function or its Fourier transform has compact support) under noise. They use the sinc-kernel estimator for the regression function, which can be understood as a Fourier-based estimator

for the case of a direct regression model. Closely related to these approaches are the results in Paparoditis (2000) and Dette and Spreckelsen (2003) which consider goodness-of-fit testing problems for the spectral density.

Similarly, goodness-of-fit tests for the parametric form of the density of independent identically distributed random variables have been discussed in the literature. In particular, Bickel and Rosenblatt (1973) 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 Neumann and Paparoditis (2000) and Bachmann and Dette (2005) to test parametric hypotheses about the marginal distribution of stationary processes. Tests for the parametric form of the density in deconvolution problems have been considered in Holzmann et al. (2007) and Butucea (2007). Moreover, testing parametric models in the presence of instrumental variables in a context which is closely related to statistical inverse problems has been considered in Holzmann (2007).

In contrast to the case of direct regression and density estimation, the problem of testing parametric hypotheses regarding the regression function f in inverse models of the type (1) is not well-studied in the literature, and it is the purpose of the present paper to develop a consistent test for this problem. A rather pragmatic approach to this problem is to replace the inverse model (1) by its direct regression analogue

$$Y_{i,n} = g(x_{i,n}) + \varepsilon_{i,n}$$

where g is defined by (2). Note that parametric assumptions on the original function f can be expressed uniquely in terms of parametric assumptions on g . If we assume that convolution with ψ is injective, in principle, all direct testing procedures (e.g. the tests suggested in Härdle and Mammen (1993) or Dette (1999)) could be applied to test such equivalent parametric assumptions on g as well (note that the observations $Y_{i,n}$ at our disposal have mean $g(x_{i,n})$). However, this procedure is generally not appropriate for detecting important 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 function f_0 has an arbitrarily large L_2 -distance to the parametric model in the domain of f , whereas the corresponding convolution $g_0 = f_0 * \psi$ is very close to the parametric model in the domain of g . Hence direct application of tests to the (observable) data $(x_{-n,n}, Y_{-n,n}), \dots, (x_{n,n}, Y_{n,n})$ will result in an inefficient procedure for those alternatives which can hardly be distinguished from the null in terms of g . In this paper we take a different approach to construct tests for parametric hypotheses in the inverse statistical regression model (1), which is based on an L_2 -distance between a nonpara-

metric and a parametric estimate of the regression function f . Therefore our approach deals directly with the original function f . In Section 2 we introduce a Fourier based estimate of the regression function and state some basic assumptions, which are required in the following discussion. In Section 3 we discuss the problem of testing simple hypothesis, i.e. $H_0 : f = f_0$, where f_0 is a given function. We establish asymptotic normality of the L_2 -type test statistic both under the hypothesis and the alternative with different rates of convergence corresponding to both cases. In Section 4 we use these results to develop a test for the composite hypothesis

$$H_0 : f \in \mathcal{M}_\Theta, \quad (3)$$

where $\mathcal{M}_\Theta := \{f(\vartheta, \cdot) | \vartheta \in \Theta \subset \mathbb{R}^n\}$ is a given parametric class. The practical performance of the proposed tests is illustrated by means of a small simulation study in Section 5. Finally, to keep the paper more readable, the proofs are deferred to Section 6.

2 Prerequisites. Estimator, notation and assumptions

In this Section we introduce the estimator and some important assumptions regarding the function f and the convolution function ψ . Throughout this paper let $\mathcal{F}\psi$ denote the Fourier transform of ψ , which is defined by

$$\mathcal{F}\psi(\omega) = \int_{\mathbb{R}} \psi(t) \exp(-i\omega t) dt.$$

This yields

$$\psi(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}\psi(\omega) \exp(i\omega t) d\omega \quad (4)$$

and

$$\mathcal{F}g = \mathcal{F}f \cdot \mathcal{F}\psi. \quad (5)$$

It follows from equations (4) and (5) that

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathcal{F}g(\omega)}{\mathcal{F}\psi(\omega)} \exp(i\omega t) d\omega. \quad (6)$$

For all purposes in this paper we further assume

$$\mathcal{F}\psi(t) \neq 0 \quad \forall \quad t \in \mathbb{R}, \quad (7)$$

which yields injectivity of the operator K_ψ (see, e.g. Holzmann et al. (2007)). On the basis of the representation (6) we propose as an estimator for the function f in model (1)

$$\hat{f}_n(t) = \frac{1}{2\pi} \int_{-\frac{1}{h}}^{\frac{1}{h}} \frac{\hat{\mathcal{F}}g(\omega)}{\mathcal{F}\psi(\omega)} \exp(i\omega t) d\omega, \quad (8)$$

(see for example Diggle and Hall (1993)), where $\hat{\mathcal{F}}g(\omega)$ is the empirical Fourier transform of g defined by

$$\hat{\mathcal{F}}g(\omega) = \frac{1}{na_n} \sum_{j=-n}^n Y_{j,n} \exp(-i\omega x_{j,n})$$

and h is a smoothing parameter satisfying $h \rightarrow 0$ as $n \rightarrow \infty$. Note that a larger value of h yields a smoother estimator \hat{f}_n . For our asymptotic considerations we require several assumptions regarding the smoothness of the functions f and ψ . Our first assumption refers to the asymptotic behavior of the Fourier transform $\mathcal{F}\psi$.

Assumption A: There exists a constant $\beta > 0$ and a constant $C \in \mathbb{C} \setminus \{0\}$ such that

$$\mathcal{F}\psi(\omega)|\omega|^\beta \longrightarrow C, \quad \text{as } |\omega| \rightarrow \infty.$$

Assumption **A** holds, e.g., for the density function ψ of the centered double-exponential (or Laplace) distribution, given by $\psi(t) = \frac{\lambda}{2} e^{-\lambda|t|}$ with

$$\mathcal{F}\psi(\omega) = \frac{1}{1 + \frac{\omega^2}{\lambda^2}}.$$

In this case we have $\beta = 2$ and $C = \lambda^2$. The second assumption requires integrability of the Fourier transform $\mathcal{F}f$ and the convolution $K_\psi f$.

Assumption B: For $\mathcal{F}f$ and $K_\psi f$ we impose that there exist constants $\alpha > \beta$ and $\gamma > 0$ such that

$$\begin{aligned} \alpha > \beta &: \int_{\mathbb{R}} |\mathcal{F}f(\omega)| |\omega|^\alpha d\omega < \infty \\ \gamma > 0 &: \int_{\mathbb{R}} |K_\psi f(t)| |t|^\gamma dt < \infty. \end{aligned}$$

3 Testing of simple hypothesis

For a test of the simple hypothesis $H_0 : f = f_0$ against the alternative $H_1 : f \neq f_0$ (for a given function f_0) we use the L_2 distance

$$T_n = \int_{\mathbb{R}} |\hat{f}_n(t) - f_0(t)|^2 dt, \tag{9}$$

where \hat{f}_n is the estimator for the regression function f defined in (8). Our first theorem demonstrates asymptotic normality of the statistic T_n under the null hypothesis and some regularity assumptions on the index β of ill-posedness of the deconvolution problem, the smoothness parameter γ of the function f and the convergence rates of the bandwidth h and a_n . For a precise statement of the result, we make the following

Assumption C: For a_n and h we assume that for $n \rightarrow \infty$,

$$\begin{aligned} na_n h^{\beta+1} &\longrightarrow \infty, \\ na_n^\gamma h^\beta &\longrightarrow c \in \mathbb{R}_0^+, \\ a_n h \ln^2(n) &\longrightarrow 0, \\ ha_n^{-\frac{1}{2}} &\longrightarrow 0, \\ a_n^\gamma h^{-\beta-1} &\longrightarrow 0. \end{aligned}$$

Assumption C ensures that the rate of convergence to zero of a_n and h is neither too fast nor too slow. Throughout this paper the symbol $a_n \sim b_n$ implies $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.

Theorem 1. *Suppose that assumptions A, B and C are satisfied. Then, under the null hypothesis $H_0 : f = f_0$*

$$n \left(a_n^3 h^{4\beta+1} \right)^{\frac{1}{2}} \cdot \left(T_n - \frac{2\sigma^2 C_1}{\pi \cdot (2\beta + 1) n a_n^2 h^{2\beta+1}} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_2),$$

where the constants $C_1, C_2 \in \mathbb{R}^+$ are defined by

$$\int_{-\frac{1}{h}}^{\frac{1}{h}} \frac{1}{|\mathcal{F}\psi(s)|^2} ds \sim C_1 \cdot \int_{-\frac{1}{h}}^{\frac{1}{h}} |s|^{2\beta} ds \quad (10)$$

$$\text{Var}(A_{n,1}) \sim \frac{C_2}{n^2 a_n^3 h^{4\beta+1}} \quad (11)$$

and the random variable $A_{n,1}$ is given by

$$A_{n,1} = \frac{1}{2\pi n^2 a_n^2} \sum_{-n \leq j \neq k \leq n} \varepsilon_j \varepsilon_k \int_{-\frac{1}{h}}^{\frac{1}{h}} \frac{1}{|\mathcal{F}\psi(s)|^2} \exp(-isx_j) \exp(isx_k) ds. \quad (12)$$

Note that the result implies the existence of the limits in (10) and (11). Moreover, similar to the case of testing parametric hypotheses in direct regression models, the bias term of the distribution grows with $(a_n h)^{-1/2}$ (see for example Dette (1999)). Hence, in a practical application of the method, the bias, which depends on the convolution kernel by the constant C_1 , is not (even asymptotically) negligible, and has to be determined either explicitly or by simulation. In the proof of Theorem 1 we will show that $A_{n,1}$ defined in (12) is the dominating term in a decomposition of the difference $T_n - ET_n$.

