Ann Inst Stat Math (2009) 61:691–714
DOI 10.1007/s10463-007-0164-y

Optimal testing for additivity in multiple nonparametric regression

Felix Abramovich · Italia De Feis · Theofanis Sapatinas

Received: 19 February 2007 / Revised: 25 September 2007 / Published online: 12 January 2008

Abstract We consider the problem of testing for additivity in the standard multiple nonparametric regression model. We derive optimal (in the minimax sense) non-adaptive and adaptive hypothesis testing procedures for additivity against the composite nonparametric alternative that the response function involves interactions of second or higher orders separated away from zero in $L^2([0, 1]^d)$-norm and also possesses some smoothness properties. In order to shed some light on the theoretical results obtained, we carry out a wide simulation study to examine the finite sample performance of the proposed hypothesis testing procedures and compare them with a series of other tests for additivity available in the literature.

Keywords Additive models · Functional hypothesis testing · Minimax testing · Nonparametric regression · Wavelets

F. Abramovich (✉)
Department of Statistics and Operations Research,
Tel Aviv University, Tel Aviv 69978, Israel
e-mail: felix@post.tau.ac.il

I. De Feis
Istituto per le Applicazioni del Calcolo "Mauro Picone",
Sezione di Napoli, Consiglio Nazionale delle Ricerche,
Via Pietro Castellino 111, 80131 Napoli, Italy

T. Sapatinas
Department of Mathematics and Statistics, University of Cyprus,
P.O. Box 20537, 1678 Nicosia, Cyprus

1 Introduction

Consider the standard multiple nonparametric regression model with a noise level σ, i.e.,

$$y_i = f(\mathbf{x}_i) + \sigma\,\varepsilon_i, \quad \mathbf{x}_i = (x_{1i}, \ldots, x_{di})' \in [0, 1]^d, \qquad i = 1, \ldots, n, \quad d \geq 2, \qquad (1)$$

where ε_i are independent $N(0, 1)$ random variables. It is well-known that for a general response function f, the direct use of various d-dimensional nonparametric estimators (e.g., splines, kernel methods, local polynomial regression, orthogonal series expansions, etc.) faces a serious "curse of dimensionality" problem, even for moderate d. Several multiple nonparametric regression techniques have been devised in response to this problem, implementing various assumptions on the structure of f.

Probably the most popular existing approach to structural multiple nonparametric regression is the use of *additive models*, where the d-dimensional response function $f(x_1, \ldots, x_d)$ is assumed to be (or, at least, well approximated by) a sum of d univariate functions $f_l(x_l),\ l = 1, \ldots, d$. Additive models can be viewed as a natural *nonparametric* generalization of the main-effects models in standard linear regression. They are efficiently tractable and allow for a simple interpretation. In particular, Stone (1985) showed that in additive models the response function can be estimated with the same rate of estimation error as in the univariate case. Additive models were popularized by Hastie and Tibshirani (1990) and nowadays there is a plethora of research work on fitting and estimating their components. See Buja et al. (1989), Hastie and Tibshirani (1990), Linton and Nielsen (1995), Linton (1997), Opsomer and Ruppert (1997, 1998), Sperlich et al. (1999), Amato and Antoniadis (2001), Amato et al. (2002), Sperlich et al. (2002), Zhang and Wong (2003), Sardy and Tseng (2004) and Fan and Jiang (2005).

On the other hand, much less attention has been paid to checking the adequacy of additivity assumption of the underlying response function from the data. Ignoring such a model check for additivity might lead evidently to misinterpretation of the data and erroneous inference. Although early work dates back to Tukey (1949), it is only recently that the problem of testing for additivity has been of real interest. We refer to Barry (1993), Eubank et al. (1995), Gozalo and Linton (2001), Dette and Derbort (2001), Dette and Wilkau (2001), Derbort et al. (2002), Sperlich et al. (2002), Li et al. (2003), De Canditiis and Sapatinas (2004) and Dette et al. (2005) for various approaches to testing for additivity in the nonparametric regression models. Despite a growing number of works, almost none of them investigated the optimality of the proposed additivity tests. Gozalo and Linton (2001) and Li et al. (2003) investigated the asymptotic power of their tests against a sequence of local alternatives. However, such a approach in nonparametric testing is known to be generally restricted to a too small class of alternative models (see, e.g., Horowitz and Spokoiny, 2001). Härdle et al. (2001) suggested testing additivity by performing a series of tests for no second order interactions for each pair $(x_l, x_l'),\ l \neq l'$. They demonstrated asymptotic optimality (in the minimax sense) of each *individual* test against a smooth nonparametric alternative but the unavoidable multiplicity effect for the *overall* testing remained unclear. Moreover, possible presence of higher order interactions was ignored.

In this paper, we derive asymptotically (as the sample size grows) optimal (in the minimax sense) *non-adaptive* and *adaptive* general tests for additivity for a wide class of alternatives. Borrowing ideas from the theory of analysis of variance, the underlying response function f in model (1), assuming $f \in L^2([0,1]^d)$, admits the following unique decomposition

$$f(\mathbf{x}) = \mu + \sum_{l=1}^{d} f_l(x_l) + f_0(\mathbf{x}), \quad \mathbf{x} \in [0,1]^d, \tag{2}$$

where μ is a constant (the *grand mean*), f_l are 1-dimensional functions of x_l (the *main effects*) and f_0 involves all interactions of second and higher orders (see, e.g., Antoniadis 1984). The components of the decomposition (2) satisfy the following identifiability conditions

$$\int_{[0,1]} f_l(x_l)\, \mathrm{d}x_l = 0, \quad l = 1, \ldots, d, \tag{3}$$

$$\int_{[0,1]^d} f_0(\mathbf{x})\, \mathrm{d}\mathbf{x} = 0, \tag{4}$$

$$\int_{[0,1]^{d-1}} f_0(\mathbf{x})\, \mathrm{d}\mathbf{x}_{(-l)} = 0, \quad l = 1, \ldots, d, \tag{5}$$

where $\mathbf{x}_{(-l)} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d)$. Testing for additivity in the model (1)–(2), subject to identifiability conditions (3)–(5), is equivalent to testing for no interactions of second or higher orders which is, in turn, equivalent to testing the null hypothesis

$$H_0 : f_0 \equiv 0. \tag{6}$$

Certain regularity assumptions on f_0 are required to be able to distinguish it from noise. In particular, we assume that f_0 belongs to a Besov ball $B^s_{p,q}(M)$ of radius $M > 0$ on $[0,1]^d$, where $1 \leq p, q \leq \infty$ and $s > d/p$. Besov classes are known to have exceptional expressive power: for particular choices of the parameters s, p and q, they include the traditional Hölder ($p = q = \infty$) and Sobolev ($p = q = 2$) classes of smooth functions, and the class of inhomogeneous functions of bounded variation sandwiched between $B^1_{1,\infty}$ and $B^1_{1,1}$. The parameter p can be viewed as a degree of function's inhomogeneity while s is a measure of its smoothness. Roughly speaking, the (not necessarily integer) parameter s indicates the number of function's (fractional) derivatives, where their existence is required in an L^p-sense, while the additional parameter q provides a further finer gradation. We refer to Meyer (1992, Chaps. 2 and 6) for rigorous definitions and a detailed study of Besov spaces.

In addition, to distinguish between the two hypotheses, the set of alternatives should be also separated away from zero by some distance ρ_n tending to zero as n tends to infinity. In this paper, we measure this distance by the $L^2([0,1]^d)$-norm of f_0, $||f_0||_2$, and the resulting nonparametric alternative hypothesis is then of the form

$$H_1 : f_0 \in \mathcal{F}(\rho_n) = \{f_0 : f_0 \in B^s_{p,q}(M),\ ||f_0||_2 \geq \rho_n\}. \tag{7}$$

As the sample size increases, alternatives closer and closer to zero can be detected without loosing accuracy and for the prescribed error probabilities of Type I (erroneous rejection of H_0) and Type II (erroneous acceptance of H_0), the rate of decay of ρ_n can be viewed as a natural measure of goodness of a test (see, e.g., Ingster 1982, 1993). The goal then is to find the fastest rate for which such testing is still possible and to construct the corresponding rate-optimal test.

To develop the testing procedures for additivity in (1), we use the minimax approach for hypothesis testing of Ingster (1982, 1993). In particular, we adapt the results for the minimax nonparametric hypothesis testing for the presence of a signal in the 1-dimensional Gaussian white noise model originated by Ingster (1982) and further developed in, e.g., Ingster (1993), Spokoiny (1996), Lepski and Spokoiny (1999), Guerre and Lavergne (2002), Ingster and Suslina (2005) for various separation distances between the two hypotheses and different smoothness function classes under the alternative. (We refer to Ingster and Suslina, 2003 for a comprehensive account on minimax testing of nonparametric hypotheses in Gaussian models.) More precisely, we derive asymptotically minimax non-adaptive and adaptive hypothesis testing procedures for additivity in the nonparametric regression model (1) against the composite nonparametric alternative hypothesis (7). The proposed tests are similar in spirit to those of Spokoiny (1996) for testing the presence of a signal in the 1-dimensional Gaussian white noise model, and are based on the empirical wavelet coefficients of the data. Wavelet decompositions allow one to characterize different types of smoothness conditions assumed on the response function by means of its wavelet coefficients for the whole range of Besov classes (see, e.g., Meyer 1992, Chap. 6).