For a concrete convolution operator specified by the function ψ the constants C_1 and C_2 can be determined either analytically or numerically. For example, if ψ is the density function of a centered Laplace distribution the constants C_1, C_2 are given by

$$C_1 = \frac{1}{\lambda^4} \quad \text{and} \quad C_2 = \frac{4}{9} \cdot \frac{C_1^2 \sigma^4}{\pi} = \frac{4}{9} \cdot \frac{\sigma^4}{\lambda^8 \pi}.$$

We obtain an asymptotic level- α -test $\varphi = \varphi(T_n)$ for the hypothesis $H_0 : f = f_0$, i.e.,

$$\varphi(T_n) = I \left\{ \frac{1}{\sqrt{C_2}} n a_n^{\frac{3}{2}} h^{\frac{9}{2}} \left(T_n - \frac{2\sigma^2 C_1}{\pi \cdot (2\beta + 1) n a_n^2 h^{2\beta+1}} \right) > u_{1-\alpha} \right\}, \quad (13)$$

where $u_{1-\alpha}$ denotes the $1-\alpha$ -quantile of the standard normal distribution and $\hat{\sigma}^2$ denotes an appropriate estimate of the variance σ^2 (see for example (1998)).

Next we consider the asymptotic behaviour of the test statistic T_n under the fixed alternative $H_1 : f \neq f_0$. Here we need slightly different assumptions on a_n and h than in the case of H_0 .

Assumption D:

$$\begin{aligned} h^{2\beta+1} \sqrt{n a_n} &\longrightarrow \infty, \\ n a_n^3 &\longrightarrow \infty \quad \text{and} \\ a_n^\gamma h^{-(2\beta+1)} &= O(1) \end{aligned}$$

as $n \rightarrow \infty$.

Theorem 2. *Assume that assumptions **A**, **B** and **D** hold. Then, under $H_1 : f \neq f_0$*

$$\sqrt{n a_n} \left(T_n - \|f - f_0\|^2 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{2\sigma^2}{\pi} \cdot \left\| \frac{\mathcal{F}f}{\mathcal{F}\psi} \right\|^2 \right).$$

where $\|f\| = (\int_{\mathbb{R}} f^2(t) dt)^{1/2}$ denotes the L_2 -norm of the function f .

Similarly to the proof of Theorem 1, the proof of Theorem 2 is based on the decomposition of the statistic $T_n - ET_n$ where the leading term is not of order $n a_n^{\frac{3}{2}} h^{2\beta+1}$ as under the null hypothesis. Under the alternative H_1 the dominating term in the decomposition of $T_n - ET_n$ is a linear statistic of the random variables $\varepsilon_{-n,n}, \dots, \varepsilon_{n,n}$, which is of order $\sqrt{n a_n}$. In other words the rates of convergence of the test statistic under the hypothesis and the alternative are different, which is typical for this type of statistics (see for example Dette (1999), or Bissantz et al. (2005)). Moreover, due to this difference in the dominating terms of the test statistics, assumption **D** on the convergence rate to zero of h are slightly weaker than under the null hypothesis, i.e. faster convergence to zero of h is permitted.

Note that the estimator $\hat{f}_n(t)$ is only asymptotically unbiased. Hence, since $T_n - ET_n = A_n + 2B_n$, where

$$A_n = \int_{\mathbb{R}} |\hat{f}_n(t) - E\hat{f}_n(t)|^2 dt - E \int_{\mathbb{R}} |\hat{f}_n(t) - E\hat{f}_n(t)|^2 dt$$

and

$$B_n = \int_{\mathbb{R}} (\hat{f}_n(t) - E\hat{f}_n(t))(E\hat{f}_n(t) - f_0(t)) dt$$

includes the bias, B_n may as well be not negligible. It is therefore tempting to use a test statistic, for which $B_n = 0$, in particular for small sample sizes. For this purpose we propose as an alternative statistic

$$\tilde{T}_n = \int_{\mathbb{R}} |\hat{f}_n(t) - \tilde{f}_0(t)|^2 dt, \quad (14)$$

where

$$\tilde{f}_0(t) = \frac{1}{2\pi} \int_{-\frac{1}{h}}^{\frac{1}{h}} \frac{\frac{1}{na_n} \sum_{j=-n}^n K_\psi f_0(x_j) \exp(-i\omega x_j)}{\mathcal{F}\psi(\omega)} \exp(i\omega t) d\omega$$

is an approximation of f_0 satisfying $E[\hat{f}_n(t)] = \tilde{f}_0(t)$. The alternative test statistic \tilde{T}_n turns out to have practical advantages over T_n at least for moderate sample sizes. For further details we refer to the discussion in section 5.

Under $H_0 : f = f_0$, we have $\tilde{f}_0(t) = E\hat{f}_n(t)$ if assumption **C** is satisfied. Using the same decompositions as for the proofs of Theorem 1 and Theorem 2 we obtain similar results for the limit distribution of the test statistic \tilde{T}_n .

Corollary 1. *If assumptions **A**, **B**, **C** and **D** hold, we have to the statistic \tilde{T}_n defined in (14) under $H_0 : f = f_0$*

$$n \left(a_n^3 h^{4\beta+1} \right)^{\frac{1}{2}} \cdot \left(\tilde{T}_n - \frac{2\sigma^2 C_1}{\pi \cdot (2\beta + 1) n a_n^2 h^{2\beta+1}} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_2)$$

and under $H_1 : f \neq f_0$

$$\sqrt{na_n} \left(\tilde{T}_n - \|f - f_0\|^2 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{2\sigma^2}{\pi} \left\| \frac{\mathcal{F}f}{\mathcal{F}\psi} \right\|^2 \right).$$

Corollary 1 yields the following alternative level- α -test $\varphi = \varphi(\tilde{T}_n)$:

$$\varphi(\tilde{T}_n) = I \left\{ \frac{1}{\sqrt{C_2}} n a_n^{\frac{3}{2}} h^{\frac{9}{2}} \left(\tilde{T}_n - \frac{2\sigma^2 C_1}{\pi \cdot (2\beta + 1) n a_n^2 h^{2\beta+1}} \right) > u_{1-\alpha} \right\}. \quad (15)$$

4 Testing of composite hypothesis

In this Section we extend the proposed test to the case of testing for a composite hypothesis of the form

$$H_0 : f \in \mathcal{M}_\Theta, \quad (16)$$

where

$$\mathcal{M}_\Theta := \{f(\vartheta, \cdot) | \vartheta \in \Theta \subset \mathbb{R}^n\}$$

is a finite dimensional parametric family of regression models. For a precise statement of the results, we assume that there exists a unique value ϑ_0 with

$$\vartheta_0 = \arg \min_{\vartheta \in \Theta} \|f - f_\vartheta\|.$$

In particular this means that, under the null hypothesis (16), there exists a unique $\vartheta_0 \in \Theta$ such that $f_{\vartheta_0} = f$. Let $\hat{\vartheta}$ an estimator for the parameter $\vartheta \in \Theta$, which yields the estimator $f(\hat{\vartheta}, \cdot) =: f_{\hat{\vartheta}}$ for $f(\vartheta, \cdot)$ in the parametric model. As test statistic $T_{n, \hat{\vartheta}}$ we now use the (squared) L_2 -distance between $f_{\hat{\vartheta}}$ and \hat{f}_n , that is

$$T_{n, \hat{\vartheta}} := \int_{\mathbb{R}} |f_{\hat{\vartheta}}(t) - \hat{f}_n(t)|^2 dt.$$

Following Neumann and Paparoditis (2000), we use the results obtained for T_n in Section 2 to derive the asymptotic properties of the statistic $T_{n, \hat{\vartheta}}$. To this end we require sufficient conditions regarding the estimate $\hat{\vartheta}$ which yield the asymptotic equivalence of T_n and $T_{n, \hat{\vartheta}}$, i.e.

$$T_n - T_{n, \hat{\vartheta}} = o_P \left(\frac{1}{na_n^{\frac{3}{2}} h^{2\beta + \frac{1}{2}}} \right). \quad (17)$$

Observe that

$$\begin{aligned} T_{n, \hat{\vartheta}} - T_n &= \int_{\mathbb{R}} (f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t))^2 dt + 2 \int_{\mathbb{R}} (\hat{f}_n(t) - f_{\vartheta_0}(t))(f_{\vartheta_0}(t) - f_{\hat{\vartheta}}(t)) dt \\ &= \int_{\mathbb{R}} (f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t))^2 dt + 2E_{n,2}, \end{aligned}$$

where the last identity defines $E_{n,2}$ in an obvious manner. In order to find a suitable set of conditions, which guarantee the approximation (17), we use the notation $f(\vartheta, t) = f_t(\vartheta)$ and the first-order Taylor expansion

$$f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t) = f'_{\vartheta_0}(t)(\hat{\vartheta} - \vartheta_0) + R(\hat{\vartheta}, \vartheta_0, t),$$

with

$$f'_{\vartheta_0}(t) = \frac{\partial}{\partial \vartheta} (f_\vartheta(t)) \Big|_{\vartheta = \vartheta_0}.$$

The Cauchy-Schwarz inequality yields

$$\begin{aligned}
|E_{n,2}| &= \left| \int_{\mathbb{R}} (\hat{f}_n(t) - f_{\vartheta_0}(t))(f_{\vartheta_0}(t) - f_{\hat{\vartheta}}(t)) dt \right| \\
&= \left| \int_{\mathbb{R}} (\hat{f}_n(t) - f_{\vartheta_0}(t))(f'_{\vartheta_0}(t)(\hat{\vartheta} - \vartheta_0) + R(\hat{\vartheta}, \vartheta_0, t)) dt \right| \\
&\leq \left(\left(\int_{\mathbb{R}} R^2(\hat{\vartheta}, \vartheta_0, t) dt \right)^{\frac{1}{2}} + |\hat{\vartheta} - \vartheta_0| \left(\int_{\mathbb{R}} |f'_{\vartheta_0}(t)|^2 dt \right)^{\frac{1}{2}} \right) \\
&\quad \times \left(\int_{\mathbb{R}} |\hat{f}_n(t) - f_{\vartheta_0}(t)|^2 dt \right)^{\frac{1}{2}} \\
&= \left(\left(\int_{\mathbb{R}} R^2(\hat{\vartheta}, \vartheta_0, t) dt \right)^{\frac{1}{2}} + |\hat{\vartheta} - \vartheta_0| \left(\int_{\mathbb{R}} |f'_{\vartheta_0}(t)|^2 dt \right)^{\frac{1}{2}} \right) \\
&\quad \times \left(O_P \left(\frac{1}{na_n^2 h^{2\beta+1}} \right) \right)^{\frac{1}{2}},
\end{aligned}$$

where the estimate in the last identity follows from Theorem 1, since

$$\int_{\mathbb{R}} |\hat{f}_n(t) - f_{\vartheta_0}(t)|^2 dt = T_n$$

under $H_0 : f = f_{\vartheta_0}$. This shows that the following assumption **E** is sufficient for the required asymptotic equivalence (17) of the statistics $T_{n,\hat{\vartheta}}$ and T_n .