The paper is organized as follows. In Sect. 2, we provide a brief background on minimax testing and multivariate wavelet analysis necessary for the proposed methodology. The minimax non-adaptive and adaptive hypothesis testing procedures for additivity in nonparametric regression are derived in Sect. 3. In Sect. 4, we carry out a wide simulation study to examine the finite sample performance of the proposed hypothesis testing procedures for testing additivity in the 2-dimensional (bivariate) nonparametric regression setting on the unit square, and make comparisons with a series of other tests for additivity available in the literature. Some concluding remarks are made in Sect. 5. The proofs of the theoretical results obtained in Sect. 2 and 3 are deferred to the Appendix.

2 Formulations and definitions

2.1 Nonparametric hypothesis testing

Testing the hypotheses (6)–(7) is a nonparametric functional hypothesis testing problem and we start with a brief discussion of several main notions.

A (non-randomized) test ϕ is a measurable function of the observations with two values $\{0, 1\}$ corresponding to accepting and rejecting the null hypothesis H_0 respectively. The probability of a Type I error is defined as

$$\alpha(\phi) = \mathbb{P}_{f_0 \equiv 0}(\phi = 1),$$

while the probability of a Type II error for the composite nonparametric alternative hypothesis is defined as

$$\beta(\phi, \rho_n) = \sup_{f_0 \in \mathcal{F}(\rho_n)} \mathbb{P}_{f_0}(\phi = 0).$$

We focus on the asymptotic behavior of nonparametric hypothesis testing procedures as $n \to \infty$. For prescribed α and β, the rate of decay to zero of ρ_n, as $n \to \infty$, is a standard measure of asymptotical goodness of a test (see, e.g., Ingster 1982, 1993). The minimax rate of testing ρ_n is defined as follows:

Definition 1 A sequence ρ_n is the minimax rate of testing if $\rho_n \to 0$ as $n \to \infty$ and the following two conditions hold

(i) for any ρ'_n satisfying

$$\rho'_n/\rho_n = o_n(1),$$

one has

$$\inf_{\phi_n} \left[\alpha(\phi_n) + \beta(\phi_n, \rho'_n))\right] = 1 - o_n(1),$$

where $o_n(1)$ is a sequence tending to zero as $n \to \infty$.

(ii) for any $\alpha > 0$ and $\beta > 0$ there exists a constant $c > 0$ and a test ϕ_n^* such that

$$\alpha(\phi_n^*) \leq \alpha + o_n(1),$$
$$\beta(\phi_n^*, c\rho_n) \leq \beta + o_n(1).$$

The first condition in Definition 1 states that testing with a rate faster than ρ_n is impossible, while the second one guarantees the existence of a test with prescribed error probabilities if the distance between the null and the alternative set is of order ρ_n.

The following theorem establishes the minimax rate for testing additivity in the standard multiple nonparametric regression model (1)–(2), subject to identifiability conditions (3)–(5), i.e., testing the hypotheses (6)–(7).

Theorem 1 *Let the parameter $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$ be known, where $1 \leq p, q \leq \infty$, $sp > d$, $s > d/4$ for $p \geq 2$, and $M > 0$. Then, the minimax rate for testing the hypotheses (6)–(7) in the model (1)–(2), subject to identifiability conditions (3)–(5), is*

$$\rho_n = n^{-\frac{2s''}{4s''+d}} \quad as\ n \to \infty, \tag{8}$$

where $s'' = s - d/(2p') + d/4$ and $p' = \min(p, 2)$.

Remark 1 As a by-product, Theorem 1 provides the minimax rate of testing the presence of a signal in the standard d-dimensional ($d \geq 1$) nonparametric regression model against the composite nonparametric alternative hypothesis that it is separated away from zero and belongs to some Besov ball. In this setup, the rate (8) extends the d-dimensional results of Guerre and Lavergne (2002) for Hölder balls to Besov balls.

2.2 Multivariate wavelet series

In this section, we briefly recall some relevant issues on the multivariate orthonormal wavelet series on $[0, 1]^d$. Let ϕ be a scaling function and ψ a mother wavelet that form an orthonormal basis in $L^2[0, 1]$. For simplicity of exposition consider a periodic wavelet series, where, in particular, $\phi \equiv 1$. Despite the poor behaviour of periodic wavelets near the boundaries, where they might yield large coefficients for non-periodic functions, they are commonly used since their numerical implementation is particular simple and does not involve boundary corrections (see, e.g., Mallat 1999, Sect. 7.5).

For any integer $h = 0, 1, \ldots, 2^d - 1$ written in the binary form as $h = h_1 \ldots h_d$, define the corresponding d-dimensional function

$$\psi^h(\mathbf{x}) = \prod_{l=1}^{d} \psi(x_l)^{h_l} \phi(x_l)^{1-h_l} = \prod_{l=1}^{d} \psi(x_l)^{h_l}.$$

For convenience of further representation, re-arrange the order of $\psi^h(\mathbf{x})$ so that the first $d+1$ functions $\psi^h(\mathbf{x})$ will be

$$\psi^0(\mathbf{x}) = \prod_{l=1}^{d} \phi(x_l) \equiv 1, \quad \psi^h(\mathbf{x}) = \psi(x_h) \prod_{l \neq h} \phi(x_l) = \psi(x_h), \quad h = 1, \ldots, d. \tag{9}$$

For any $j \geq 0$, define the j-th index set $\mathcal{J}_j = \{(j, k_1, \ldots, k_d),\ k_i = 0, 1, \ldots, 2^j - 1;\ i = 1, \ldots, d\}$. Obviously, $\mathrm{Card}(\mathcal{J}_j) = 2^{jd}$. For each $I \in \mathcal{J}_j$, let

$$\psi_I^h(\mathbf{x}) = 2^{jd/2} \psi^h(2^j x_1 - k_1, \ldots, 2^j x_d - k_d).$$

The set of functions $\{\psi^0, \psi_I^h,\ h = 0, 1, \ldots, 2^d - 1,\ j \geq 0,\ I \in \mathcal{J}_j\}$ constitutes an orthonormal (periodic) wavelet basis in $[0, 1]^d$ (see, e.g., Mallat 1999, Sect. 7.7) and any f in the model (1) can be represented as

$$f(\mathbf{x}) = w_0 + \sum_{h=1}^{2^d-1} \sum_{j \geq 0} \sum_{I \in \mathcal{J}_j} w_I^h \psi_I^h(\mathbf{x}),$$

where

$$w_0 = \int_{[0,1]^d} f(\mathbf{x})\psi^0(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{[0,1]^d} f(\mathbf{x})\mathrm{d}\mathbf{x} \quad \text{and} \quad w_I^h = \int_{[0,1]^d} f(\mathbf{x})\psi_I^h(\mathbf{x})\mathrm{d}\mathbf{x}.$$

Using decomposition (2), identifiability conditions (3)–(5) and exploiting (9), f allows the following wavelet series representation

$$f(\mathbf{x}) = \underbrace{w_0}_{\mu} + \sum_{h=1}^{d} \underbrace{\sum_{j\geq 0} \sum_{I\in\mathcal{J}_j} w_I^h \psi_I^h(x_h)}_{f_h(x_h)} + \underbrace{\sum_{h=d+1}^{2^d-1} \sum_{j\geq 0} \sum_{I\in\mathcal{J}_j} w_I^h \psi_I^h(\mathbf{x})}_{f_0(\mathbf{x})}. \tag{10}$$

Multivariate wavelet series as defined above form unconditional bases in various Besov spaces on $[0, 1]^d$ and the Besov norm of a function is related to a sequence space norm of its wavelet coefficients. In particular, let the original (1-dimensional) mother wavelet ψ be of regularity $r > 0$. Then, in view of (10), for any $0 < s < r,\ 1 \leq p, q \leq \infty$ one has (see, e.g., Meyer 1992, Sect. 6.10)

$$||f_0||_{B^s_{p,q}} \asymp \begin{cases} \left(\sum_{j\geq 0} 2^{j(s+d/2-d/p)q} \left(\sum_{h=d+1}^{2^d-1} \sum_{I\in\mathcal{J}_j} |w_I^h|^p\right)^{q/p}\right)^{1/q}, & 1 \leq q < \infty, \\ \sup_{j\geq 0} \left\{2^{j(s+d/2-d/p)} \left(\sum_{h=d+1}^{2^d-1} \sum_{I\in\mathcal{J}_j} |w_I^h|^p\right)^{1/p}\right\}, & q = \infty. \end{cases} \tag{11}$$

In the wavelet domain, the null hypothesis (6) (additivity) is equivalent to

$$H_0: \sum_{h=d+1}^{2^d-1} \sum_{j=0}^{\infty} \sum_{I\in\mathcal{J}_j} (w_I^h)^2 = 0.$$

2.3 Discrete wavelet coefficients

In practice, however, one typically deals with *discrete* data as in the model (1). Assume for simplicity that the sample size is $n = 2^{dJ}$ for some integer $J > 0$. The corresponding d-dimensional orthogonal *discrete wavelet transform* (DWT) of the *sampled* function values $f(\mathbf{x}_i),\ i = 1, \ldots, n$, yields the set of DWT coefficients $\tilde{w}_I^h,\ h = 0, 1, \ldots, 2^d - 1,\ j = 0, 1, \ldots, J - 1,\ I \in \mathcal{J}_j$. For the considered periodic case and equispaced design, no boundary corrections and preconditioning data are needed for the DWT. The DWT coefficients $\tilde{w}_I^h$ provide a close approximation to the continuous wavelet coefficients w_I^h (see, e.g., Shenza 1992; Johnstone and Silverman 2004, 2005). In particular, from Proposition 5 of Johnstone and Silverman (2004) it follows that the equivalence between f_0's Besov norm and the sequence space norm of w_I^h in (11) remains true (up to probably different constants) for $\tilde{w}_I^h$ as well. In terms of the DWT coefficients, $\sum_{h=d+1}^{2^d-1}\sum_{j=0}^{J-1}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h)^2 = 0$ under the null hypothesis.

3 Main results

3.1 A Non-adaptive minimax test for additivity

We now construct a rate-optimal test for testing the hypotheses (6)–(7) where the parameters $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$ are assumed to be known and $1 \le p, q \le \infty,\ sp > d,\ s > d/4$ for $p \ge 2,\ M > 0$. We exploit the equivalence between the Besov norm of f_0 and the corresponding sequence space norm of its wavelet coefficients (11) and perform testing within the wavelet domain.

Given the sampled data in the model (1), where for simplicity of exposition we assume that $n = 2^{dJ}$ for some positive integer J, choose a mother wavelet ψ of regularity $r > s$ with $r-1$ vanishing moments, and take the corresponding d-dimensional DWT of (1). In the wavelet domain one then has

$$Y_I^h = \tilde{w}_I^h + \frac{\sigma}{\sqrt{n}}\,\xi_I^h, \quad I \in \mathcal{J}_j,\ \ j = 0, 1, \ldots, J-1,\ \ h = 0, 1, \ldots, 2^d - 1,$$

where Y_I^h are the empirical wavelet coefficients of the data, $\tilde{w}_I^h$ are the discrete wavelet transform of the sampled function values $\{f(\mathbf{x}_i)\}$, and ξ_I are independent $N(0, 1)$ random variables.

The test will be based on the set of $\{Y_I^h\}$. Define the levels j_n and j_θ as

$$j_n = (1/d)\log_2 n\ (= J), \quad j_\theta = \frac{2}{4s'' + d}\log_2 n, \tag{12}$$

where assume that j_θ in (12) is integer; otherwise, we take the corresponding integer part. Since $sp > d$ and $s > d/4$ for $p \ge 2$, one can easily verify that $j_\theta < j_n$.

Let $\mathcal{J}_n = \mathcal{J}_- \cup \mathcal{J}_+$, where $\mathcal{J}_- = \{0, 1, \ldots, j_\theta - 1\}$ (coarse levels) and $\mathcal{J}_+ = \{j_\theta, \ldots, j_n - 1\}$ (fine levels). For each $j \in \mathcal{J}_-$, define S_j to be

$$S_j = \sum_{h=d+1}^{2^d-1}\sum_{I\in\mathcal{J}_j}\left((Y_I^h)^2 - \frac{\sigma^2}{n}\right)$$

while, for each $j \in \mathcal{J}_+$ and for a given threshold $\lambda > 0$, define $S_j(\lambda)$ to be

$$S_j(\lambda) = \sum_{h=d+1}^{2^d-1}\sum_{I\in\mathcal{J}_j}\left((Y_I^h)^2\mathbf{1}\left\{|Y_I^h| > \frac{\sigma}{\sqrt{n}}\,\lambda\right\} - \frac{\sigma^2}{n}\,b(\lambda)\right),$$

where

$$b(\lambda) = \mathbb{E}\left[\xi^2 \mathbf{1}(|\xi| > \lambda)\right] = 2(\Phi(-\lambda) + \lambda\varphi(\lambda)),$$

$\mathbf{1}(A)$ is the indicator function of the set A, ξ is a standard normal $N(0, 1)$ random variable, and Φ and φ denote respectively the cumulative distribution and probability density functions of ξ.

With the above notations, introduce the following test statistics

$$T(j_\theta) = \sum_{j \in \mathcal{J}_-} S_j \tag{13}$$

and

$$Q(j_\theta) = \sum_{j \in \mathcal{J}_+} S_j(\lambda_j), \tag{14}$$

where

$$\lambda_j = 4\sqrt{d(j - j_\theta + 8)\ln 2}. \tag{15}$$

Let $V_0^2(j_\theta)$ and $W_0^2(j_\theta)$ be the variances of $T(j_\theta)$ and $Q(j_\theta)$, respectively, under H_0. It is easy to see that

$$V_0^2(j_\theta) = 2\,\frac{\sigma^4}{n^2}\,2^{j_\theta d} \quad \text{and} \quad W_0^2(j_\theta) = \frac{\sigma^4}{n^2} \sum_{j \in \mathcal{J}_+} 2^{jd} d(\lambda_j),$$

where

$$d(\lambda) = \mathbb{E}\left[\xi^4 \mathbf{1}(|\xi| > \lambda)\right] - b^2(\lambda)$$

and

$$\mathbb{E}\left[\xi^4 \mathbf{1}(|\xi| > \lambda)\right] = 6\Phi(-\lambda) + 2\lambda(3 + \lambda^2)\varphi(\lambda).$$

Finally, for a given significance level $\alpha \in (0, 1)$, define the following test

$$\phi_n^* = \mathbf{1}\left\{\frac{T(j_\theta) + Q(j_\theta)}{\sqrt{V_0^2(j_\theta) + W_0^2(j_\theta)}} > z_{1-\alpha}\right\}, \tag{16}$$

where $z_{1-\alpha}$ is $(1-\alpha)100\%$-th percentile of the $N(0, 1)$ distribution.

The resulting test statistic has a clear intuitive meaning and is essentially the standardized sum of squares of the thresholded empirical wavelet coefficients Y_I^h with the properly chosen level-dependent thresholds, where coefficients on the coarse levels $j \in \mathcal{J}_-$ are not thresholded. The resulting coefficients are centered to yield $\mathbb{E}(S_j) = 0$ and $\mathbb{E}(S_j(\lambda)) = 0$ under H_0. The null hypothesis is rejected when the above test statistic is large.

The following theorem establishes the asymptotic optimality of the proposed (non-adaptive) testing procedure ϕ_n^* defined in (16).

Theorem 2 *Let the mother wavelet ψ be of regularity $r > s$, and let the parameter $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$ be known, where $1 \le p, q \le \infty,\ sp > d$, $s > d/4$ for $p \ge 2$, and $M > 0$. Then, for a given significance level $\alpha \in (0, 1)$, the test ϕ_n^* defined in (16) for testing*

$$H_0 : f_0 \equiv 0 \quad versus \quad H_1 : f_0 \in \mathcal{F}(\rho_n) = \{f_0 : f_0 \in B^s_{p,q}(M),\ ||f_0||_2 \ge \rho_n\}$$

is a level-α asymptotically minimax test as $n \to \infty$. That is, for any $\beta \in (0, 1)$, it attains the minimax testing rate (8)

$$\rho_n = n^{-\frac{2s''}{4s''+d}} \quad as \quad n \to \infty,$$

where $s'' = s - d/(2p') + d/4$ and $p' = \min(p, 2)$.

Remark 2 For $p \ge 2$, which corresponds to “spatially homogeneous” functions whose wavelet coefficients are concentrated on coarse resolution levels, the above optimal test (16) can be simplified by truncating the wavelet series at level $j_\theta - 1$. Indeed, by working along the lines of Sect. 3.2.1 in Abramovich et al. (2004) for testing the presence of a signal in the 1-dimensional Gaussian white noise model, the level-α asymptotically minimax test ϕ_n^* for $p \ge 2$ can be then simplified as

$$\phi_n^* = \mathbf{1}\left\{\frac{T(j_\theta)}{V_0(j_\theta)} > z_{1-\alpha}\right\}.$$

3.2 An adaptive minimax test for additivity

The rate-optimal test derived in the previous section relies on the knowledge of the parameters $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$. In practice, however, they are typically unknown. We now consider the *adaptive* local testing problem where the above parameters are not specified *a priori* but assumed to lie within a given range. We first construct the adaptive test and then show its asymptotic optimality.

Assume now that $\theta = (s, p, q, M)$ are unknown but $d/4 < s \le s_{\max}$, $1 \le p \le p_{\max}$, $1 \le q \le \infty$, $sp > d$ and $0 < M_{\min} \le M \le M_{\max} < \infty$. Denote such a range of θ by $\mathcal{T}$. For each given set of parameters θ one may determine j_θ from (12). In fact, the range $\mathcal{T}$ derives essentially a range of admissible levels of the form $j_{\min} \le j_\theta \le j_{\max}$. One performs then a series of tests of type (16) for each admissible level and rejects the null hypothesis if it is rejected at least for one of them.

More precisely, recall that $j_n = (1/d)\log_2 n$ from (12), and let $j_{\min} = \frac{2}{4s''_{\max}+d}\log_2 n$, $j_{\max} = j_n - 1$ with $s''_{\max} = s_{\max} - d/(2p'_{\max}) + d/4$, where $p'_{\max} = \min(p_{\max}, 2)$. Again, we assume that the right-hand sides of j_n and $j_{\min}$ are integers; otherwise, we take the corresponding integer parts.