Assumption E:

$$\begin{aligned}
\int_{\mathbb{R}} |f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t)|^2 dt &= o_P \left(\frac{1}{na_n^{\frac{3}{2}} h^{2\beta+\frac{1}{2}}} \right) \\
\int_{\mathbb{R}} R^2(\hat{\vartheta}, \vartheta_0, t) dt &= o_P \left(\frac{1}{na_n h^{2\beta}} \right) \\
\int_{\mathbb{R}} (f'_{\vartheta_0}(t))^2 dt &< \infty \\
|\vartheta_0 - \hat{\vartheta}| &= O_P \left(\frac{1}{\sqrt{na_n h^{2\beta}}} \right)
\end{aligned}$$

Theorem 3. *Assume that assumptions **A,B,C** and **E** are satisfied, and that the null hypothesis (16) holds. Then*

$$n (a_n^3 h^{4\beta+1})^{\frac{1}{2}} \left(T_{n,\hat{\vartheta}} - \frac{2\sigma^2 C_1}{\pi(2\beta+1)na_n^2} \left(\frac{1}{h} \right)^{2\beta+1} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_2),$$

where the constants C_1 and C_2 are defined in Theorem 1.

Theorem 3 can be summarized in the following way. Under the null hypothesis the test statistic $T_{n,\hat{\vartheta}}$ has the same limit distribution as T_n provided that the estimate $\hat{\vartheta}$ is consistent with a sufficiently fast rate.

Example 1. To provide an example for a parametric model where assumption **E** is satisfied, consider the case of simple linear regression with

$$\mathcal{M}_{\Theta} = \{f_{\vartheta} = \vartheta \cdot m \mid \vartheta \in \mathbb{R}, \}.$$

where $m \in L_2(\mathbb{R})$ is a given function. Under the hypothesis (16) the function of interest is given by $f = \vartheta_0 \cdot m$ for a unique $\vartheta_0 \in \mathbb{R}$ and the model (1) takes the form

$$Y_i = \vartheta_0 \cdot Km(x_i) + \varepsilon_i, \quad i = -n, \dots, n.$$

For the least squares estimator $\hat{\vartheta}$ for the parameter ϑ , given by

$$\hat{\vartheta} = \arg \min_{\vartheta \in \mathbb{R}} \sum_{i=-n}^n (Y_{i,n} - \vartheta Km(x_{i,n}))^2 = \sum_{j=-n}^n Y_{j,n} \frac{Km(x_{j,n})}{\sum_{k=-n}^n (Km(x_{k,n}))^2},$$

we have

$$|\vartheta - \vartheta_0| = O_P \left(\frac{1}{\sqrt{na_n}} \right),$$

since $m \in L_2(\mathbb{R})$ and $Var(\varepsilon_{i,n}) < \infty$. The first condition of assumption **E** is satisfied since

$$\int |f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t)|^2 dt = |\vartheta - \vartheta_0| \int |m(t)|^2 dt = O_P \left(\frac{1}{n} \right).$$

Moreover,

$$f_{\hat{\vartheta}}(t) - f_{\vartheta_0}(t) = m(t) \cdot (\vartheta - \vartheta_0)$$

with $f'_{\vartheta_0} = m$. Thus the second and the third condition are satisfied as well.

To prove asymptotic normality of the statistic $T_{n,\hat{\vartheta}}$ under a fixed alternative, we can use the same arguments as above to reduce the proof to the proof of Theorem 2. The following assumption yields the asymptotic equivalence of the statistics

$$T_n = \int_{\mathbb{R}} \left| \hat{f}_n(t) - f_{\vartheta_0}(t) \right|^2 dt \quad \text{and} \quad T_{n,\hat{\vartheta}}$$

under a fixed alternative $f \notin \mathcal{M}_{\vartheta}$.

Assumption F

$$\begin{aligned} |\hat{\vartheta} - \vartheta_0| &= o_P \left(\frac{1}{\sqrt{na_n}} \right) \\ \int_{\mathbb{R}} R^2(\hat{\vartheta}, \vartheta_0, t) &= o_P \left(\frac{1}{na_n} \right) \\ \int_{\mathbb{R}} |f'_{\vartheta_0}(t)| dt &< \infty \end{aligned}$$

Theorem 4. *Assume that assumptions **A**, **B**, **D** and **F** are satisfied. Then, under $H_1 : f \notin \mathcal{M}_{\Theta}$*

$$\sqrt{na_n} \left(T_{n,\hat{\vartheta}} - \|f - f_{\vartheta_0}\|^2 \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \frac{2\sigma^2}{\pi} \left\| \frac{\mathcal{F}f}{\mathcal{F}\psi} \right\|^2 \right).$$

5 Finite sample properties

In this Section we present a simulation study of the finite sample properties of the estimator and the testing procedures suggested in the previous sections.

5.1 Simulation setup

We assume that the observations follow model (1), i.e.

$$Y_{k,n} = K_\psi f(x_{k,n}) + \varepsilon_{k,n}, \quad k = -n, \dots, n,$$

where the noise terms $\varepsilon_{k,n}$ are normally distributed with variance σ^2 , and $x_{k,n} = \frac{k}{na_n}$ are the design points. We consider the cases $n = 100$ and $n = 500$, corresponding to sample sizes of 201 and 1001, respectively, and noise levels $\sigma = 0.1$ and $\sigma = 0.25$. The design parameter a_n was chosen as $a_n = \frac{2}{9}$ for sample size 201 and $a_n = \frac{2}{11}$ for sample size 1001. For a practical application, a rule-of-thumb for the selection of a_n would be to choose it such that the observed signal is indistinguishable from noise outside of the support $\left[-\frac{1}{a_n}, \frac{1}{a_n}\right]$ of the design. The operator K_ψ constitutes the convolution on \mathbb{R} of the function of interest f with the function ψ . We consider the functions

$$\begin{aligned} f_1(x) &= \exp(-0.85 \cdot (x - 1)^2) \\ f_2(x) &= \exp(-0.5(x - 0.6)^2) + 1.2 \exp(-1.3(x + 1.45)^2), \end{aligned}$$

and the convolution function ψ is the density of a Laplace distribution with parameter $\lambda = 1$, i.e. $\psi(x) = \frac{1}{2} \exp(-|x|)$.

Figure 1 shows the functions of interest f_1 and f_2 and typical estimates of each of these functions, both for sample sizes 201 and 1001. Here and in the following for each set of values of parameters n and σ the bandwidth was chosen as follows. First, the bandwidth used for these estimates was chosen by visual inspection from 5 sets of data and kept fixed in all subsequent simulation with the same set of parameters n, σ . For sample size 201 the bandwidth is ≈ 0.3 , and for sample size 1001 it is ≈ 0.25 , with only slightly different values for the two different noise levels. Second, for the subsequent simulations of the proposed tests, the bandwidth was chosen $\approx 20\%$ smaller since the test statistic T_n requires undersmoothing estimates.

5.2 Distribution of the test statistic and a bootstrap test

In the following we consider both the original test statistic T_n and its alternative version \tilde{T}_n defined in (9) and (14), respectively. Figure 2 shows simulated distributions of both test statistics

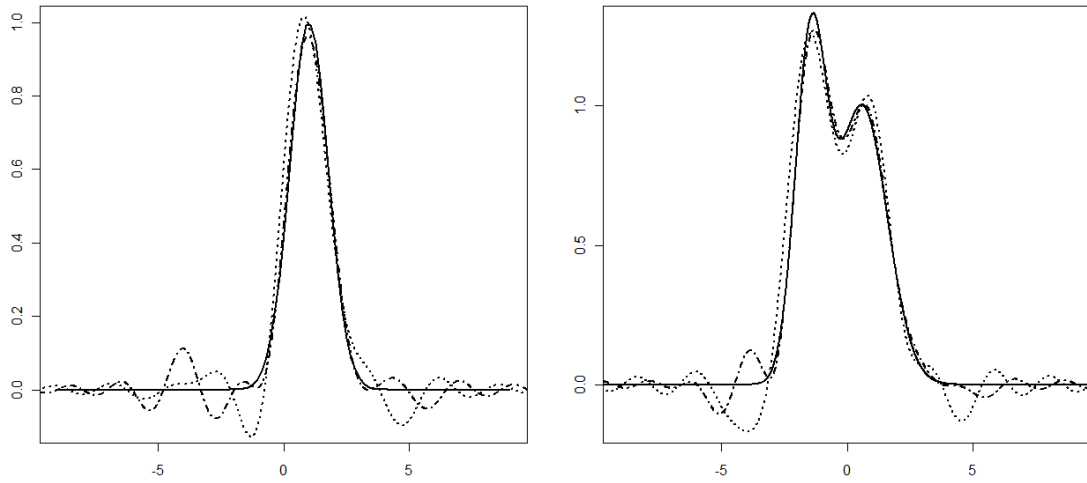


Figure 1: *Function of interest f_i (solid line) and typical estimates $\hat{f}_{i,n}$ from 201 (dotted line) and 1001 observations (dash dotted line), respectively, for $i = 1$ (left panel) and $i = 2$ (right panel). The noise level for data was chosen as $\sigma = 0.1$.*