Choose a mother wavelet of regularity $r > s_{\max}$. Since the number of admissible levels is $\mathcal{O}(\ln n)$, a Bonferroni type correction for multiple testing leads to the following asymptotic adaptive test

$$\phi_n^a = \mathbf{1}\left\{\max_{j_{\min}\le j_\theta\le j_{\max}} \frac{T(j_\theta)+Q(j_\theta)}{\sqrt{V_0^2(j_\theta)+W_0^2(j_\theta)}} > \sqrt{2\ln\ln n}\right\}. \tag{17}$$

The following theorem establishes the asymptotic optimality of the proposed (adaptive) testing procedure ϕ_n^a defined in (17).

Theorem 3 *Let the mother wavelet ψ be of regularity $r > s_{\max}$, and let the parameter $\theta = (s, p, q, M)$ of the Besov ball $B^s_{p,q}(M)$ be unknown but $1 \le p \le p_{\max}$, $1 \le q \le \infty$, $sp > d$, $d/4 < s \le s_{\max}$, and $0 < M_{\min} \le M \le M_{\max} < \infty$. Then, the rate ρ_n of the test ϕ_n^a defined in (17) for testing*

$$H_0 : f_0 \equiv 0 \quad \textit{versus} \quad H_1 : f_0 \in \mathcal{F}(\rho_n) = \{f_0 : f_0 \in B^s_{p,q}(M),\ ||f_0||_2 \ge \rho_n\}$$

is

$$\rho_n = n^{-\frac{2s''}{4s''+d}}(\ln\ln n)^{\frac{s'}{4s''+d}} \quad \textit{as } n \to \infty,$$

where $s'' = s - d/(2p') + d/4$, $s' = s + d/2 - d/p'$ and $p' = \min(p, 2)$.

Moreover, there exists a constant c such that

$$\alpha(\phi_n^a) = o_n(1) \quad \text{and} \quad \sup_{\mathcal{T}} \beta(\phi_n^a, c\rho_n)) = o_n(1),$$

where $\mathcal{T}$ is the range of θ.

Theorem 3 establishes that the adaptive test ϕ_n^a is *nearly* rate-optimal (up to a ln ln n-factor). The results of Spokoiny (1996) indicate there is no adaptive testing without loss of efficiency of the test and such an extra log-log factor is unavoidable (though not essential) price for adaptivity. Furthermore, Theorem 3 demonstrates the degenerate behavior of the error probabilities of ϕ_n^a which is typical for adaptive testing (see, e.g., Ingster 1993; Spokoiny 1996).

Remark 3 Similar to the nonadaptive testing (see Remark 2), if, in addition, it is known that $p \ge 2$, the adaptive test ϕ_n^a for $p \ge 2$ can be simplified as

$$\phi_n^a = \mathbf{1}\left\{\max_{j_{\min}\le j_\theta\le j_{\max}} \frac{T(j_\theta)}{V_0(j_\theta)} > \sqrt{2\ln\ln n}\right\}.$$

4 Numerical experiments

We conducted a wide simulation study to evaluate the finite sample performance of the proposed adaptive minimax test (17) ($Amin$) for testing additivity in the standard 2-dimensional (bivariate) nonparametric regression setting. We have applied $Amin$ to a battery of test functions considered previously in the literature.

We have also compared $Amin$ with a series of other known procedures for testing additivity, namely the V_d and V_f tests of Eubank et al. (1995), the $\hat{T}_n^{(1,2)}$ (with order $l = 1$) test of Dette and Derbort (2001), the $M_{2,\alpha}$ (with order $l = 2$) test of Derbort et al. (2002), and the T_{1n}, T_{2n}, T_{3n} and T_{4n} tests of Dette et al. (2005). The V_d and V_f tests are based on truncated Fourier series estimation of the interactive term and its further testing for the significant difference from zero. The $\hat{T}_n^{(1,2)}$ and $M_{2,\alpha}$ tests use similar ideas but involve various estimates of the L^2-distance between the full model (2) and the additive model under the null hypothesis (6). The T_{1n}, T_{2n}, T_{3n} and T_{4n} tests are based on residuals from an internal marginal integration estimation.

The computational algorithms related to wavelet analysis were performed using Version 8 of the WaveLab toolbox for MATLAB that is freely available from http://www-stat.stanford.edu/software/software.html. The entire study was carried out using the MATLAB programming environment.

The test functions used in the simulation study are the following:

$$
\begin{aligned}
m_1(t_1, t_2) &= 0,\\
m_2(t_1, t_2) &= t_1 + t_2,\\
m_3(t_1, t_2) &= \exp(t_1) + \sin(\pi t_2),\\
m_4(t_1, t_2) &= \sin(\pi t_1) + \sin(\pi t_2),\\
m_5(t_1, t_2) &= \exp(t_1) + \exp(t_2),
\end{aligned}
$$

(see Derbort et al. 2002),

$$
\begin{aligned}
m_6(t_1, t_2) &= t_1\, t_2,\\
m_7(t_1, t_2) &= \exp(5(t_1 + t_2))\, /\, (1 + \exp(5(t_1 + t_2)))\ - 1,\\
m_8(t_1, t_2) &= 0.5\ (1 + \sin(2\pi(t_1 + t_2))),\\
m_9(t_1, t_2) &= 64\ (t_1 t_2)^3 (1 - t_1 t_2)^3,\\
m_{10}(t_1, t_2) &= (t_1 + t_2)\, /\, 2 + (1 \text{ outlier}),\\
m_{11}(t_1, t_2) &= G(t_1) G(t_2)/36,
\end{aligned}
$$

where

$$
G(t) = \begin{cases} 15t, & 0 \le t \le 0.2,\\ 5 - 10t, & 0.2 \le t \le 0.4,\\ -9 + 25t, & 0.4 \le t \le 0.6,\\ 18 - 20t, & 0.6 \le t \le 0.8,\\ -2 + 5t, & 0.8 \le t \le 1, \end{cases}
$$

(see Barry 1993; Eubank et al. 1995; Derbort et al. 2002), and

$$
m_{12}^{[ij]}(t_1, t_2) = h_i(t_1) + h_j(t_2) + \gamma\ h_i(t_1) h_j(t_2), \quad i, j = 1, 2, 3, 4, 5,
$$

$$
m_{13}^{[ij]}(t_1, t_2) = (h_i(t_1) + h_j(t_2))^{\delta}, \quad i, j = 1, 2, 3, 4, 5,
$$

where the parameters $\gamma \neq 0$ and $\delta \neq 1$ specify the deviation from additivity, and h_1, h_2, h_3, h_4 and h_5 are the Blip, Heavisine, Spikes, Corner and Doppler functions respectively (see, e.g., Antoniadis et al. 2001). We set $\gamma = 1, 2$ and $\delta = 1/2, 1/4$. The latter functions are examples of inhomogeneous functions often arising in signal processing.

For each test function the data were generated on a two-dimensional regular lattice $\{(t_{1i}, t_{2j}) : \; i = 0, 1, \ldots, n-1; \;\; j = 0, 1, \ldots, n-1\}$ in $[0, 1]^2$ for a small grid size ($n = 32$), a medium grid size ($n = 64$) and a larger grid size ($n = 256$). The comparative study of various testing procedures was performed by an empirical power analysis. The empirical power functions of the tests were computed from $M = 500$ Monte Carlo replications as a function of the signal-to-noise ratio (SNR), for SNR varying within the interval [0.01:0.04:2] (50 values overall).

The tests $T_{1n},\ T_{2n},\ T_{3n}$ and T_{4n} are based on the residuals of the Internalized Marginal Integration and the Internalized Nadaraya-Watson estimators of the response function. As it was pointed out also in Dette et al. (2005), for finite samples these methods involve cross-validation techniques for choosing tuning parameters and bootstrap for calculating quantiles for the corresponding test statistics that makes them computationally extremely expensive. To perform $T_{1n},\ T_{2n},\ T_{3n}$ and T_{4n} on 500 Monte Carlo replications for 50 values of SNR even for a grid size $n = 16$ required a tremendous computing time while the procedure computationally collapsed for $n = 32$. Moreover, for such a small grid size, the approximations of the quantiles were very poor. Due to these hard computational restrictions we decided to omit these tests from the study.

For the $Amin$ test for all test functions we used the ranges $2 \le j_\theta \le 4 (= \log_2(32) - 1)$ for $n = 32$, $2 \le j_\theta \le 5\ (= \log_2(64) - 1)$ for $n = 64$ and $2 \le j_\theta \le 7 (= \log_2(256) - 1)$ for $n = 256$ respectively. We should also note the adaptive test $Amin$ in (17) was based on the asymptotic properties of the maxima of weakly dependent Gaussian random variables, where the number of admissible resolution levels was supposed to be sufficiently large. Such a condition, obviously, holds only for very large samples. In practice, to perform $Amin$ at the given significant level α, one needs to derive the percentile $\zeta_{1-\alpha}$ of the *exact* distribution of the maxima under the null hypothesis:

$$\tilde{\phi}_n^a = \mathbf{1}\left\{ \max_{j_{\min} \le j_\theta \le j_{\max}} \frac{T(j_\theta) + Q(j_\theta)}{\sqrt{V_0^2(j_\theta) + W_0^2(j_\theta)}} > \zeta_{1-\alpha} \right\}.$$

Estimation of quantiles $\zeta_{1-\alpha}$ for various α was made by independent Monte Carlo studies. For the chosen ranges given above, the empirical quantiles were $\zeta_{0.90} = 2.03$, $\zeta_{0.95} = 2.52$ and $\zeta_{0.99} = 3.57$ for $n = 32$; $\zeta_{0.90} = 2.11$, $\zeta_{0.95} = 2.55$ and $\zeta_{0.99} = 3.46$ for $n = 64$, and $\zeta_{0.90} = 2.24$, $\zeta_{0.95} = 2.62$ and $\zeta_{0.99} = 3.40$ for $n = 256$. In all cases, the evaluation of the quantiles was based on 100,000 runs.