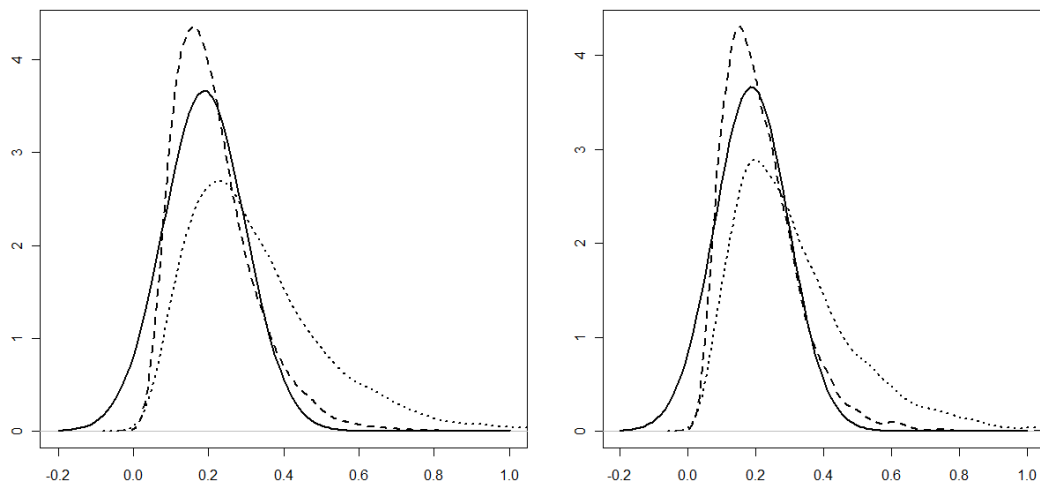


Figure 2: *Empirical densities of the test statistics T_n (on the left) and \tilde{T}_n (on the right) under the null hypothesis H_0 from samples of size 201 (dotted line) and 1001 (dashed line), in comparison with the asymptotic densities (solid line) for noise level $\sigma = 0.1$.*

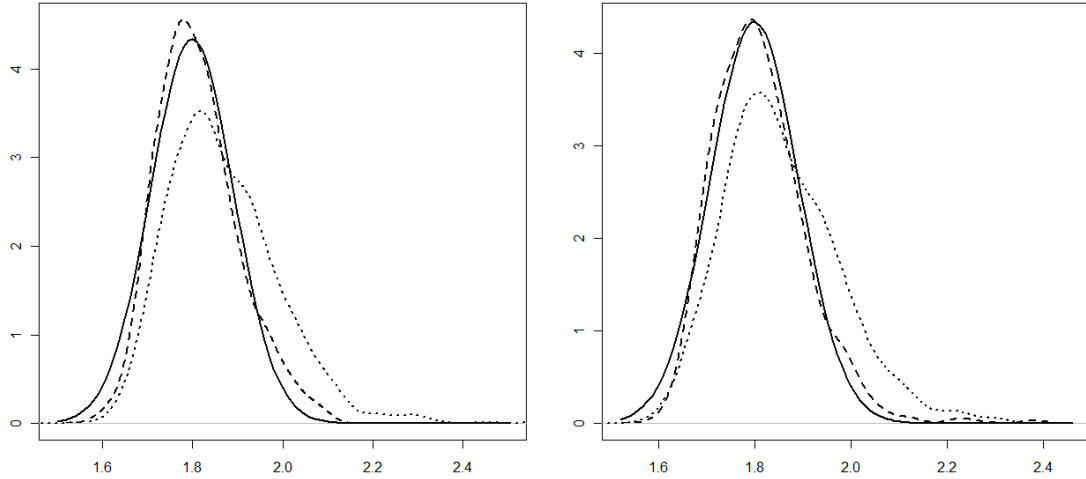


Figure 3: Empirical densities of the test statistics T_n (on the left) and \tilde{T}_n (on the right) under an alternative H_1 from samples of size 201 (dotted line) and 1001 (dashed line), in comparison with the asymptotic densities (solid line) for noise level $\sigma = 0.1$. For the test, we used the null hypothesis $H_0 : f = f_1$, but generated the data from f_2 .

in comparison to their asymptotic densities. In direct regression problems it is well known that the asymptotic distribution of L_2 -type statistics does not yield an accurate approximation of the nominal level (see Härdle and Mammen (1993) or Fan and Linton (2003)) and in the case of inverse regression this difference is even more visible. We observe substantial differences between the asymptotic distribution and the distribution for finite sample sizes. As a consequence the tests using quantiles from the asymptotic distribution do not yield an adequate approximation of the nominal level and the corresponding results are not displayed for the sake of brevity.

In Figure 3 we compare the test statistics with their asymptotic counterparts under an alternative H_1 . Here the empirical densities are closer to the asymptotic densities. Hence the asymptotic distribution under fixed alternatives can be used to estimate the type II error of the test if the null hypothesis can not be rejected (see Dette (1999)).

Because the approximation of the nominal level by the asymptotic distribution is rather poor we propose to use a bootstrap procedure for the determination of the quantiles. In the remaining part of this section we will investigate the finite sample properties of the tests $\varphi(T_n)$ and $\varphi(\tilde{T}_n)$ defined in (13) and (15), respectively. To be precise let

$$\hat{g}(x_i) = \frac{\sum_{j=-n}^n \tilde{K}\left(\frac{x_i - x_j}{h}\right) Y_j}{\sum_{k=-n}^n \tilde{K}\left(\frac{x_i - x_k}{h}\right)}$$

denote the Nadaraya-Watson estimate of the function $g = K_\psi f$, where $\tilde{K}(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$

Table 1: *Simulated level (in %) of the bootstrap version of the tests $\varphi(T_n)$ and $\varphi(\tilde{T}_n)$ defined in (13) and (15), respectively. Data was generated from $f = f_1$ and $f = f_2$ with sample sizes $n = 201$ and 1001, and with noise levels $\sigma = 0.1$ and $\sigma = 0.25$.*

n	σ	Level:	$\varphi(T_n)$			$\varphi(\tilde{T}_n)$		
			5%	10%	20%	5%	10%	20%
100	0.1	$f = f_1$	2.3	5.2	12.4	2.0	4.7	11.0
		$f = f_2$	1.3	2.9	7.2	1.3	3.5	9
100	0.25	$f = f_1$	5.6	9.5	19.2	4.1	8.3	17.4
		$f = f_2$	3.0	6.2	13.8	3.3	6.9	15.1
500	0.1	$f = f_1$	2.5	5.0	9.5	2.5	6.3	12.9
		$f = f_2$	1.3	3.2	8.0	1.9	4.5	10.7
500	0.25	$f = f_1$	2.4	6.6	14	3.9	9.3	17.8
		$f = f_2$	2.9	6.5	13.9	3.9	9.0	17.2

denotes the Gaussian kernel. In the next step the bootstrap residuals

$$\tilde{\varepsilon}_i = Y_i - \hat{g}(x_i)$$

are calculated, where - for the sake of a reliable comparison - the local optimal bandwidth is used in the Nadaraya Watson estimate. The bootstrap sample of residuals $\varepsilon_{-n}^*, \dots, \varepsilon_n^*$ is now drawn with replacement from the set $\{\tilde{\varepsilon}_{-n}, \dots, \tilde{\varepsilon}_n\}$, which gives the bootstrap observations

$$Y_i^* = \hat{g}(x_i) + \varepsilon_i^*.$$

To determine the power of the test we performed the following steps for 200 datasets generated according to model 1. First, for each of the datasets we used 300 bootstrap replications to determine the critical values of the test which were used in the second step to estimate the power of the test in another 300 simulations. Hence, in total we performed 60000 simulations for each combination of parameters n , σ^2 and $f_{i..}$.

In Table 1 we investigate the performance of the tests under the null hypothesis H_0 for data generated from f_1 and f_2 . We observe that both tests are conservative for small variances. Overall the bootstrap test based on \tilde{T}_n yields a better approximation of the nominal level for a larger standard deviation, i.e. $\sigma = 0.25$.