Finally, the noise level σ was also assumed unknown and estimated from the data. For the V_d, V_f, $M_{2,\alpha}$ and $\hat{T}_n^{(1,2)}$ tests, σ was estimated by various consistent estimators (see the corresponding papers for more details), while for the wavelet-based $Amin$ test, it was estimated by the median absolute deviation of the empirical wavelet

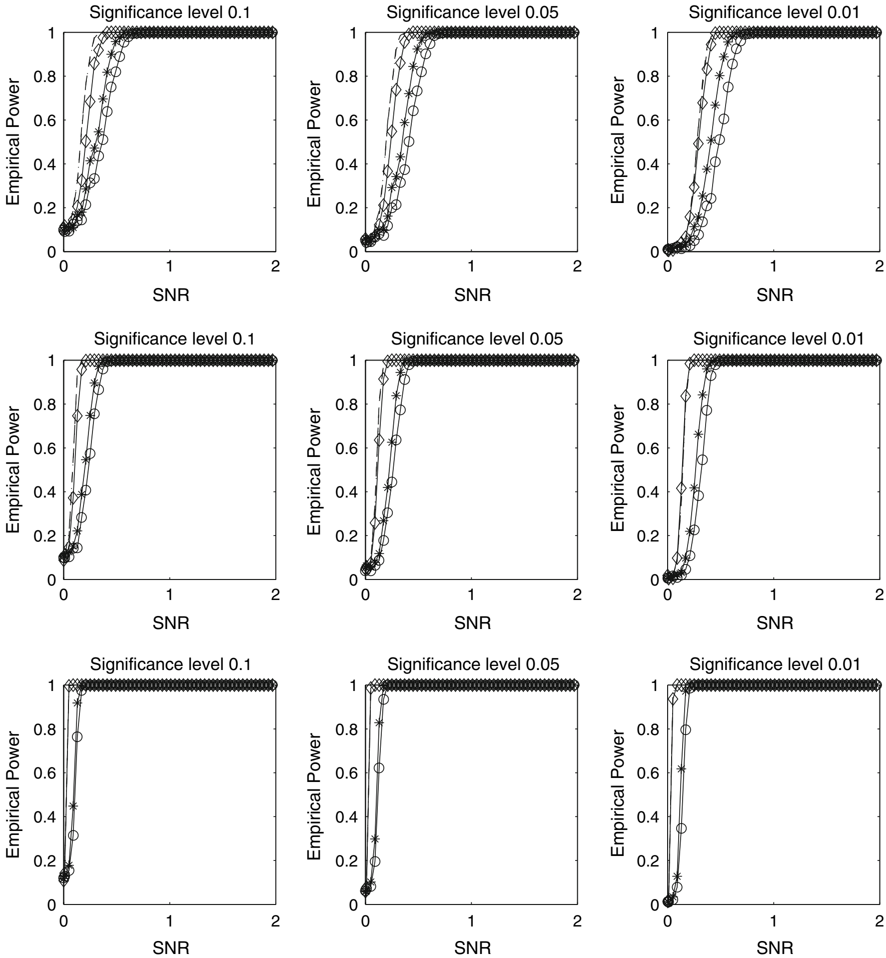

Fig. 1 Empirical power for the five tests for the test function m_9, $n = 32$ (first row), $n = 64$ (second row), and $n = 256$ (third row). The *dash* line refers to the V_d test; the *dash-dot* line refers to the V_f test; the *solid-circle* line refers to the $M_{2,\alpha}$ (with order $l = 2$) test; the *solid-asterisk* line refers to the $\hat{T}_n^{(1,2)}$ (with order $l = 1$) test; the *solid-diamond* line refers to the $Amin$ test

coefficients on the finest resolution level divided by 0.6745 as proposed by Donoho and Johnstone (1994).

Relative performances of the testing procedures for different test functions and various significance levels α were similar in magnitude and we report below only the results for m_9 (smooth function) and m_{13} ($i = 2$, $j = 4$, $\delta = 1/4$) (inhomogeneous function) for all significance levels α. Figure 1 shows the empirical powers of the simulation study for m_9, $n = 32$, $n = 64$ and $n = 256$, for the V_d, V_f, $M_{2,\alpha}$, $\hat{T}_n^{(1,2)}$ and $Amin$ tests, while Fig. 2 shows the same results for m_{13}. For the test function m_9, and for all chosen grid sizes $n = 32$, $n = 64$ and $n = 256$, $Amin$ competes with V_d and V_f, while $M_{2,\alpha}$ and $\hat{T}_n^{(1,2)}$ yield consistently worse results. For the test function

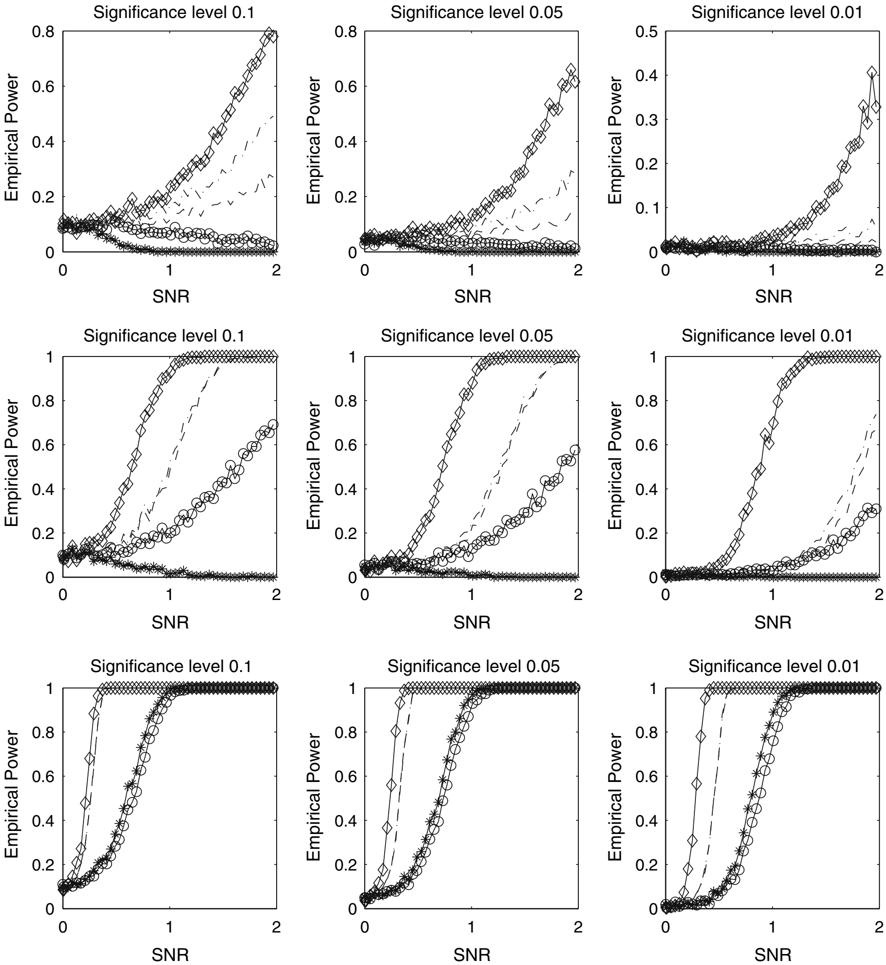

Fig. 2 Empirical power for the five tests for the test function m_{13}, $n = 32$ (first row), $n = 64$ (second row), and $n = 256$ (third row). The *dash* line refers to the V_d test; the *dash-dot* line refers to the V_f test; the *solid-circle* line refers to the $M_{2,\alpha}$ (with order $l = 2$) test; the *solid-asterisk* line refers to the $\hat{T}_n^{(1,2)}$ (with order $l = 1$) test; the *solid-diamond* line refers to the $Amin$ test

m_{13}, and for all grid sizes $n = 32$, $n = 64$ and $n = 256$, $Amin$ clearly outperforms all its counterparts. The reason of the different behavior of $Amin$ relatively to V_d and V_f for m_9 and m_{13} lies in the properties of Fourier and wavelet series expansions. It is well-known that Fourier series are more suitable for the analysis of smooth homogeneous functions while wavelets are much more appropriate for representing inhomogeneous signals. The poor performance of $M_{2,\alpha}$ and $\hat{T}_n^{(1,2)}$ for m_{13} is also due to the characteristics of these tests designed for very smooth functions.

Overall, the simulation results show good finite sample properties of the suggested wavelet-based minimax adaptive testing procedure especially applied to inhomogeneous response functions. From a computational point of view, $Amin$ resulted in the best performance. Note also that the considered grid sizes may be still too small to appreciate all the potentialities of the proposed (asymptotically in nature) minimax testing methodology.

5 Concluding remarks

We have considered the problem of testing for additivity in the standard multiple nonparametric regression model. We have derived asymptotically optimal (in the minimax sense) nonadaptive and adaptive tests for additivity against a wide set of alternatives. These tests are based on the empirical wavelet coefficients of the data and are computationally fast. The empirical power analysis of the tests in comparison with several existing counterparts demonstrates their good performance in finite sample situations, especially for inhomogeneous functions. The established asymptotic results obtained for the standard multiple nonparametric regression model can be straightforwardly modified to the corresponding d-dimensional Gaussian white noise model. For either model, the developed approach can be easily adapted for testing the significance of any *particular* interaction(s) of any order(s) by testing the significance of the corresponding subset of wavelet coefficients.

Acknowledgment This work was supported by a bilateral agreement between the Department of Mathematics and Statistics of the University of Cyprus and the School of Mathematical Sciences of Tel Aviv University. Felix Abramovich would like to thank Theofanis Sapatinas for warm hospitality while visiting Nicosia to carry out part of this work. Finally, we would like to thank an anonymous referee for suggestions on improvements to this paper.

Appendix

Throughout the proofs in the Appendix, we use C to denote a generic positive constant, not necessarily the same each time it is used, even within a single equation.

Proof of Theorem 1

We prove here only the lower bound, i.e., that the rate (8) in Theorem 1 cannot be improved. The upper bound is a consequence of Theorem 2 which shows the test ϕ_n^* proposed in (16) achieves this lower bound. The notations defined in Sect. 3.1 are adopted in the proof that follows.

We are going to prove that for any ρ_n' such that $c_n := \rho_n'/\rho_n = o_n(1)$ and for any test ϕ_n, one has

$$\inf_{\phi_n}\{\alpha(\phi_n)) + \beta(\phi_n, \rho_n')\} \geq 1 - o_n(1). \tag{18}$$

Following standard arguments, we replace the minimax problem by a Bayes problem. Let π_n be prior measures on the alternative set $\mathcal{F}(\rho_n') = \{f_0 : f_0 \in B^s_{p,q}(M),\ ||f_0||_2 \geq \rho_n'\}$ satisfying

$$\pi_n(\mathcal{F}(\rho_n')) \to 1, \quad \text{as } n \to \infty. \tag{19}$$

Let also $\mathbb{P}_{\pi_n}$ be the corresponding Bayes measure for the model (1), i.e., $\mathbb{P}_{\pi_n} = \int_{[0,1]^d} \mathbb{P}_f \pi_n \mathrm{d}f$, and define $L_{\pi_n} = \frac{\mathrm{d}\mathbb{P}_{\pi_n}}{\mathrm{d}\mathbb{P}_0}$, where the measure $\mathbb{P}_0$ corresponds to the null hypothesis. Ingster (1993) showed that to prove (18) it is sufficient to verify that, under the $\mathbb{P}_0$-measure, one has

$$L_{\pi_n} \to 1 \quad \text{as } n \to \infty. \tag{20}$$

We now construct the measures π_n satisfying (19) and (20).

Consider first the case $p \geq 2$. Note that in this case $p' = 2$ and $s'' = s$ and recall from (12) that $j_\theta = \frac{2}{4s''+d} \log_2 n$. Define a random function

$$f_0(\mathbf{x}) = u_n\, c_n \sum_{h=d+1}^{2^d-1} \sum_{j \in \mathcal{J}_-} \sum_{I \in \mathcal{J}_j} \xi_I^h \psi_I^h(\mathbf{x}),$$

where $u_n = n^{-(2s+d)/(4s+d)}$ and ξ_I^h are independent random variables taking the values ± 1 with probabilities 1/2. Let a measure π_n correspond to the distribution of such a random function f_0.

Exploiting the embedding properties of Besov spaces and the equivalence between the Besov norm of a function and the corresponding sequence space norm of its wavelet coefficients (11), for j_θ and u_n defined above, we have

$$\begin{aligned} ||f_0||_{B^s_{p,q}} &\leq C||f_0||_{B^s_{p,1}} \asymp u_n c_n \sum_{j \in \mathcal{J}_-} 2^{j(s+d/2-d/p)} \left(\sum_{h=d+1}^{2^d-1} \sum_{I \in \mathcal{J}_j} |\xi_I^h| \right)^{\frac{1}{p}} \\ &= u_n c_n l_d 2^{j_\theta(s+d/2)} < \infty, \end{aligned}$$

where $l_d = \#\{h : h = d+1, d+2, \ldots, 2^d - 1\} = 2^d - d - 1$. Furthermore, by Parseval's identity, one can easily verify that $||f_0||_2 = C c_n \rho_n = C\rho_n'$. Hence, for the constructed measures, $\pi_n\{f_0 \in \mathcal{F}(\rho_n')\} = 1$.

Consider now the case $p < 2$, where $p' = p$ and $s'' = s - d/(2p) + d/4$. Consider a random function

$$f_0(\mathbf{x}) = n^{-1/2} c_n \sum_{h=d+1}^{2^d-1} \sum_{j \in \mathcal{J}_-} \sum_{I \in \mathcal{J}_j} \xi_I^h \psi_I^h(\mathbf{x}),$$

where ξ_I^h are now independent random variables taking the values ± 1 and 0 with probabilities $p_I^h/2$ and $1 - p_I^h$ respectively, with $p_I^h = 2^{-j_\theta d/2}$. Let a measure π_n corresponds to the distribution of such a random function f_0. By the law of large numbers,

$$l_d^{-1} 2^{-jd} 2^{j_\theta d/2} \sum_{h=d+1}^{2^d-1} \sum_{I \in \mathcal{J}_j} |\xi_I^h| \xrightarrow{P} 1. \tag{21}$$

Similarly to the case $p \geq 2$, using (21), a straightforward calculus implies that, with high probability, $||f_0||_{B^s_{p,1}} < \infty$ and $||f_0||_2 = Cc_n\rho_n = C\rho_n'$. Hence, the constructed measures π_n satisfy (19).

Following the arguments of Ingster (1993) with obvious modifications for the case $d > 1$, one can show that, for each of the above prior measures π_n, (20) holds. Thus, the lower bound is obtained. This, together with the upper bound obtained in Theorem 2, completes the proof of Theorem 1. □

Proofs of Theorems 2 and 3

In order to prove Theorems 2 and 3, we need the following lemmas that provide the necessary bounds for the statistics $\mathbb{E}_\mu(T(j_\theta) + Q(j_\theta))$ and $\sqrt{\mathbb{V}_\mu(T(j_\theta) + Q(j_\theta))}$, where $\mathbb{E}_\mu$ and $\mathbb{V}_\mu$ denotes the expectation and variance operators respectively when the true parameter is μ.

Lemma 1 *Let $\mathcal{F}(\rho_n)$, $T(j_\theta)$ and $Q(j_\theta)$ be defined as in (7), (13) and (14), respectively. Then, for any $f_0 \in \mathcal{F}(\rho_n)$,*

$$\mathbb{E}_{f_0}(T(j_\theta) + Q(j_\theta)) \geq \frac{1}{2}||f_0||_2^2 - C\left(n^{-\frac{4s''}{4s''+d}} + n^{-\frac{2s'}{d}}\right)$$

where $s'' = s - d/(2p') + d/4$, $s' = s + d/2 - d/p'$ and $p' = \min(p, 2)$.

Proof of Lemma 1 Obviously,

$$\mathbb{E}_{f_0}(T(j_\theta)) = \sum_{h=d+1}^{2^d-1} \sum_{j \in \mathcal{J}_-} \sum_{I \in \mathcal{J}_j} (\tilde{w}_I^h)^2. \tag{22}$$

Using Lemma 4.4 in Spokoiny (1996), one has

$$\mathbb{E}_{f_0}\left[(Y_I^h)^2 \mathbf{1}\left\{|Y_I^h| > \frac{\sigma}{\sqrt{n}}\lambda_j\right\} - \frac{\sigma^2}{n} b(\lambda)\right] \geq \frac{1}{2}(\tilde{w}_I^h)^2 \mathbf{1}\left\{|\tilde{w}_I^h| > \frac{\sigma}{\sqrt{n}}\lambda_j\right\}$$

and therefore,

$$\mathbb{E}_{f_0}(Q(j_\theta)) \geq \frac{1}{2} \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \sum_{I\in\mathcal{J}_j} (\tilde{w}_I^h)^2 \mathbf{1}\left\{|\tilde{w}_I^h| \geq \frac{\sigma}{\sqrt{n}}\lambda_j\right\}$$
$$\geq \frac{1}{2}\left(\sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \sum_{I\in\mathcal{J}_j} (\tilde{w}_I^h)^2 - \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \sum_{I\in\mathcal{J}_j} (\tilde{w}_I^h)^2 \mathbf{1}\left\{|\tilde{w}_I^h| \leq \frac{\sigma}{\sqrt{n}}\lambda_j\right\}\right). \quad (23)$$

A straightforward multidimensional extension of (59) from Johnstone and Silverman (2005) implies that, for any $f_0 \in B^s_{p,q}(M)$,

$$\sum_{I\in\mathcal{J}_j} |\tilde{w}_I^h|^{p'} \leq C2^{-js'p'} \quad \text{for all } j \in \mathcal{J}_n.$$