In the second part of the simulation study we study the power of the bootstrap test and investigate the hypotheses $H_0 : f = f_1$ or $H_0 : f = f_2$. The data has been generated from the

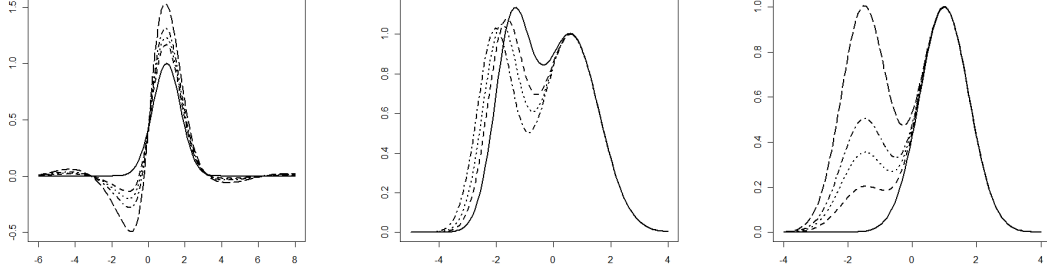


Figure 4: *Left: f_1 (solid line) and **scale disturbed** versions $f_{1,a}$ for $a = 0.6$ (dashed line), $a = 0.85$ (dotted line), $a = 1.1$ (dash dotted line) and $a = 1.35$ (long dashed line). Middle: f_2 (solid line) and **location disturbed** versions $f_{2,b}$ for $b = 0.3$ (dashed line), $b = 0.45$ (dotted line), $b = 0.95$ (dash dotted line). Right: f_1 (solid line) and **modality disturbed** versions $f_{1,c}$ for $c = 0.3$, (dashed line), $c = 0.45$ (dotted line), $c = 0.6$ and $c = 1.0$.*

following three types of alternatives $f_{1,a}$, $f_{2,b}$ and $f_{1,c}$:

Scale disturbance:

$$f_{1,a}(x) = f_1(x) + a \cdot \frac{\sin(x)}{x^2 + 1}$$

where parameter values are chosen as $a \in \{0.6, 0.85, 1.1\}$ for $\sigma = 0.1$ and $a \in \{1.1, 1.35\}$ for $\sigma = 0.25$.

Location disturbance:

$$f_{2,b}(x) = \exp(-0.5 \cdot (x - 0.6)^2) + 1.2 \exp(-1.3 \cdot (x + 1.45 - b)^2)$$

where the parameter values chosen as $b \in \{0.3, 0.45\}$ for $\sigma = 0.1$ and $b \in \{0.65, 0.95\}$ for $\sigma = 0.25$.

Modality disturbance:

$$f_{1,c}(x) = f_1(x) + c \cdot \exp(-(x + 1.5)^2)$$

where parameter values are chosen as $c \in \{0.3, 0.45, 0.6\}$ for $\sigma = 0.1$ and $c \in \{0.6, 1.0\}$ for $\sigma = 0.25$.

Figure 4 shows the different alternative models and in Tables 2-4 we present the power of the bootstrap tests based on T_n and \tilde{T}_n for the different alternatives. In all cases the alternatives are detected with reasonable probabilities. It is worthwhile to mention that for *scale disturbances* the bootstrap test based on T_n yields slightly more power than the test based on \tilde{T}_n . On the

Table 2: *Simulated power (in %) of the bootstrap version of the tests $\varphi(T_n)$ and $\varphi(\tilde{T}_n)$ for the hypothesis $H_0 : f = f_1$ in the case of **scale disturbances**. The sample sizes are given by 201 and 1001 and noise levels are $\sigma = 0.1$ and $\sigma = 0.25$. Data is generated from $f_{1,a}$, but the null hypothesis $f_0 = f_1$.*

n	σ	Level:	Power of $\varphi(T_n)$			Power of $\varphi(\tilde{T}_n)$		
			5%	10%	20%	5%	10%	20%
100	0.1	$a = 0.6$	8.0	17.6	46.4	7.6	16.6	42.3
		$a = 0.85$	18.4	39.6	72.1	21.0	39.7	74.4
		$a = 1.1$	72.2	91.6	99.7	75.5	96.2	99.4
100	0.25	$a = 1.1$	4.6	10.7	24.5	4.1	14.7	26.3
		$a = 1.35$	9.4	19	37.1	6.0	17.1	26.3
500	0.1	$a = 0.6$	25.5	49.6	76.6	22.7	46.5	75.8
		$a = 0.85$	64.1	94.5	100	63.5	88.5	100
		$a = 1.1$	91.0	98.2	100	100	100	100
500	0.25	$a = 1.1$	7.1	17.9	37.5	5.0	16.2	34.2
		$a = 1.35$	12.4	21.3	48.0	11.8	21.6	50.2

other hand for *location disturbances* and *modality disturbances* the test based on \tilde{T}_n shows a substantially better performance. Moreover, *location disturbances* appear to be more simple to detect than *scale disturbances*. A possible explanation of this observation is that a *scale disturbed* function $f_{1,a}$ to a significant extend resembles a typical estimate for the function f for a undersmoothing bandwidth h .

Acknowledgements. The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation and the BMBF project INVERS.

Table 3: Simulated power (in %) of bootstrap version of the tests $\varphi(T_n)$ and $\varphi(\tilde{T}_n)$ for the hypothesis $H_=: f = f_2$ in the case of **location disturbances**. The sample sizes are given by 201 and 1001 and noise levels are $\sigma = 0.1$ and $\sigma = 0.25$. Data is generated from $f_{2,b}$.

n	σ	Level:	Power of $\varphi(T_n)$			Power of $\varphi(\tilde{T}_n)$		
			5%	10%	20%	5%	10%	20%
100	0.1	$b = 0.3$	7.4	15.5	35.5	11.9	23.6	45.5
		$b = 0.45$	22.2	40.1	67.4	25.0	50.8	80.0
100	0.25	$b = 0.65$	5.7	13.7	29	10.6	20.0	34.7
		$b = 0.95$	13.7	26.4	46.2	12.4	30.3	52.4
500	0.1	$b = 0.3$	12.0	33.3	64.4	21.4	44.6	73.0
		$b = 0.45$	81.6	95.7	99.8	90.1	98.0	100
500	0.25	$b = 0.65$	10.5	17.9	38.9	15.2	27.4	46.0
		$b = 0.95$	36.3	60.6	83.3	43.6	67.6	89.1

Table 4: Simulated power (in %) of the bootstrap version of the tests $\varphi(T_n)$ and $\varphi(\tilde{T}_n)$ for the hypothesis $H_0 : f = f_1$ in the case of **modality disturbances**. The sample sizes are given by 201 and 1001 and noise levels are $\sigma = 0.1$ and $\sigma = 0.25$. Data is generated from $f_{1,c}$.

n	σ	Level:	Power of $\varphi(T_n)$			Power of $\varphi(\tilde{T}_n)$		
			5%	10%	20%	5%	10%	20%
100	0.1	$c = 0.3$	5.2	9.2	22.5	5.2	11.0	25.3
		$c = 0.45$	12.8	19.1	41.5	10.3	25.8	46.7
		$c = 0.6$	34.2	59.1	84.7	35.5	63.1	90.9
100	0.25	$c = 0.6$	5.9	13.8	27.3	7.1	12.5	25.0
		$c = 1.0$	15.1	32.4	58.1	15.5	28.7	59.4
500	0.1	$c = 0.3$	5.0	11.8	27.9	7.5	16.1	37.5
		$c = 0.45$	29.2	55.6	88.4	33.8	57.1	85.5
		$c = 0.6$	92.9	99.8	100	97.8	99.9	100
500	0.25	$c = 0.6$	8.4	12.6	31.3	12.4	22.1	36.7
		$c = 1.0$	41.2	69.3	93	57.5	81.0	97.0

6 Proofs

Note that $\{Y_{i,n} \mid i = -n, \dots, n; n \in \mathbb{N}\}$ is a triangular array of random variables. Throughout this proof we omit the second index n for the sake of a transparent notation, i.e. we write Y_i instead of $Y_{i,n}$ and similarly x_i, ε_i instead of $x_{i,n}, \varepsilon_{i,n}$, respectively.

6.1 Proof of Theorem 1

The proof is based on the following decomposition of the difference $T_n - ET_n$ between observed and expected test statistic under $H_0 : f = f_0$:

$$T_n - ET_n = A_n + 2B_n,$$

where

$$\begin{aligned} A_n &:= \int_{\mathbb{R}} |\hat{f}_n(t) - E\hat{f}_n(t)|^2 dt - E \int_{\mathbb{R}} |\hat{f}_n(t) - E\hat{f}_n(t)|^2 dt \\ B_n &:= \int_{\mathbb{R}} (\hat{f}_n(t) - E\hat{f}_n(t))(E\hat{f}_n(t) - f(t)) dt. \end{aligned}$$

Next, Parseval's equation with the notation $c_n = \frac{1}{2\pi n a_n}$ yields

$$\begin{aligned} A_n &= 2\pi \int_{\mathbb{R}} c_n^2 \frac{1}{|\mathcal{F}\psi(s)|^2} \sum_{-n \leq j \neq k \leq n} \varepsilon_j \varepsilon_k \exp(-isx_j) \exp(isx_k) I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds \\ &+ 2\pi \int_{\mathbb{R}} c_n^2 \frac{1}{|\mathcal{F}\psi(s)|^2} \sum_{j=-n}^n (\varepsilon_j^2 - \sigma^2) I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds \\ &= \sum_{-n \leq j \neq k \leq n} \varepsilon_j \varepsilon_k a_{jk} + \sum_{j=-n}^n (\varepsilon_j^2 - \sigma^2) a_{jj} =: A_{n,1} + A_{n,2}, \end{aligned}$$

where the last line defines the statistics $A_{n,1}$ and $A_{n,2}$ in the obvious manner and

$$a_{jk} := 2\pi \int_{\mathbb{R}} c_n^2 \frac{1}{|\mathcal{F}\psi(s)|^2} \exp(-isx_j) \exp(isx_k) I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds.$$

Hence, the $T_n - ET_n$ can be decomposed into a sum of three terms as

$$T_n - ET_n = A_{n,1} + A_{n,2} + 2B_n.$$

The proof of Theorem 1 is now performed in several steps. First, we show that under the null hypothesis $H_0 : f = f_0$ the dominating term of A_n is the random variable $A_{n,1}$ which is of order $(na_n^{3/2} h^{2\beta+1/2})^{-1}$. This is as consequence of Lemmas 6.1, 6.2 and 6.4, and the fact that

$$\begin{aligned} A_{n,2} &= \sum_{j=-n}^n (\varepsilon_j^2 - \sigma^2) a_{jj} \sim \frac{C_1}{(2\beta+1)\pi n^2 a_n^2 h^{2\beta+1}} \sum_{j=-n}^n (\varepsilon_j^2 - \sigma^2) \\ &= \frac{C_1}{(2\beta+1)\pi a_n^2 n^{\frac{3}{2}} h^{2\beta+1}} \frac{1}{\sqrt{n}} \sum_{j=-n}^n (\varepsilon_j^2 - \sigma^2), \end{aligned}$$

which implies (by the central limit theorem and assumption **C**)

$$A_{n,2} = o_P \left(\frac{1}{na_n^{\frac{3}{2}} h^{2\beta + \frac{1}{2}}} \right).$$