Thus, similar to Spokoiny (1996), using the definition (12) for j_θ, one has

$$\sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \sum_{I\in J_j} (\tilde{w}_I^h)^2 \mathbf{1}\left\{|\tilde{w}_I^h| \leq \frac{\sigma}{\sqrt{n}}\lambda_j\right\} \leq \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \sum_{I\in\mathcal{J}_j} |\tilde{w}_I^h|^{p'} \times \left(\frac{\sigma}{\sqrt{n}}\lambda_j\right)^{2-p'} \mathbf{1}\left\{|\tilde{w}_I^h| \leq \frac{\sigma}{\sqrt{n}}\lambda_j\right\}$$
$$\leq n^{-(1-p'/2)} \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_+} \lambda_j^{2-p'} \sum_{I\in\mathcal{J}_j} |\tilde{w}_I^h|^{p'}$$
$$\leq Cn^{-(1-p'/2)} 2^{-j_\theta s'p'} = Cn^{-\frac{4s''}{4s''+s}}. \quad (24)$$

Note that

$$\sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_n} \sum_{I\in\mathcal{J}_j} (\tilde{w}_I^h)^2 \geq \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_n} \sum_{I\in\mathcal{J}_j} (w_I^h)^2 - \frac{1}{2} \sum_{h=d+1}^{2^d-1} \sum_{j\in\mathcal{J}_n} \sum_{I\in\mathcal{J}_j} (\tilde{w}_I^h - w_I^h)^2, \quad (25)$$

where the d-dimensional extension (adapted to periodic boundary conditions) of (70) of Johnstone and Silverman (2005) implies, in particular, that the second term in the right-hand side of (25) is $O(n^{-2s'/d})$. Furthermore, from (11) and the fact that l_r norms decrease as r increases, it follows that

$$\sum_{h=d+1}^{2^d-1}\sum_{j=j_n}^{\infty}\sum_{I\in\mathcal{J}_j}(w_I^h)^2 \le \sum_{h=d+1}^{2^d-1}\sum_{j=j_n}^{\infty}\left(\sum_{I\in\mathcal{J}_j}|w_I^h|^{p'}\right)^{2/p'} \le Cn^{-\frac{2s'}{d}}. \qquad (26)$$

Hence, combining (22)–(26), we get

$$\begin{aligned}\mathbb{E}_{f_0}(T(j_\theta)+Q(j_\theta)) &\ge \frac{1}{2}\left(||f_0||_2^2 - \sum_{h=d+1}^{2^d-1}\sum_{j=j_n}^{\infty}\sum_{I\in\mathcal{J}_j}(w_I^h)^2\right.\\ &\quad\left. - \sum_{h=d+1}^{2^d-1}\sum_{j\in\mathcal{J}_-}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h - w_I^h)^2\right)\\ &\quad - \frac{1}{2}\sum_{h=d+1}^{2^d-1}\sum_{j\in\mathcal{J}_+}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h)^2\mathbf{1}\left\{|\tilde{w}_I^h| \le \frac{\sigma}{\sqrt{n}}\lambda_j\right\}\\ &\ge \frac{1}{2}||f_0||_2^2 - C\left(n^{-\frac{4s''}{4s''+d}} + n^{-\frac{2s'}{d}}\right).\end{aligned}$$

This completes the proof of Lemma 1. □

Lemma 2 *Let $\mathcal{F}(\rho_n)$, $T(j_\theta)$ and $Q(j_\theta)$ be defined as in (7), (13) and (14), respectively. Then, for any $f_0 \in \mathcal{F}(\rho_n)$,*

$$\mathbb{V}_{f_0}(T(j_\theta)+Q(j_\theta)) \le C\left(n^{-1}||f_0||_2^2 + n^{-\frac{2s'}{d}} + n^{-\frac{8s''}{4s''+d}}\right),$$

where $s'' = s - d/(2p') + d/4$, $s' = s + d/2 - d/p'$ and $p' = \min(p, 2)$.

Proof of Lemma 2 From the well-known properties of the non-central χ^2-distribution, one has

$$\mathbb{V}_{f_0}(T(j_\theta)) = 2\sigma^4 n^{-2}2^{j_\theta d} + 4\sigma^2 n^{-1}\sum_{h=d+1}^{2^d-1}\sum_{j\in J_-}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h)^2, \qquad (27)$$

where using the definition (12) for j_θ one can easily verify that the first term in the right-hand side of (27) is $O(n^{-8s''/(4s''+d)})$.

Consider now $\mathbb{V}_f(Q(j_\theta))$. Exploiting Lemma 4.5 in Spokoiny (1996), we get

$$\mathbb{V}_f\left((Y_I^h)^2\mathbf{1}\left\{|Y_I^h| > \frac{\sigma}{\sqrt{n}}\lambda_j\right\}\right) \le 4\sigma^2 n^{-1}(\tilde{w}_I^h)_I^2 + 2n^{-2}\sigma^4\mathbf{1}\left\{|\tilde{w}_I^h| > \frac{\sigma}{\sqrt{n}}\,\lambda_j/2\right\}$$
$$+\sigma^4 n^{-2}\lambda_j^4 e^{-\lambda_j^2/8}. \tag{28}$$

A straightforward calculus now yields

$$\sum_{h=d+1}^{2^d-1}\sum_{j\in\mathcal{J}_+}\sum_{I\in\mathcal{J}_j}\mathbf{1}\left\{|\tilde{w}_I^h| > \frac{\sigma}{\sqrt{n}}\frac{\lambda_j}{2}\right\} \le \sum_{h=d+1}^{2^d-1}\sum_{j\in\mathcal{J}_+}\sum_{I\in\mathcal{J}_j}\left(\frac{\sigma}{\sqrt{n}}\frac{\lambda_j}{2}\right)^{-p'}|\tilde{w}_I^h|^{p'}$$
$$\le C\, n^{p'/2}2^{-j_\theta s'p'} = C\, n^{-\frac{8s''+d}{4s''+d}}. \tag{29}$$

Furthermore, for the thresholds λ_j defined in (15), we have

$$\sigma^4 n^{-2}\sum_{h=d+1}^{2^d-1}\sum_{j\in\mathcal{J}_+}\sum_{I\in\mathcal{J}_j}\lambda_j^4 e^{-\lambda_j^2/8} \le Cn^{-2}2^{j_\theta d} = Cn^{-\frac{8s''}{4s''+d}}. \tag{30}$$

Combining (28), (29) and (30), we obtain the following upper bound on $\mathbb{V}_{f_0}(Q(j_\theta))$

$$\mathbb{V}_{f_0}(Q(j_\theta)) \le C\left(n^{-1}\sum_{h=d+1}^{2^d-1}\sum_{j\in J_+}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h)^2 + n^{-\frac{8s''}{4s''+d}}\right). \tag{31}$$

By combining (27) and (31), exploiting again the d-dimensional extension (adapted to periodic boundary conditions) of (70) of Johnstone and Silverman (2005), and noting that

$$\sum_{h=d+1}^{2^d-1}\sum_{j\in J_n}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h)^2 \le 2\left(\sum_{h=d+1}^{2^d-1}\sum_{j\in J_n}\sum_{I\in\mathcal{J}_j}(w_I^h)^2 + \sum_{h=d+1}^{2^d-1}\sum_{j\in J_n}\sum_{I\in\mathcal{J}_j}(\tilde{w}_I^h - w_I^h)^2\right)$$
$$\le 2||f_0||^2 + Cn^{-\frac{2s'}{d}},$$

we easily get

$$\mathbb{V}_{f_0}(T(j_\theta) + Q(j_\theta)) \le C\left(n^{-1}||f_0||_2^2 + n^{-\frac{2s'}{d}} + n^{-\frac{8s''}{4s''+d}}\right).$$

This completes the proof of Lemma 2. □

We are now in the position to prove the main Theorems 2 and 3.

Proof of Theorem 2 The statistics $T(j_\theta)$ and $Q(j_\theta)$ are the sums of respectively j_θ and $j_n - j_\theta$ independent, squared integrable, random variables that under the null hypothesis have zero means and variances $V_0^2(j_\theta)$ and $W_0^2(j_\theta)$. By the central limit theorem, the resulting standardized test statistic in (16) as $n \to \infty$ is then asymptotically standard normal and the significance level of ϕ_n^* is therefore asymptotically α.

Consider now the Type II error of the test. It is straightforward to see that, for any specific $f_0 \in \mathcal{F}(\rho_n)$, asymptotically one has

$$\mathbb{P}_{f_0}(\phi_n^* = 0) = \Phi\left(\sqrt{\frac{V_0^2(j_\theta) + W_0^2(j_\theta)}{\mathbb{V}_{f_0}(T(j_\theta) + Q(j_\theta))}} z_{1-\alpha} - \frac{\mathbb{E}_{f_0}(T(j_\theta) + Q(j_\theta))}{\sqrt{\mathbb{V}_{f_0}(T(j_\theta)) + Q(j_\theta))}}\right) + o_n(1).$$

Since the variances ratio $(V_0^2(j_\theta) + W_0^2(j_\theta))/(\mathbb{V}_{f_0}(T(j_\theta) + Q(j_\theta))$ is evidently bounded above by one, the asymptotic behavior of $\mathbb{P}_{f_0}(\phi_n^* = 0)$ depends only on the ratio $\mathbb{E}_{f_0}(T(j_\theta) + Q(j_\theta))/\sqrt{\mathbb{V}_{f_0}(T(j_\theta) + Q(j_\theta))}$.