Secondly, we show that the term B_n is asymptotically negligible, i.e.

$$B_n = o_P \left(\frac{1}{na_n^{\frac{3}{2}}} h^{-2\beta - \frac{1}{2}} \right), \quad (18)$$

see Lemma 6.3. Next, we use Theorem 5.2 of de Jong (1987) to establish asymptotic normality of $A_{n,1}$ (see Lemma 6.4), which yields

$$na_n^{\frac{3}{2}} h^{2\beta + \frac{1}{2}} (T_n - ET_n) = na_n^{\frac{3}{2}} h^{2\beta + \frac{1}{2}} A_{n,1} + o_P(1) + o_P(1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_2).$$

With the following auxiliary results we present a rigorous justification of these arguments. Finally, the bias will be computed as

$$E[T_n] \sim \frac{C_1}{\pi(2\beta + 1)} \frac{\sigma^2}{na_n^2} \left(\frac{1}{h} \right)^{2\beta + 1}.$$

This completes the proof of Theorem 1.

Lemma 6.1. *Suppose that $na_n h \rightarrow \infty$ for $n \rightarrow \infty$, then*

$$\text{Var}(A_{n,1}) \sim \frac{C_2}{n^2 a_n^3} \cdot \left(\frac{1}{h} \right)^{4\beta + 1} \text{ for a constant } C_2 \in \mathbb{R}^+.$$

Proof: A straightforward calculation yields

$$\begin{aligned} \text{Var}(A_{n,1}) &= \sum_{-n \leq k < j \leq n} 4 \text{Var}(\varepsilon_k \varepsilon_j) |a_{kj}|^2 = \sum_{-n \leq k < j \leq n} 4\sigma^4 |a_{kj}|^2 = \sum_{-n \leq k \neq j \leq n} 2\sigma^4 |a_{kj}|^2 \\ &= \sum_{-n \leq k, j \leq n} 2\sigma^4 |a_{kj}|^2 - \sum_{k=-n}^n 2\sigma^4 |a_{kk}|^2 = \sum_{-n \leq k, j \leq n} 2\sigma^4 |a_{kj}|^2 + O \left(\frac{1}{n^3 a_n^4} \left(\frac{1}{h} \right)^{4\beta + 2} \right). \end{aligned}$$

$$\begin{aligned} \sum_{j, k=-n}^n |a_{jk}|^2 &\sim 4C_1^2 \pi^2 c_n^4 \sum_{j=-n}^n \sum_{k=-n}^n \int_{\mathbb{R}} \int_{\mathbb{R}} |s|^{2\beta} |z|^{2\beta} I_{[-\frac{1}{h}, \frac{1}{h}]}(s) I_{[-\frac{1}{h}, \frac{1}{h}]}(z) \cdot \\ &\quad \cdot \exp(ix_k(s-z)) \cdot \exp(-ix_j(s-z)) ds dz \\ &= 4C_1^2 c_n^2 \int_{\mathbb{R}} \int_{\mathbb{R}} |s|^{2\beta} |z|^{2\beta} I_{[-\frac{1}{h}, \frac{1}{h}]}(s) I_{[-\frac{1}{h}, \frac{1}{h}]}(z) \cdot \frac{\sin^2 \left(\frac{s-z}{a_n} \right)}{(s-z)^2} ds dz \\ &= \frac{4C_1^2 c_n^2}{h^{4\beta}} \int_{-1}^1 \int_{-1}^1 |sz|^{2\beta} \cdot \frac{\sin^2 \left(\frac{s-z}{a_n h} \right)}{(s-z)^2} ds dz \\ &= \frac{8C_1^2 c_n^2}{h^{4\beta}} \left(\int_0^1 \int_0^1 |sz|^{2\beta} \cdot \frac{\sin^2 \left(\frac{s-z}{a_n h} \right)}{(s-z)^2} ds dz + \int_0^1 \int_0^1 |sz|^{2\beta} \cdot \frac{\sin^2 \left(\frac{s+z}{a_n h} \right)}{(s+z)^2} ds dz \right) \\ &= \frac{C_1^2}{4^{2\beta} \pi^2 n^2 a_n^2 h^{4\beta}} \left(\int_0^1 \int_{-v}^v |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2 \left(\frac{u}{a_n h} \right)}{u^2} du dv \right. \\ &\quad + \int_1^2 \int_{v-2}^{2-v} |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2 \left(\frac{u}{a_n h} \right)}{u^2} du dv \\ &\quad \left. + \int_{-1}^1 \int_{|u|}^{2-|u|} |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2 \left(\frac{v}{a_n h} \right)}{v^2} dv du \right), \quad (19) \end{aligned}$$

where we substituted $s + z = v$ and $s - z = u$. Next, we show

$$\lim_{n \rightarrow \infty} a_n h \int_0^1 \int_0^1 |sz|^{2\beta} \cdot \frac{\sin^2(\frac{s-z}{a_n h})}{(s-z)^2} ds dz = \tilde{C} \in \mathbb{R}^+$$

for the first of these integrals. For the second and third integral a similar limiting property can be established.

For this purpose define $\tau_n := \lfloor v/(\pi a_n h) \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of $v/(\pi a_n h)$. We find

$$\begin{aligned} & a_n h \int_0^1 \int_{-v}^v |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2(\frac{u}{a_n h})}{u^2} du dv \\ &= 2a_n h \int_0^1 \sum_{j=0}^{\tau_n} \int_{j\pi a_n h}^{(j+1)\pi a_n h} |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2(\frac{u}{a_n h})}{u^2} du dv - \Delta_n^1, \end{aligned}$$

where Δ_n^1 is a non-negative error term. From this decomposition of the inner integral it follows immediately that whenever

$$u^2 = j^2 \pi^2 a_n^2 h^2 + \Delta_{j,n,u}^2 \in [j^2 \pi^2 a_n^2 h^2, j^2 \pi^2 a_n^2 h^2 + (2j+1)\pi^2 a_n^2 h^2],$$

we have for the non-negative error term $\Delta_{j,n,u}^2$

$$0 \leq \Delta_{j,n,u}^2 \leq (2j+1)\pi^2 a_n^2 h^2 \leq \left(\frac{2v}{\pi a_n h} + 1 \right) \pi^2 a_n^2 h^2 \leq 2\pi a_n h (1 + \pi a_n h),$$

or

$$\Delta_{j,n,u}^2 = O(a_n h)$$

uniformly in u and j . Using this approximation for u^2 in the expression $|v^2 - u^2|$ of the integrand we find

$$\begin{aligned} & a_n h \int_0^1 \int_{-v}^v |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2(\frac{u}{a_n h})}{u^2} du dv \\ & \sim 2 \int_0^1 \sum_{j=0}^{\tau_n} \int_{j\pi}^{(j+1)\pi} |v^2 - (j\pi a_n h)^2|^{2\beta} \cdot \frac{\sin^2(\tilde{u})}{\tilde{u}^2} d\tilde{u} dv - \Delta_n^1 \\ & = 2 \int_0^1 \sum_{j=0}^{\tau_n} c_j |v^2 - (j\pi a_n h)^2|^{2\beta} dv - \Delta_n^1, \end{aligned}$$

where the sequence $\{c_j\}$ is summable, since

$$\sum_{j=0}^{\infty} c_j = \int_0^{\infty} \frac{\sin^2(\tilde{u})}{\tilde{u}^2} d\tilde{u} = \frac{\pi}{2}.$$

It remains to show the existence of the outer integral. Let

$$A_n(v) := \sum_{j=0}^{\tau_n} c_j |v^2 - (j\pi a_n h)^2|^{2\beta} \cdot I_{[0,1]}(v).$$

Since $v \in [0, 1]$ and $j \in \{0, \dots, \tau_n\}$ we have $|v^2 - (j\pi a_n h)^2|^{2\beta} \leq 1$. This implies the convergence of A_n to a (Lebesgue-)measurable function A and

$$0 \leq A_n \leq \frac{\pi}{2} \cdot I_{[0,1]} =: g.$$

An application of Lebesgue's Theorem of dominated convergence gives

$$\lim_{n \rightarrow \infty} \int A_n(v) dv = \int \lim_{n \rightarrow \infty} A_n(v) dv = \int A(v) dv \leq \frac{\pi}{2},$$

which shows that the limit exists.

Moreover, a straightforward calculation yields

$$\Delta_n^1 = 2a_n h \int_0^1 \int_v^{\lfloor \frac{v}{\pi a_n h} \rfloor + 1} \pi a_n h |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2(\frac{u}{a_n h})}{u^2} du dv = O\left((a_n h)^{2\beta}\right) = o(1),$$

and thus

$$0 \leq l := \lim_{n \rightarrow \infty} a_n h \int_0^1 \int_0^1 |sz|^{2\beta} \cdot \frac{\sin^2(\frac{s-z}{a_n h})}{(s-z)^2} ds dz = \int A(v) dv.$$

In the final step, we show that the limit l is positive. To this end let $n_0 \in \mathbb{N}$ such that $\frac{1}{a_n h} \geq \frac{\pi}{\sqrt{2}}$ for all $n \geq n_0$, then it follows

$$\begin{aligned} a_n h \int_0^1 \int_{-v}^v |v^2 - u^2|^{2\beta} \cdot \frac{\sin^2(\frac{u}{a_n h})}{u^2} du dv &= \int_0^1 \int_{-\frac{v}{a_n h}}^{\frac{v}{a_n h}} |v^2 - (ua_n h)^2|^{2\beta} \cdot \frac{\sin^2(u)}{u^2} du dv \\ &\geq \int_{\frac{1}{2}}^1 \int_{-\frac{\pi}{4}}^{\frac{\pi}{4}} |v^2 - (ua_n h)^2|^{2\beta} \cdot \frac{\sin^2(u)}{u^2} du dv > 0. \end{aligned}$$

Together with a similar computation for the second and third integral in decomposition (19) this concludes the proof of the Lemma. \square

For a proof of (18) we start by showing that $\hat{f}_n(t)$ is asymptotically unbiased uniformly in t .

Lemma 6.2. *For $\mathcal{F}f \in L_2(\mathbb{R})$ and under assumptions **A**, **B** and **C** we have*

$$E[\hat{f}_n(t)] - f(t) = o(h^\alpha) + O\left(\frac{1}{na_n h^{\beta+1}}\right) + o\left(a_n^\gamma \cdot \frac{1}{h^{\beta+1}}\right) o(1),$$

uniformly with respect to t .

Proof: Note that

$$E[\hat{f}_n(t)] - f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(ist) \left[\frac{\frac{1}{na_n} \sum_{j=-n}^n Kf(x_j) \exp(-isx_j)}{\mathcal{F}\psi(s)} \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) - \mathcal{F}f(s) \right] ds.$$

Using the estimate

$$\frac{1}{na_n} \sum_{j=-n}^n Kf(x_j) \exp(-isx_j) = \int_{-\frac{1}{a_n}}^{\frac{1}{a_n}} Kf(z) \exp(-isz) dz + O\left(\frac{1}{na_n}\right),$$

(uniformly with respect to s) we obtain

$$\begin{aligned}
E[\hat{f}_n(t)] - f(t) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(ist) \left[\frac{\int_{-\frac{1}{a_n}}^{\frac{1}{a_n}} Kf(z) \exp(-isz) dz + O\left(\frac{1}{na_n}\right)}{\mathcal{F}\psi(s)} \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) - \mathcal{F}f(s) \right] ds \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(ist) \left[\frac{\mathcal{F}(Kf)(s)}{\mathcal{F}\psi(s)} \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) - \mathcal{F}f(s) \right] ds \\
&\quad - \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\exp(ist)}{\mathcal{F}\psi(s)} \left[\int_{(-\infty, -\frac{1}{a_n}] \cup [\frac{1}{a_n}, \infty)} Kf(z) \exp(-isz) dz + O\left(\frac{1}{na_n}\right) \right] \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds
\end{aligned}$$

From equation (5) we have $\frac{\mathcal{F}(Kf)}{\mathcal{F}\psi} = \mathcal{F}f$, which yields

$$\begin{aligned}
E[\hat{f}_n(t)] - f(t) &= \frac{1}{2\pi} \int_{\mathbb{R}} \exp(ist) \left[\mathcal{F}f(s) \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) - \mathcal{F}f(s) \right] ds + O\left(\frac{1}{na_n h^{\beta+1}}\right) \\
&\quad - \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\exp(ist)}{\mathcal{F}\psi(s)} \left[\int_{(-\infty, -\frac{1}{a_n}] \cup [\frac{1}{a_n}, \infty)} Kf(z) \exp(-isz) dz \right] \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds \\
&= -\frac{1}{2\pi} \int_{(-\infty, -\frac{1}{h}] \cup [\frac{1}{h}, \infty)} \exp(ist) \mathcal{F}f(s) ds + O\left(\frac{1}{na_n h^{\beta+1}}\right) \\
&\quad - \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\exp(ist)}{\mathcal{F}\psi(s)} \left[\int_{(-\infty, -\frac{1}{a_n}] \cup [\frac{1}{a_n}, \infty)} Kf(z) \exp(-isz) dz \right] \cdot I_{[-\frac{1}{h}, \frac{1}{h}]}(s) ds.
\end{aligned}$$

Now we estimate the integrals as follows.

$$\frac{1}{a_n^\gamma} \left| \int_{[\frac{1}{a_n}, \infty)} Kf(z) \exp(-isz) dz \right| \leq \int_{[\frac{1}{a_n}, \infty)} \left| Kf(z) \exp(-isz) \frac{1}{a_n^\gamma} \right| dz \leq \int_{[\frac{1}{a_n}, \infty)} |Kf(z)| \cdot |z|^\gamma dz.$$

From assumption **B** and Lebesgue's theorem of dominated convergence it follows that

$$\lim_{n \rightarrow \infty} \int_{[\frac{1}{a_n}, \infty)} |Kf(z)| \cdot |z|^\gamma dz = 0,$$

and we obtain

$$\int_{(-\infty, \frac{1}{a_n}] \cup [\frac{1}{a_n}, \infty)} |Kf(z)| dz = o(a_n^\gamma).$$

The same line of arguments yields

$$\int_{(-\infty, -\frac{1}{h}] \cup [\frac{1}{h}, \infty)} \exp(ist) \mathcal{F}f(s) ds = o(h^\alpha).$$

Finally, we conclude

$$E\hat{f}_n(t) - f(t) = o(h^\alpha) + O\left(\frac{1}{na_n h^{\beta+1}}\right) + o\left(a_n^\gamma \cdot \frac{1}{h^{\beta+1}}\right),$$

uniformly with respect to t .

□

Lemma 6.3. Under assumptions **A**, **B** and **C**, and for $\mathcal{F}f \in L_2(\mathbb{R})$,

$$B_n = \int_{\mathbb{R}} \left(\hat{f}_n(t) - E\hat{f}_n(t) \right) \left(E\hat{f}_n(t) - f(t) \right) dt = o_P \left(h^{-2\beta - \frac{1}{2}} n^{-1} a_n^{-\frac{3}{2}} \right).$$

Proof: From the Cauchy-Schwarz equality we have

$$|B_n| \leq \left(\int_{\mathbb{R}} \left| \hat{f}_n(t) - E\hat{f}_n(t) \right|^2 dt \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}} \left| E\hat{f}_n(t) - f(t) \right|^2 dt \right)^{\frac{1}{2}},$$

and an application of the Markov-inequality yields

$$\left(\int_{\mathbb{R}} \left| \hat{f}_n(t) - E\hat{f}_n(t) \right|^2 dt \right)^{\frac{1}{2}} = O_P \left(h^{-2\beta - \frac{1}{2}} n^{-1} a_n^{-\frac{3}{2}} \right).$$

Finally, using Lemma 6.2 provides

$$\int_{\mathbb{R}} \left| E\hat{f}_n(t) - f(t) \right|^2 dt = o(1),$$

which completes the proof of lemma 6.3. □

Lemma 6.4. Under the assumptions of Theorem 1 we have

$$na_n^{\frac{3}{2}} h^{2\beta + \frac{1}{2}} A_{n,1} \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_2).$$

Proof: The asymptotic normality is established using Theorem 5.2 in de Jong (1987). Straight-forward calculations yield

$$\frac{1}{h^\beta} \cdot \frac{1}{\text{Var}(A_{n,1})} \cdot \max_{|i| \leq n} \sum_{\substack{|j| \leq n \\ j \neq i}} a_{ij}^2 \rightarrow 0 \quad \text{for } n \rightarrow \infty,$$

which implies assumption (1) of Theorem 5.2 in de Jong (1987) with the choice $K(n) = h^{-\frac{\beta}{4}}$.

Moreover, assumption (2) of the latter Theorem is an immediate consequence of $E|\varepsilon_i| < \infty$ and $\frac{1}{h} \rightarrow \infty$ for $n \rightarrow \infty$. It remains to show assumption (3) of Theorem 5.2 in de Jong (1987). From Gershgorin's circle Theorem (see Horn and Johnson (1985), p. 344-346) we obtain

$$|\mu_k| \leq \sum_{\substack{j=-n \\ j \neq k}}^n |a_{jk}| \quad \forall k = -n, \dots, n,$$

where μ_{-n}, \dots, μ_n denote the eigenvalues of the matrix $(a_{ij})_{i,j=-n}^n$. Further we have for all $i \in \{-n, \dots, n\}$

$$|\mu_i| \leq \max_{|k| \leq n} \sum_{\substack{j=-n \\ j \neq k}}^n |a_{jk}| = O \left(\frac{\ln(n)}{na_n h^{2\beta}} \right),$$

which is uniformly in k . From this estimate we have, under assumption **C**,

$$\frac{1}{\text{Var}(A_{n,1})} \cdot \max_{|i| \leq n} |\mu_i|^2 = O(h(1 + \ln(2n))) = o(1).$$

Hence, assumption (3) of de Jong (1987), Theorem 5.2 is satisfied, too, which establishes asymptotic normality of $A_{n,1}$.

□

6.2 Proof of Theorem 2

For a proof of the asymptotic normality of T_n under the alternative we use *Parseval's equation* to obtain the decomposition

$$\begin{aligned} T_n - ET_n &= \int_{\mathbb{R}} |\hat{f}_n(t) - f_0(t)|^2 dt - E \int_{\mathbb{R}} |\hat{f}_n(t) - f_0(t)|^2 dt \\ &= \sum_{j \neq k = -n}^n (Y_j Y_k - Kf(x_j)Kf(x_k))a_{jk} + \sum_{j=-n}^n (Y_j^2 - Kf^2(x_j) - \sigma^2)a_{jj} \\ &\quad + \int_{-\frac{1}{h}}^{\frac{1}{h}} c_n \sum_{j=-n}^n \varepsilon_j \frac{\exp(-isx_j)}{\mathcal{F}\psi(s)} \overline{\mathcal{F}f_0(s)} ds + \int_{-\frac{1}{h}}^{\frac{1}{h}} c_n \sum_{j=-n}^n \varepsilon_j \frac{\exp(isx_j)}{\mathcal{F}\psi(s)} \mathcal{F}f_0(s) ds \\ &=: D_1 + D_2 + D_3 + \overline{D_3}, \end{aligned}$$

where the last identity defines the random variables D_1 , D_2 and D_3 in an obvious manner.