Recall that $||f_0||_2 \geq \rho_n$ for all $f_0 \in \mathcal{F}(\rho_n)$. Substituting $\rho_n = n^{-2s''/(4s''+d)}$, Lemmas 1 and 2 imply therefore that, for sufficiently large n, for any given β, there exists a constant c_β such that

$$\inf_{f_0 \in \mathcal{F}(c_\beta \rho_n)} \frac{\mathbb{E}_{f_0}(T(j_\theta) + Q(j_\theta))}{\sqrt{\mathbb{V}_{f_0}(T(j_\theta) + Q(j_\theta))}} > \tilde{c}_\beta,$$

where $\tilde{c}_\beta > 0$ satisfies $\Phi(z_{1-\alpha} - \tilde{c}_\beta) = \beta$ and, hence, $\tilde{c}_\beta = z_{1-\alpha} + z_{1-\beta}$. Thus,

$$\beta(\phi_n^*, c_\beta \rho_n)) \leq \beta + o_n(1),$$

showing that the test ϕ_n^* achieves the lower bound (8) for the minimax rate and it is, therefore, rate-optimal. This completes the proof of Theorem 2. □

Proof of Theorem 3

Under the null hypothesis, $\{(T(j_\theta) + Q(j_\theta))/\sqrt{V_0^2(j_\theta) + W_0^2(j_\theta)},\ j_{\min} \leq j_\theta \leq j_{\max}\}$ is a sequence of $\mathcal{O}(\ln n)$ weakly dependent, asymptotically $N(0, 1)$ random variables. Applying the well-known extreme value results for asymptotically $N(0, 1)$ random variables (see, e.g., Leadbetter et al. 1986, Chap. 4), one has

$$\alpha(\phi_n^a) = \mathbb{P}_{f_0 \equiv 0}\left\{\max_{j_{\min} \leq j_\theta \leq j_{\max}} \frac{T(j) + Q(j)}{\sqrt{V_0^2(j) + W_0^2(j)}} > \sqrt{2 \ln \ln n}\right\} \to 0 \quad \text{as } n \to \infty.$$

Choose now any set of parameters $\theta = (s, p, q, M) \in \mathcal{T}$ and define

$$j_\theta^* = \frac{2}{4s'' + d} \log_2 \left(n (\ln \ln n)^{-1/p'}\right).$$

For any f_0 from the alternative set, we then have

$$\mathbb{P}_{f_0}(\phi_n^a = 0) \le \mathbb{P}_{f_0}\left(\frac{T(j_\theta^*) + Q(j_\theta^*)}{\sqrt{V_0^2(j_\theta^*) + W_0^2(j_\theta^*)}} \le \sqrt{2 \ln \ln n}\right)$$
$$\le \Phi\left(\sqrt{2 \ln \ln n} - \frac{\mathbb{E}_{f_0}(T(j_\theta^*) + Q(j_\theta^*))}{\sqrt{\mathbb{V}_{f_0}(T(j_\theta^*) + Q(j_\theta^*))}}\right) + o_n(1). \qquad (32)$$

Repeating the arguments in the proof of Theorem 2, and substituting j_θ^* and $\rho_n = n^{-\frac{2s''}{4s''+d}} (\ln \ln n)^{\frac{s'}{4s''+d}}$ in (32), straightforward calculus imply that it is always possible to find a constant c such that, for any $f_0 \in \mathcal{F}(c\rho_n)$,

$$\frac{\mathbb{E}_{f_0}(T(j_\theta^*) + Q(j_\theta^*))}{\sqrt{\mathbb{V}_{f_0}(T(j_\theta^*) + Q(j_\theta^*))}} > \sqrt{2 \ln \ln n}.$$

Hence,

$$\beta(\phi_n^a, c\rho_n) = \sup_{f_0 \in \mathcal{F}(c\rho_n)} P_{f_0}(\phi_n^a = 0) \to 0, \quad \text{as } n \to \infty.$$

This completes the proof of Theorem 3. □

References

Abramovich, F., Antoniadis, A., Sapatinas, T., Vidakovic, B. (2004). Optimal testing in a fixed-effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing*, *2*, 323–349.

Amato, U., Antoniadis, A. (2001). Adaptive wavelet series estimation in separable nonparametric regression models. *Statistics and Computing*, *11*, 373–394.

Amato, U., Antoniadis, A., De Feis, I. (2002). Fourier series approximation of separable models. *Journal of Computational and Applied Mathematics*, *146*, 459–479.

Antoniadis, A., Bigot, J., Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software*, *6*(6), 1–83.

Barry, D. (1993). Testing for additivity of a regression function. *Annals of Statistics*, *21*, 235–254.

Buja, A., Hastie, T., Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, *17*, 453–555.

De Canditiis, D., Sapatinas, T. (2004). Testing for additivity and joint effects in multivariate nonparametric regression using Fourier and wavelet methods. *Statistics and Computing*, *14*, 235–249.

Derbort, S., Dette, H., Munk, A. (2002). A test for additivity in nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *54*, 60–82.

Dette, H., Derbort, S. (2001). Analysis of variance in nonparametric regression models. *Journal of Multivariate Analysis*, *76*, 110–137.

Dette, H., Wilkau, C. V. U. (2001). Testing additivity by kernel-based methods—what is a reasonable test? *Bernoulli*, *7*, 669–697.

Dette, H., Wilkau, C. V. U., Sperlich, S. (2005). A comparison of different nonparametric methods for inference on additive models. *Journal of Nonparametric Statistics*, *17*, 57–81.

Donoho, D. L., Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, *81*, 425–455.

Eubank, R. L., Hart, J. D., Simpson, D. G., Stefanski, L. A. (1995). Testing for additivity in nonparametric regression. *Annals of Statistics*, *23*, 1896–1920.

Fan, J., Jiang, J. (2005). Nonparametric inferences for additive models. *Journal of the Americal Statistical Association*, *100*, 890–907.
Gozalo, P. L., Linton, O. B. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics*, *104*, 1–48.
Guerre, E., Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, *18*, 1139–1171.
Härdle, W., Sperlich, S., Spokoiny, V. (2001). Structural tests in additive regression. *Journal of the American Statistical Association*, *96*, 1333–1347.
Hastie, T. J., Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman and Hall.
Horowitz, J. L., Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, *69*, 599–631.
Ingster, Yu. I. (1982). Minimax nonparametric detection of signals in white Gaussian noise. *Problems of Information Transmission*, *18*, 130–140.
Ingster, Yu. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives: I, II, III. *Mathematical Methods of Statistics*, *2*, 85–114; *3*, 171–189; *4*, 249–268.
Ingster, Yu. I., Suslina, I. A. (2003). *Nonparametric goodness-of-fit testing under Gaussian models*. Lecture Notes in Statistics, *169*, New York: Springer.
Ingster, Yu. I., Suslina, I. A. (2005). On estimation and detection of smooth functions of many variables. *Mathematical Methods of Statistics*, *14*, 299–331.
Johnstone, I. M., Silverman, B. W. (2004). Boundary coiflets for wavelet shrinkage in function estimation. *Journal of Applied Probability*, *41A*, 81–98.
Johnstone, I. M., Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, *33*, 1700–1752.
Leadbetter, M. R., Lindgren, G., Rootzén, H. (1986). *Extremes and related properties of random sequences and processes*. New York: Springer.
Lepski, O. V., Spokoiny, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, *5*, 333–358.
Li, Q., Hsiao, C., Zinn, J. (2003). Consistent specification tests for semiparametric/nonparametric models based on series of estimation methods. *Journal of Econometrics*, *112*, 295–325.
Linton, O. B. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika*, *84*, 469–473.
Linton, O. B., Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, *82*, 93–100.
Mallat, S. G. (1999). *A wavelet tour of signal processing* (2nd edn.). San Diego: Academic Press.
Meyer, Y. (1992). *Wavelets and operators*. Cambridge: Cambridge University Press.
Opsomer, J. D., Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics*, *25*, 186–211.
Opsomer, J. D., Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association, 93*, 605–619.
Sardy, S., Tseng, P. (2004). AMlet, RAMlet, and GAMlet: automatic nonlinear fitting of additive models, robust and generalized, with wavelets. *Journal of Computational and Graphical Statistics*, *13*, 283–309.
Shensa, M. J. (1992). The discrete wavelet transform: wedding the á trous and Mallat algorithms. *IEEE Transactions on Signal Processing*, *40*, 2464–2482.
Sperlich, S., Linton, O. B., Härdle, W. (1999). Integration and backfitting methods in additive models—finite sample properties and comparison. *Test*, *8*, 419–458.
Sperlich, S., Tjøstheim, D., Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, *18*, 197–251.
Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *Annals of Statistics*, *24*, 2477–2498.
Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, *13*, 689–705.
Tukey, J. (1949). One degree of freedom test for non-additivity. *Biometrics*, *5*, 232–242.
Zhang, S., Wong, M.-Y. (2003). Wavelet threshold estimation for additive regression models. *Annals of Statistics*, *31*, 152–173.