Straightforward calculations give the representation

$$D_1 = \sum_{-n \leq j \neq k \leq n} \varepsilon_j \varepsilon_k a_{jk} + 2 \sum_{-n \leq j \neq k \leq n} Kf(x_j) \varepsilon_k a_{jk} =: D_{1,1} + 2D_{1,2},$$

which yields

$$T_n - ET_n = D_{1,1} + 2D_{1,2} + D_2 + D_3 + \overline{D_3}. \quad (20)$$

Obviously, $E[D_{1,2}] = 0$, and for the variance we obtain

$$\begin{aligned} \text{Var}(D_{1,2}) &= \sigma^2 \sum_{k=-n}^n \left| \sum_{\substack{j=-n \\ j \neq k}}^n Kf(x_j) a_{jk} \right|^2 = \sigma^2 \sum_{k=-n}^n \left| \sum_{j=-n}^n Kf(x_j) a_{jk} - Kf(x_k) a_{kk} \right|^2 \\ &= \sigma^2 \sum_{k=-n}^n \left| \sum_{j=-n}^n Kf(x_j) a_{jk} \right|^2 + O\left(\frac{1}{n^2 a_n^2 h^{2\beta+1}}\right) + O\left(\frac{1}{n^3 a_n^3 h^{4\beta+2}}\right). \end{aligned}$$

By an integral approximation and the definition of a_{jk} we obtain

$$\begin{aligned}
\sum_{k=-n}^n \left| \sum_{j=-n}^n Kf(x_j) a_{jk} \right|^2 &= 4\pi^2 c_n^4 \sum_{k=-n}^n \left| \sum_{j=-n}^n Kf(x_j) \int_{-\frac{1}{h}}^{\frac{1}{h}} \frac{\exp(-isx_j) \exp(isx_k)}{|\mathcal{F}\psi(s)|^2} ds \right|^2 \\
&= 2\pi c_n^3 \left(\int_{-\frac{1}{an}}^{\frac{1}{an}} \left| \int_{-\frac{1}{h}}^{\frac{1}{h}} \sum_{j=-n}^n Kf(x_j) \frac{\exp(-isx_j) \exp(isz)}{|\mathcal{F}\psi(s)|^2} ds \right|^2 dz + O\left(\frac{1}{na_n}\right) \right) \\
&= 2\pi c_n^3 \int_{\mathbb{R}} \left| \int_{-\frac{1}{h}}^{\frac{1}{h}} \sum_{j=-n}^n Kf(x_j) \frac{\exp(-isx_j) \exp(isz)}{|\mathcal{F}\psi(s)|^2} ds \right|^2 dz + O\left(\frac{1}{n^4 a_n^4}\right) \\
&= 4\pi^2 c_n^3 \int_{-\frac{1}{h}}^{\frac{1}{h}} \left| \sum_{j=-n}^n Kf(x_j) \frac{\exp(-isx_j)}{|\mathcal{F}\psi(s)|^2} \right|^2 ds + O\left(\frac{1}{n^4 a_n^4}\right),
\end{aligned}$$

where we have applied Parseval's equation for the fourth equality. Another integral approximation of the remaining sum yields

$$\begin{aligned}
\text{Var}(D_{1,2}) &= 4\sigma^2 \pi^2 c_n^3 n^2 a_n^2 \int_{-\frac{1}{h}}^{\frac{1}{h}} \left| \int_{-\frac{1}{an}}^{\frac{1}{an}} Kf(z) \frac{\exp(-isz)}{|\mathcal{F}\psi(s)|^2} dz \right|^2 ds + o\left(\frac{1}{na_n}\right) + O\left(\frac{1}{n^2 a_n^2 h^{2\beta+1}}\right) \\
&\quad + O\left(\frac{1}{n^3 a_n^3 h^{4\beta+2}}\right) + O\left(\frac{1}{n^3 a_n^3 h^{4\beta+1}}\right) \\
&= \frac{\sigma^2}{2\pi n a_n} \int_{-\frac{1}{h}}^{\frac{1}{h}} \left| \frac{\mathcal{F}f(s)}{\mathcal{F}\psi(s)} \right|^2 ds + o\left(\frac{1}{na_n}\right) + O\left(\frac{1}{n^2 a_n^2 h^{2\beta+1}}\right) + O\left(\frac{1}{n^3 a_n^3 h^{4\beta+2}}\right) \\
&= \frac{\sigma^2}{2\pi n a_n} \left(\left\| \frac{\mathcal{F}f}{\mathcal{F}\psi} \right\|^2 + o(1) \right),
\end{aligned}$$

where the last estimate follows from assumption **D**, which implies $O(n^{-2} a_n^{-2} h^{-2\beta-1}) = o(n^{-1} a_n^{-1})$ and $O(n^{-3} a_n^{-3} h^{-4\beta-2}) = o(n^{-1} a_n^{-1})$. If assumption **D** is fulfilled, straightforward computations show that the remaining terms in (20) are asymptotically negligible, that is

$$D_{1,1} = o_P\left(\frac{1}{\sqrt{na_n}}\right), \quad D_2 = o_P\left(\frac{1}{\sqrt{na_n}}\right), \quad D_3 = o_P\left(\frac{1}{\sqrt{na_n}}\right), \quad \overline{D}_3 = o_P\left(\frac{1}{\sqrt{na_n}}\right).$$

Finally, the assertion of Theorem 2 follows by an application of *Lyapunov's Theorem* to the leading term $D_{1,2}$, i.e.

$$\sqrt{na_n}(T_n - ET_n) = \sqrt{na_n} \left(2D_{1,2} - \|f - f_0\|^2 + o_P\left(\frac{1}{\sqrt{na_n}}\right) \right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{2\sigma^2}{\pi} \left\| \frac{\mathcal{F}f}{\mathcal{F}\psi} \right\|^2\right).$$

□

References

Bachmann, D. and Dette, H. (2005). A Note on the Bickel-Rosenblatt Test in Autoregressive Time Series. *Stat. Probab. Lett.* **74**, 221–234.

- Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1**, 1071–1095.
- Birke, M., Bissantz, N. and Holzmann, H. (2009). Confidence bands for inverse regression models. Submitted.
- Bissantz, N. and Birke, M. (2009). Asymptotic normality and confidence intervals for inverse regression models with convolution-type operators. *J. Multivariate Anal.*, **100**, 2364–2375.
- Bissantz, N., Hohage, T., Munk, A. and Ruymgaart, F. (2007), Convergence rates of general regularization methods for statistical inverse problems. *SIAM J. Num. Anal.*, **45**, 2610–2636.
- Bissantz, N., Holzmann, H. and Munk, A. (2005). Testing parametric assumptions on band- or time-limited signals under noise. *IEEE Trans. Inf. Theory* **51**, 3796–3805.
- Butucea, C. (2007). Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35**, 1907–1930.
- Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123**, 323–354.
- Claxton, N. S., Fellers, T. J. and Davidson, M. W. (2005). Laser scanning confocal microscopy, Department of Optical Microscopy and Digital Imaging, National High Magnetic Field Laboratory, Florida State University, 37 p., Technical Report (<http://www.olympusfluoview.com/th>
- De Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Th. Rel. Fields* **75**, 261–277.
- 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.
- Dette, H., Munk, A. and Wagner, T. (1998). Estimating the variance in nonparametric regression by quadratic forms - what is a reasonable choice? *J. Roy. Statist. Soc. Ser. B* **60**, 319–346.
- Dette, H., Spreckelsen, I. (2003). A note on a specification test for time series models based on spectral density estimation. *Scand. J. Statist.* **30**, 481–491.

- Diggle, P. J. and Hall P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55**, 523–531.
- Fan, Y., and Linton, O. (2003). Some higher order theory for a consistent nonparametric model specification test. *Journal of Statistical Planning and Inference* **109**, 125-154.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–1947.
- Holzmann, H. (2007). Testing parametric models in the presence of instrumental variables. *Stat. Probab. Lett.* **78**, 629–636.
- Holzmann, H., Bissantz, N. and Munk, A. (2007). Density testing in a contaminated sample. *J. Multivariate Anal.* **98**, 57–75.
- Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.
- Huang, L. S. and Fan J. (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli* **5**, 927–949.
- Lauer, T. R., Faber, S. M., Gebhardt, K., Richstone, D., Tremaine, S., Ajhar, E. A., Aller, M. C., Bender, R., Dressler, A., Filippenko, A. V., Green, R., Grillmair, C. J., Ho, L. C., Kormendy, J., Magorrian, J., Pinkney, J., Siopis, C. (2005). The centers of early-type galaxies with Hubble Space Telescope. V. New WFPC2 Photometry, *Astron. Journ.* **129**, 2138–2185.
- Mair, B. A. and Ruymgaart, F. H. (1996), Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.* **56**, 1424–1444.
- Neumann, M. H. and Paparoditis, E. (2000). On bootstrapping L_2 -type statistics in density testing. *Statist. Probab. Lett.* **50**, 137–147.
- Paparoditis, E. (2000). Spectral density based goodness of fit tests for time series models. *Scand. J. Statist.* **27**, 145–176.
- Spring, K. R. and Inoué, S. (1997). *Video Microscopy: The Fundamentals*, New York: Plenum Press.

